



ISTITUTO NAZIONALE DI FISICA NUCLEARE

Laboratori Nazionali di Frascati

INFN-12-10/LNF

July 09, 2012

**Fondamenti Matematici Della Fisica Macroscopica
(un percorso geometrico), Parte I**

Camillo Lo Surdo

INFN, Laboratori Nazionali di Frascati, P.O. Box 13, I-00044 Frascati, Italy

*Published by SIDS-Pubblicazioni
Laboratori Nazionali di Frascati*

FONDAMENTI MATEMATICI DELLA FISICA
MACROSCOPICA
(UN PERCORSO GEOMETRICO)

C. LO SURDO

INDICE

- 0.0 PRESENTAZIONE
 - 0.0.1 CONSIDERAZIONI GENERALI E PIANO DI LAVORO
 - 0.0.2 I CONTENUTI CAPITOLO PER CAPITOLO: SINTESI E COMMENTI
 - 0.0.3 “ISTRUZIONI PER L’USO”
- 0.1 INTRODUZIONE: ALCUNE RIFLESSIONI SULLE TEORIE FISICO-MATEMATICHE
 - 0.1.1 PREMessa
 - 0.1.2 RAGIONAMENTO ASSIOMATICO-DEDUTTIVO VS. RAGIONAMENTO INDUTTIVO
 - 0.1.3 MATEMATICA SIGNIFICANTE VS. MATEMATICA NON SIGNIFICANTE
 - 0.1.4 SCIENZE MATEMATIZZABILI VS. SCIENZE NON MATEMATIZZABILI, E ALTRI COMMENTI

PARTE PRIMA

GEOMETRIA EUCLIDEA E PSEUDOEUCLEIDEA

CAP.1 LA GEOMETRIA EUCLIDEA

- 1.1 LA GEOMETRIA COME PROTOTIPO DI TEORIA FISICO-MATEMATICA
 - 1.1.1 GENERALITÀ E INQUADRAMENTO STORICO
 - 1.1.2 GEOMETRIA EUCLIDEA SINTETICA E ANALITICA
- 1.2 UNA FORMALIZZAZIONE METRICA DELLA GEOMETRIA EUCLIDEA
 - 1.2.1 INTRODUZIONE
 - 1.2.2 I PRIMI NOVE ASSIOMI E LA GEOMETRIA “SPECIALE” DELLA RETTA
 - 1.2.3 I SUCCESSIVI OTTO ASSIOMI E LE GEOMETRIE DIADICHE DEL PIANO E DELLO SPAZIO
 - 1.2.4 GLI ULTIMI DUE ASSIOMI: LA GEOMETRIA ASSOLUTA E LA GEOMETRIA EUCLIDEA
- 1.3 L’ISOMORFISMO TRA H_n E LO SPAZIO CARTESIANO REALE N DIMENSIONALE \mathbf{R}^n
 - 1.3.1 \mathbf{R}^n COME MODELLO NORMALE DI H_n : PARTE PRIMA
 - 1.3.2 \mathbf{R}^n COME MODELLO NORMALE DI H_n : PARTE SECONDA
- 1.4 QUESTIONI DI ESTENSIONE E ORIENTAMENTO, CONCLUSIONI
 - 1.4.1 ESTENSIONE
 - 1.4.2 ORIENTAMENTO
 - 1.4.3 CONSIDERAZIONI CONCLUSIVE

APP. SPEC. CAP. 1

- 1.A LE RELAZIONI D'ORDINE
- 1.B LE FUNZIONI Cos E Sin

CAP. 2 LA GEOMETRIA PSEUDOLUCLIDEA

- 2.1 PREMESSA: LO SPAZIO-TEMPO DI EUCLIDE-NEWTON
 - 2.1.1 ELEMENTI DI CINEMATICA EUCLIDEA-NEWTONIANA
 - 2.1.2 ELEMENTI DI DINAMICA NEWTONIANA
- 2.2 INTRODUZIONE ALLA CINEMATICA E ALLA DINAMICA RELATIVISTICHE SPECIALI
 - 2.2.1 CINEMATICA RELATIVISTICA SPECIALE E TRASFORMAZIONI DI LORENTZ
 - 2.2.2 RUDIMENTI DI DINAMICA RELATIVISTICA SPECIALE
- 2.3 INTRODUZIONE ALLA GEOMETRIA PSEUDOEUCLEIDEA
 - 2.3.1 GLI SPAZI PSEUDOEUCLEIDEI (NOZIONI DI BASE)
 - 2.3.2 LO SPAZIO DI MINKOWSKI (NOZIONI DI BASE)
- 2.4 ALGEBRA E GEOMETRIA DELLO SPAZIO DI MINKOWSKI
 - 2.4.1 PARTE PRIMA
 - 2.4.2 PARTE SECONDA
- 2.5 INTRODUZIONE ALL'ELETTROMAGNETISMO MAXWELLIANO
 - 2.5.1 LE EQUAZIONI DI MAXWELL-LORENTZ
 - 2.5.2 LA TEORIA EM NEL LINGUAGGIO DEL CALCOLO TENSORIALE
 - 2.5.3 L'ELETTROMAGNETISMO NEI CONTINUI MATERIALI

APP. SPEC. CAP. 2

- 2.A SPOSTAMENTI RIGIDI E CINEMATICA CLASSICA
- 2.B PROCEDURE DI SINCRONIZZAZIONE
- 2.C COMPLEMENTI SULLE TRASFORMAZIONI DI LORENTZ
- 2.D LE FORMULE DI TRASFORMAZIONE DI LORENTZ PARALLELA
- 2.E INDUZIONE DI LEGGI DI CONSERVAZIONE IN MECCANICA RELATIVISTICA SPECIALE (SECONDO LEWIS E TOLMAN)
- 2.F ANCORA SULLA RELAZIONE TRA MASSA DI QUIETE E MASSA DI MOTO
- 2.G DESCRIZIONE LAGRANGIANA E DESCRIZIONE EULERIANA DEL MOTO DI UN CONTINUO
- 2.H ELEMENTI DI DINAMICA CLASSICA E RELATIVISTICA DEI MEZZI MATERIALI CONTINUI
- 2.I I PARADOSSI DELLE VELOCITÀ SUPERLUMINALI IN RELATIVITÀ SPECIALE

PARTE SECONDA

STRUMENTI MATEMATICI DI BASE

CAP. 3 STRUMENTI MATEMATICI I

3.1 ELEMENTI DI ALGEBRA MULTILINEARE E TENSORIALE IN SPAZI PSEUDOEUCLEIDEI

3.1.1 FORME κ -LINEARI

3.1.2 κ -TENSORI

3.1.3 FORME SIMMETRICHE E ANTISIMMETRICHE

3.2 SVILUPPI E APPLICAZIONI DELL'ALGEBRA TENSORIALE

3.2.1 SULLE TRASFORMAZIONI ORTOGONALI E PSEUDORTOGONALI

3.2.2 PSEUDOTENSORE DI LEVI-CIVITA

3.2.3 ALCUNE APPLICAZIONI ALLA FISICA

3.2.4 B -ORTOGONALITÀ E SOTTOSPAZI

3.2.5 GRAMIANI E PROIETTORI ORTOGONALI

3.2.6 OPERATORI LINEARI SIMMETRICI E HERMITIANI, AUTOPROBLEMI

3.2.7 INVARIANTI SCALARI

3.3 ANALISI TENSORIALE LOCALE IN SPAZI PSEUDOEUCLEIDEI E IN LORO VARIETÀ IMMERSE I

3.3.1 CAMPI TENSORIALI, BASI CARTESIANE E BASI LOCALI

3.3.2 DERIVAZIONE DI UN CAMPO TENSORIALE IN UNA BASE LOCALE

3.4 ANALISI TENSORIALE LOCALE IN SPAZI PSEUDOEUCLEIDEI E IN LORO VARIETÀ IMMERSE II

3.4.1 DERIVAZIONE DI CAMPI TENSORIALI INTERNA A VARIETÀ IMMERSE

3.4.2 IL TENSORE DI RIEMANN IN VARIETÀ IMMERSE

3.5 TEORIA DELLA CURVATURA PER VARIETÀ EMBEDDED IN UNO SPAZIO EUCLIDEO m -DIMENSIONALE

3.5.1 INTRODUZIONE

3.5.2 CASO DELLE CURVE, $n = 1$

3.5.3 CASO DELLE (IPER)SUPERFICI, $2 \leq n = m-1$

3.5.4 CASO DELLE VARIETÀ, $2 \leq n < m-1$

APP. SPEC. CAP. 3

3.A NOTA SUL TENSORE FONDAMENTALE IN RELATIVITÀ SPECIALE

3.B ALCUNI ASPETTI DELLA GEOMETRIA DIFFERENZIALE DI UNA SUPERFICIE IMMERSA IN \mathbb{R}^3

CAP 4 STRUMENTI MATEMATICI II

4.1 VARIETÀ TOPOLOGICHE E DIFFERENZIABILI

- 4.1.1 NOZIONI DI BASE I
- 4.1.2 NOZIONI DI BASE II
- 4.1.3 ESEMPI E COMMENTI

4.2 CALCOLO DIFFERENZIALE SU VARIETÀ

- 4.2.1 APPLICAZIONI DI UNA VARIETÀ IN UNA VARIETÀ
- 4.2.2 CALCOLO DIFFERENZIALE DEL 1° ORDINE SU O TRA VARIETÀ
- 4.2.3 APPROCCI ALTERNATIVI AL CALCOLO DIFFERENZIALE SU O TRA VARIETÀ

4.3 ALGEBRA E ANALISI DIFFERENZIALE DI CAMPI $\langle a, b \rangle$ -TENSORIALI SU VARIETÀ ASTRATTE

- 4.3.1 ALGEBRA DEI $\langle a, b \rangle$ -TENSORI
- 4.3.2 CAMPI TENSORIALI IN VARIETÀ A CONNESSIONE AFFINE
- 4.3.3 CAMPI TENSORIALI RELATIVI E CAMPI PSEUDOTENSORIALI

4.4 ALGEBRE DELLE FORME SIMMETRICHE E ANTISIMMETRICHE

- 4.4.1 INTRODUZIONE
- 4.4.2 FORME (COMPLETAMENTE) SIMMETRICHE ED ANTISIMMETRICHE
- 4.4.3 ALGEBRE DI GRASSMANN

4.5 INTRODUZIONE AL CALCOLO DIFFERENZIALE ESTERNO

- 4.5.1 DALLE FORME DIFFERENZIALI ESTERNE AL LEMMA DI POINCARÉ INVERSO
- 4.5.2 TEOREMI DEL TIPO "FROBENIUS"
- 4.5.3 IL TEOREMA DI GAUSS-BONNET

APP. SPEC. CAP. 4

- 4.A ESTENSIONE DELLE FORMULE DI FRENET-SERRET A UNA VARIETÀ PSEUDORIEMANNIANA
- 4.B COMPLEMENTI: OLTRE IL PULL-BACK
- 4.C LE GEOMETRIE A CONNESSIONE AFFINE CON TENSORE FONDAMENTALE. CENNI ALLE TEORIE FISICHE UNITARIE

CAP 5 STRUMENTI MATEMATICI III

5.1 INTEGRAZIONE

- 5.1.1 INTEGRAZIONE SU SPAZI CARICHI
- 5.1.2 J-MISURA DI SOTTOINSIEMI NOTEVOLI DI \mathbf{R}^n
- 5.1.3 INTEGRAZIONE SU SOTTOVARIETÀ DI \mathbf{R}^n
- 5.1.4 ESTENSIONI AGLI L-INTEGRALI

5.2 IL RAPPORTO DIFFERENZIAZIONE-INTEGRAZIONE I

- 5.2.1 IL TEOREMA DI GAUSS-OSTROGRADSKIJ
- 5.2.2 ESEMPI ED APPLICAZIONI I

- 5.2.3 ESEMPI ED APPLICAZIONI II (TEORIA DEL POTENZIALE NEWTONIANO)
- 5.2.4 ESEMPI ED APPLICAZIONI III (TEORIA DEI POTENZIALI ELETTROMAGNETICI)

5.3 IL RAPPORTO DIFFERENZIAZIONE-INTEGRAZIONE II:

- 5.3.1 EQUAZIONI DIFFERENZIALI E LORO SISTEMI: GENERALITÀ
- 5.3.2 IL PROBLEMA DI CAUCHY NORMALE NELLA TEORIA DEI SDP
- 5.3.3 IL PROBLEMA DI CAUCHY GENERALIZZATO NELLA TEORIA DEI SDP

5.4 L'EQUAZIONE DIFFERENZIALPARZIALE DEL 1° ORDINE:

- 5.4.1 IL CASO PROTOTIPO CON 2 VARIABILI INDIPENDENTI
- 5.4.2 IL CASO GENERALE CON $n \geq 2$ VARIABILI INDIPENDENTI
- 5.4.3 L'INTEGRALE COMPLETO

APP. SPEC. CAP. 5

- 5.A I PROBLEMI DI DIRICHLET E DI NEUMANN, INTERNI ED ESTERNI

CAP. 6 STRUMENTI MATEMATICI IV

6.1 ELEMENTI DI CALCOLO DELLE VARIAZIONI I

- 6.1.1 GENERALITÀ
- 6.1.2 LE EQUAZIONI DI EULERO-LAGRANGE NEL CASO UNIDIM
- 6.1.3 PROBLEMI VARIAZIONALI UNIDIM CONDIZIONATI

6.2 APPLICAZIONI DEL CDV UNIDIM

- 6.2.1 RASSEGNA DI PROBLEMI CLASSICI
- 6.2.2 GEODETICHE DI UNA VARIETÀ RIEMANNIANA

6.3 DINAMICA ANALITICA CLASSICA I

- 6.3.1 DALLA DINAMICA DEI SISTEMI DI PUNTI MATERIALI ALLA DINAMICA LAGRANGIANA
- 6.3.2 LE BASI CONCETTUALI DELLA DINAMICA HAMILTONIANA
- 6.3.3 TRASFORMAZIONI CANONICHE I
- 6.3.4 TRASFORMAZIONI CANONICHE II (PARENTESI DI LAGRANGE E DI POISSON)

6.4 DINAMICA ANALITICA CLASSICA II

- 6.4.1 L'EQUAZIONE DI HAMILTON-JACOBI
- 6.4.2 APPLICAZIONI DELL'EQUAZIONE DI HAMILTON-JACOBI
- 6.4.3 ELEMENTI DI DINAMICA CELESTE I
- 6.4.4 ELEMENTI DI DINAMICA CELESTE II

APP. SPEC. CAP. 6

- 6.A SULLA RAPPRESENTAZIONE POLARE DELLE CONICHE
- 6.B FORMULAZIONE LAGRANGIANA/HAMILTONIANA DELLA DINAMICA RELATIVISTICA SPECIALE DI UN PUNTO MATERIALE CARICO

CAP. 7 STRUMENTI MATEMATICI V

- 7.1 ELEMENTI DI CALCOLO DELLE VARIAZIONI II
 - 7.1.1 FONDAMENTI DEL CDV MULTIDIMENSIONALE
 - 7.1.2 ESEMPI DI APPLICAZIONE DEL CDV MULTIDIMENSIONALE I
 - 7.1.3 ESEMPI DI APPLICAZIONE DEL CDV MULTIDIMENSIONALE II
 - 7.2 ELEMENTI DI CALCOLO DELLE VARIAZIONI III
 - 7.2.1 COMPLEMENTI DI CDV
 - 7.2.2 SUL TEOREMA DI NOETHER
 - 7.3 PROBLEMI VARIAZIONALI OMOGENEI DEL 1° ORDINE
 - 7.3.1 PROBLEMI OMOGENEI UNIDIMENSIONALI
 - 7.3.2 PROBLEMI OMOGENEI MULTIDIMENSIONALI (CENNI)
- APP. SPEC. CAP. 7
- 7.A DISCONTINUITÀ DI SOLUZIONI DI SDP QUASI-LINEARI

PARTE TERZA

COMPLEMENTI DI GEOMETRIA DIFFERENZIALE E DI RELATIVITÀ SPECIALE, RELATIVITÀ GENERALE

CAP. 8 COMPLEMENTI DI GEOMETRIA DIFFERENZIALE

- 8.1 COMPLEMENTI DI GEOMETRIA DIFFERENZIALE LOCALE I
 - 8.1.1 n -SUPERFICI IMMERSE IN UNO SPAZIO EUCLIDEO: UNA RIVISITAZIONE ALTERNATIVA
 - 8.1.2 PARALLELISMO GEODETICO E COORDINATE SEMIGEODETICHE
 - 8.1.3 SUL TRASPORTO PARALLELO (DI UN VETTORE LUNGO UNA CURVA SU UNA SUPERFICIE)
 - 8.1.4 2-SUPERFICI IMMERSE IN UNO SPAZIO MINKOWSKIANO 3-DIM
- 8.2 COMPLEMENTI DI GEOMETRIA DIFFERENZIALE LOCALE II
 - 8.2.1 GEOMETRIA DELLE VARIETÀ ELEMENTARI PSEUDORIEMANNIANE
 - 8.2.2 GEOMETRIA DELLE VARIETÀ ELEMENTARI A CONNESSIONE AFFINE
- 8.3 COMPLEMENTI DI GEOMETRIA DIFFERENZIALE LOCALE III
 - 8.3.1 ALGEBRA DEI CAMPI TENSORIALI SU VARIETÀ
 - 8.3.2 DERIVAZIONE DI CAMPI TENSORIALI SU VARIETÀ ORDINARIE. DERIVATA DI LIE

- 8.4 ELEMENTI DI TEORIA DELL'INTEGRAZIONE SU VARIETÀ ELEMENTARI
 - 8.4.1 PROPEDEUTICA
 - 8.4.2 IL TEOREMA DI POINCARÉ-STOKES
 - 8.4.3 ESEMPI ED APPLICAZIONI
- 8.5 INTEGRAZIONE DI FORME DIFFERENZIALI ESTERNE SU VARIETÀ: COMPLEMENTI
 - 8.5.1 κ -FORME E DUALITÀ DI HODGE
 - 8.5.2 CODIFFERENZIAZIONE
 - 8.5.3 IL "PROBLEMA ∂ - δ "
 - 8.5.4 I TEOREMI DI DE RHAM
- APP. SPEC. CAP. 8
 - 8.A SUI MODELLI CANONICI DEL PIANO ELLITTICO E DI QUELLO IPERBOLICO

CAP. 9 COMPLEMENTI DI RELATIVITÀ SPECIALE, RELATIVITÀ GENERALE

- 9.1 NOTA STORICA: DA GAUSS A EINSTEIN
 - 9.1.1 VERSO LE GEOMETRIE NON EUCLIDEE
 - 9.1.2 LA RELATIVITÀ SPECIALE
 - 9.1.3 LA RELATIVITÀ GENERALE
 - 9.1.4 LA RELATIVITÀ E I FATTI OSSERVATIVI
- 9.2 SULLA GEOMETRIA DI UNA VARIETÀ LORENTZIANA
 - 9.2.1 PARTE PRIMA: APPROFONDIMENTI ALGEBRICI
 - 9.2.2 PARTE SECONDA: APPROFONDIMENTI ANALITICI
 - 9.2.3 TETRADI E MATRICI LORENTZIANE, TRASPORTO ALLA FERMI-WALKER
 - 9.2.4 COMPLEMENTI SUL TENSORE DI RIEMANN
- 9.3 LA TEORIA RELATIVISTICA GENERALE
 - 9.3.1 PRELIMINARI
 - 9.3.2 FONDAMENTI DI RELATIVITÀ GENERALE: PARTE PRIMA
 - 9.3.3 FONDAMENTI DI RELATIVITÀ GENERALE: PARTE SECONDA
 - 9.3.4 FONDAMENTI DI RELATIVITÀ GENERALE: PARTE TERZA
 - 9.3.5 FONDAMENTI DI RELATIVITÀ GENERALE: PARTE QUARTA
- 9.4 APPLICAZIONI E COMPLEMENTI I
 - 9.4.1 ANALISI GEODETICA IN RIFERIMENTI NON INERZIALI
 - 9.4.2 IL PARADOSSO DEI GEMELLI E LA SUA SOLUZIONE MEDIANTE LA TRASFORMAZIONE DI MØLLER
 - 9.4.3 SUL TENSORE ENERGETICO TOTALE
 - 9.4.4 MEZZO MATERIALE CONTINUO CON SFORZI "DI CONTIGUITÀ"
- 9.5 APPLICAZIONI E COMPLEMENTI II
 - 9.5.1 LE METRICHE, ESTERNA ED INTERNA, DI SCHWARZSCHILD
 - 9.5.2 CAMPI VETTORIALI DI KILLING, K-SIMMETRIE

- 9.5.3 IL TEOREMA DI BIRKHOFF
- 9.5.4 DINAMICA RELATIVISTICA GENERALE DEL PUNTO MATERIALE E DEL FOTONE

APP. SPEC. CAP. 9

- 9.A SULLA DEDUZIONE EINSTEINIANA DELLE TRASFORMAZIONI DI LORENTZ SPECIALI
- 9.B TRASFORMAZIONI DI LORENTZ PARALLELE DI κ -TENSORI 4-DIMENSIONALI
- 9.C TRASFORMAZIONE DI LORENTZ PARALLELA DEL 2-TENSORE DEGLI SFORZI MECCANICI
- 9.D L-TRASFORMAZIONI E L-BOOSTS
- 9.E MOTI RADIALI E CIRCOLARI DI PUNTI MATERIALI O DI FOTONI IN UNA VARIETÀ DI SCHWARTZSCHILD ESTERNA
- 9.F SULLE VARIETÀ CARATTERISTICHE DELLE EQUAZIONI DI EH
- 9.G NOTA SULLE COORDINATE PSEUDOARMONICHE

APPENDICI GENERALI

APP. GEN. A NOZIONI ELEMENTARI DI LOGICA E TEORIA DEGLI INSIEMI

- A.0 PREMESSA
- A.1 LOGICA E TEORIA DEGLI INSIEMI: NOTAZIONI, MORFOLOGIA, INTERPRETAZIONE INTUITIVA
- A.2 GENERALITÀ SULLA TEORIA DELLA DEDUZIONE
- A.3 GLI ASSIOMI E SCHEMI DI ASSIOMI DELLA TEORIA DEGLI INSIEMI, E ALCUNE DELLE LORO CONSEGUENZE
- A.4 ALCUNE CONSIDERAZIONI SULLA FONDAZIONE E LA FORMALIZZAZIONE DELLE TEORIE MATEMATICHE
- A.5 SEMANTICA DEI SISTEMI FORMALI E TEORIA DEI MODELLI (CENNI)

APP. GEN. B GLOSSARIO RAGIONATO DI TOPOLOGIA

- B.0 PREMESSA
- B.1 DEFINIZIONI FONDAMENTALI
- B.2 CONTINUITÀ, CONNESSIONE, COMPATTEZZA, CICLI, GRUPPO FONDAMENTALE, ECC.

APP. GEN. C STRUTTURE DI MISURA

- C.0 PREMESSA
- C.1 MISURA DI LEBESGUE SUL QUADRATO UNITARIO
- C.2 STRUTTURE DI PRE-MISURA
- C.3 L-MISURE E J-MISURE ASTRATTE

APP. GEN. D INTRODUZIONE ALLA SCIENZA COMPUTAZIONALE

- D.1 RISOLUZIONE APPROSSIMATA DI EQUAZIONI
- D.2 RISOLUZIONE APPROSSIMATA DI EQUAZIONI: ESEMPI ED APPLICAZIONI
- D.3 GENERALITÀ SULL'ANALISI FUNZIONALE-NUMERICA
- D.4 ESEMPI DI METODI COSTRUTTIVI PER LA SOLUZIONE DI EQUAZIONI ASTRATTE
- D.5 I SISTEMI DINAMICI E IL CAOS DETERMINISTICO (CENNI)

APP. GEN. E BREVE STORIA RAGIONATA DEI FONDAMENTI DELLA TERMODINAMICA CLASSICA

- E.1 IL 1° PRINCIPIO
- E.2 LA TESI DI CARNOT E LA TEMPERATURA ASSOLUTA
- E.3 IL 2° PRINCIPIO, L'ENTROPIA E I POTENZIALI TERMODINAMICI
- E.4 CONSIDERAZIONI CONCLUSIVE (DAL PUNTO DI VISTA FISICO-MATEMATICO)

APP. GEN. F ELEMENTI DI TEORIA COSMOLOGICA MACROSCOPICA

- F.1 CENNI STORICI
- F.2 I MODELLI COSMOLOGICI DI EINSTEIN E DI DE SITTER
- F.3 IL MODELLO COSMOLOGICO STANDARD
- F.4 APPLICAZIONI DEL MODELLO COSMOLOGICO STANDARD: FRIEDMANN E LEMAITRE

BIBLIOGRAFIA GENERALE

GLOSSARI

INDICE DEI NOMI

INDICE GENERALE

MATHEMATICAL FOUNDATIONS OF
MACROSCOPIC PHYSICS
(A GEOMETRICAL ROUTE)

C. LO SURDO

CONTENTS

- 0.0 INTRODUCTION
 - 0.0.1 GENERAL CONSIDERATIONS, PLAN OF THE WORK
 - 0.0.2 CONTENTS AND COMMENTS
 - 0.0.3 “DIRECTIONS FOR USE”
- 0.1 SOME CONSIDERATIONS ABOUT THE PHYSICO-MATHEMATICAL THEORIES
 - 0.1.1 PRELIMINARIES
 - 0.1.2 AXIOMATIC-DEDUCTIVE VS. INDUCTIVE REASONING
 - 0.1.3 SIGNIFICANT VS. NON-SIGNIFICANT MATHEMATICS
 - 0.1.4 MATHEMATIZABLE VS. NON-MATHEMATIZABLE SCIENCES

PART ONE

EUCLIDEAN AND PSEUDOEUCLEIDEAN GEOMETRY

CHPT. 1 EUCLIDEAN GEOMETRY

- 1.1 GEOMETRY AS A PROTOTYPE OF THE PHYSICO-MATHEMATICAL THEORIES
 - 1.1.1 OUTLINE AND HISTORICAL FRAMING
 - 1.1.2 SYNTHETIC VS. ANALYTICAL EUCLIDEAN GEOMETRY
- 1.2 A METRIC FORMALIZATION OF EUCLIDEAN GEOMETRY
 - 1.2.1 INTRODUCTION
 - 1.2.2 THE FIRST NINE AXIOMS AND THE “SPECIAL” GEOMETRY OF THE STRAIGHT-LINE
 - 1.2.3 THE SUBSEQUENT EIGHT AXIOMS AND THE DYADIC PLANE/SPACE GEOMETRIES
 - 1.2.4 THE LAST TWO AXIOMS: ABSOLUTE VS. EUCLIDEAN GEOMETRY
- 1.3 ISOMORPHISM BETWEEN H_n AND THE REAL CARTESIAN SPACE \mathbf{R}^n
 - 1.3.1 \mathbf{R}^n AS A NORMAL MODEL OF H_n : PART 1
 - 1.3.2 \mathbf{R}^n AS A NORMAL MODEL OF H_n : PART 2
- 1.4 EXTENSION AND ORIENTATION, CONCLUDING REMARKS
 - 1.4.1 EXTENSION
 - 1.4.2 ORIENTATION
 - 1.4.2 CONCLUDING REMARKS

SPEC. APP.S TO CHPT. 1

- 1.A ORDER RELATIONS
- 1.B THE SPECIAL FUNCTIONS Cos AND Sin

CHPT. 2 PSEUDOEUCCLIDEAN GEOMETRY

- 2.1 PRELIMINARIES: THE EUCLID-NEWTON SPACE-TIME
 - 2.1.1 ELEMENTS OF CLASSICAL KINEMATICS
 - 2.1.2 ELEMENTS OF CLASSICAL DYNAMICS
- 2.2 INTRODUCTION TO SPECIAL-RELATIVISTIC KINEMATICS AND DYNAMICS
 - 2.2.1 SPECIAL-RELATIVISTIC KINEMATICS, SPECIAL LORENTZ'S TRANSFORMATIONS
 - 2.2.2 RUDIMENTS OF SPECIAL-RELATIVISTIC DYNAMICS
- 2.3 INTRODUCTION TO PSEUDOEUCCLIDEAN SPACES
 - 2.3.1 PSEUDOEUCCLIDEAN SPACES (BASIC CONCEPTS)
 - 2.3.2 MINKOWSKI'S SPACE (BASIC CONCEPTS)
- 2.4 ALGEBRA AND GEOMETRY OF THE MINKOWSKI SPACE
 - 2.4.1 PART 1
 - 2.4.2 PART 2
- 2.5 INTRODUCTION TO CLASSICAL ELECTROMAGNETISM
 - 2.5.1 THE MAXWELL-LORENTZ EQUATIONS
 - 2.5.2 ELECTROMAGNETIC THEORY IN TENSOR LANGUAGE
 - 2.5.3 ELECTROMAGNETIC THEORY IN MATERIAL CONTINUOUS MEDIA

SPEC. APP.S TO CHPT. 2

- 2.A RIGID DISPLACEMENTS AND CLASSICAL KINEMATICS
- 2.B SYNCHRONIZATION PROCEDURES
- 2.C COMPLEMENTS ON LORENTZ'S TRANSFORMATIONS
- 2.D PARALLEL LORENTZ TRANSFORMATIONS
- 2.E INDUCTION OF THE CONSERVATION LAWS IN SPECIAL RELATIVISTIC MECHANICS
- 2.F MORE ON THE RELATIONSHIP BETWEEN "REST" AND "MOTION" MASS
- 2.G LAGRANGE VS. EULER DESCRIPTION OF THE MOTION OF A CONTINUOUS MEDIUM
- 2.H CLASSICAL AND RELATIVISTIC DYNAMICS OF A CONTINUOUS MATERIAL MEDIUM
- 2.I THE PARADOXES ARISING FROM SUPERLUMINAL VELOCITIES IN SPECIAL RELATIVITY

PART TWO

BASIC MATHEMATICAL TOOLS

CHPT. 3 MATHEMATICAL TOOLS I

3.1 MULTILINEAR AND TENSOR ALGEBRA IN PSEUDOEUCCLIDEAN SPACES

3.1.1 κ -LINEAR FORMS

3.1.2 κ -TENSORS

3.1.3 SYMMETRIC AND ANTISYMMETRIC FORMS

3.2 TENSOR ALGEBRA AND APPLICATIONS

3.2.1 ORTHOGONAL AND PSEUDORTHOGONAL TRANSFORMATIONS

3.2.2 LEVI-CIVITA'S PSEUDOTENSOR

3.2.3 SOME PHYSICAL APPLICATIONS

3.2.4 B -ORTHOGONALITY AND SUBSPACES

3.2.5 GRAMIANS AND ORTHOGONAL PROJECTORS

3.2.6 SYMMETRIC AND HERMITIAN OPERATORS, EIGENPROBLEMS

3.2.7 SCALAR INVARIANTS

3.3 TENSOR ANALYSIS IN PSEUDOEUCCLIDEAN SPACES AND IMMERSED MANIFOLDS I

3.3.1 TENSOR FIELDS, CARTESIAN AND LOCAL BASES

3.3.2 DERIVATION OF A TENSOR FIELD IN A LOCAL BASE

3.4 TENSOR ANALYSIS IN PSEUDOEUCCLIDEAN SPACES AND IMMERSED MANIFOLDS II

3.4.1 "INNER" DERIVATION OF A TENSOR FIELD (IN AN IMMERSSED MANIFOLD)

3.4.2 RIEMANN TENSOR (IN AN IMMERSSED MANIFOLD)

3.5 CURVATURE THEORY IN A MANIFOLD EMBEDDED IN A m -DIM EUCLIDEAN SPACE

3.5.1 INTRODUCTION

3.5.2 CASE OF A CURVE, $n = 1$

3.5.3 CASE OF A (HYPER)-SURFACE, $2 \leq n = m-1$

3.5.3 CASE OF A MANIFOLD, $2 \leq n < m-1$

SPEC. APP.S TO CHPT. 3

3.A A NOTE ON THE FUNDAMENTAL TENSOR IN SPECIAL RELATIVITY

3.B SOME ASPECTS OF THE DIFFERENTIAL GEOMETRY OF A SURFACE IMMERSSED IN \mathbb{R}^3

CHPT. 4 MATHEMATICAL TOOLS II

- 4.1 TOPOLOGICAL AND DIFFERENTIABLE MANIFOLDS
 - 4.1.1 BASIC CONCEPTS I
 - 4.1.2 BASIC CONCEPTS II
 - 4.1.2 EXAMPLES AND COMMENTS
- 4.2 DIFFERENTIAL CALCULUS ON MANIFOLDS
 - 4.2.1 APPLICATIONS OF A MANIFOLD INTO A MANIFOLD
 - 4.2.2 1ST-ORDER DIFFERENTIAL CALCULUS ON, OR BETWEEN, MANIFOLDS
 - 4.2.3 ALTERNATIVE APPROACHES TO DIFFERENTIAL CALCULUS ON, OR BETWEEN, MANIFOLDS
- 4.3 ALGEBRAIC AND DIFFERENTIAL ANALYSIS OF $\langle a, b \rangle$ -TENSOR FIELDS ON MANIFOLDS
 - 4.3.1 $\langle a, b \rangle$ -TENSOR ALGEBRA
 - 4.3.2 MANIFOLDS WITH AN AFFINE CONNECTION AND TENSOR FIELDS ON THEM
 - 4.3.3 RELATIVE TENSOR FIELDS AND PSEUDOTENSOR FIELDS
- 4.4 ALGEBRAS OF THE SYMMETRIC/ANTISYMMETRIC FORMS
 - 4.4.1 PRELIMINARIES
 - 4.4.2 (COMPLETELY) SYMMETRIC/ANTISYMMETRIC FORMS
 - 4.4.3 GRASSMANN'S ALGEBRAS
- 4.5 INTRODUCTION TO EXTERIOR DIFFERENTIAL CALCULUS
 - 4.5.1 FROM THE EXTERIOR DIFFERENTIAL FORMS TO THE INVERSE POINCARÉ'S LEMMA
 - 4.5.2 THEOREMS OF THE FROBENIUS TYPE
 - 4.5.3 THE GAUSS-BONNET THEOREM
- SPEC. APP.S TO CHPT. 4
 - 4.A EXTENSION OF FRENET-SERRET'S FORMULAS TO A PSEUDORIEMANNIAN MANIFOLD
 - 4.B COMPLEMENTS: BEYOND THE PULL-BACK
 - 4.C AFFINELY CONNECTED GEOMETRIES WITH A FUNDAMENTAL TENSOR. UNITARY PHYSICAL THEORIES (SHORT ACCOUNT)

CHPT. 5 MATHEMATICAL TOOLS III

- 5.1 INTEGRATION
 - 5.1.1 INTEGRATION IN "CHARGED" SPACES
 - 5.1.2 J-MEASURE OF TYPICAL SUBSETS OF \mathbf{R}^n
 - 5.1.3 INTEGRATION ON SUBMANIFOLDS OF \mathbf{R}^n
 - 5.1.4 EXTENSIONS TO L-INTEGRALS
- 5.2 THE RELATIONSHIP BETWEEN DIFFERENTIATION AND INTEGRATION I
 - 5.2.1 THE GAUSS-OSTROGRADSKIJ THEOREM
 - 5.2.2 EXAMPLES AND APPLICATIONS I
 - 5.2.3 EXAMPLES AND APPLICATIONS II (NEWTONIAN POTENTIAL THEORY)
 - 5.2.4 EXAMPLES AND APPLICATIONS III (ELECTROMAGNETIC POTENTIAL THEORY)

5.3 THE RELATIONSHIP BETWEEN DIFFERENTIATION AND INTEGRATION II

- 5.3.1 DIFFERENTIAL EQUATIONS AND SYSTEMS: AN OUTLINE
- 5.3.2 THE NORMAL CAUCHY PROBLEM IN PDS THEORY
- 5.3.3 THE GENERALIZED CAUCHY PROBLEM IN PDS THEORY

5.4 THE 1ST-ORDER PARTIAL DIFFERENTIAL EQUATION

- 5.4.1 SIMPLEST CASE WITH TWO INDEPENDENT VARIABLES
- 5.4.2 GENERAL CASE WITH $n \geq 2$ INDEPENDENT VARIABLES
- 5.4.3 THE COMPLETE INTEGRAL

SPEC. APP.S TO CHPT. 5

5.A INTERIOR/EXTERIOR DIRICHLET/NEUMANN PROBLEMS

CHPT. 6 MATHEMATICAL TOOLS IV

6.1 CALCULUS OF VARIATIONS I

- 6.1.1 PRELIMINARIES
- 6.1.2 EULER-LAGRANGE EQUATIONS IN THE ONE-DIMENSIONAL CASE
- 6.1.3 CONDITIONED VARIATIONAL ONE-DIMENSIONAL PROBLEMS

6.2 APPLICATIONS OF THE ONE-DIMENSIONAL COV

- 6.2.1 SOME CLASSICAL PROBLEMS
- 6.2.2 GEODETICS OF A RIEMANN MANIFOLD

6.3 CLASSICAL ANALYTICAL DYNAMICS I

- 6.3.1 FROM THE DYNAMICS OF A SYSTEM OF MATERIAL POINTS TO LAGRANGE'S FORMALISM
- 6.3.2 FUNDAMENTALS OF HAMILTON'S DYNAMICS
- 6.3.3 CANONICAL TRANSFORMATIONS I
- 6.3.4 CANONICAL TRANSFORMATIONS II (POISSON AND LAGRANGE BRACKETS)

6.4 CLASSICAL ANALYTICAL DYNAMICS II

- 6.4.1 THE HAMILTON-JACOBI EQUATION
- 6.4.2 APPLICATIONS OF THE HAMILTON-JACOBI EQUATION
- 6.4.3 ELEMENTS OF CELESTIAL DYNAMICS I
- 6.4.4 ELEMENTS OF CELESTIAL DYNAMICS II

SPEC. APP.S TO CHPT. 6

- 6.A POLAR REPRESENTATION OF THE CONIC CURVES
- 6.B THE LAGRANGE/HAMILTON FORMULATION OF THE SPECIAL-RELATIVISTIC DYNAMICS FOR A MATERIAL, CHARGED POINT

CHPT. 7 MATHEMATICAL TOOLS V

7.1 CALCULUS OF VARIATIONS II

- 7.1.1 FUNDAMENTALS OF THE TO MULTI-DIM COV
- 7.1.2 SOME APPLICATIONS OF MULTI-DIM COV I
- 7.1.3 SOME APPLICATIONS OF MULTI-DIM COV II

7.2 CALCULUS OF VARIATIONS III

- 7.2.1 COMPLEMENTS OF COV
- 7.2.2 NOETHER'S THEOREM

7.3 HOMOGENEOUS VARIATIONAL PROBLEMS OF THE 1ST ORDER

- 7.3.1 ONE-DIMENSIONAL CASE
- 7.3.2 HOMOGENEOUS MULTI-DIMENSIONAL PROBLEMS (SHORT ACCOUNT)

SPEC. APP.S TO CHPT. 7

- 7.A DISCONTINUITIES OF SOLUTIONS OF QUASI-LINEAR PDS

PART THREE

COMPLEMENTS OF DIFFERENTIAL GEOMETRY AND SPECIAL RELATIVITY. GENERAL RELATIVITY

CHPT. 8 COMPLEMENTS OF DIFFERENTIAL GEOMETRY

8.1 COMPLEMENTS OF LOCAL DIFFERENTIAL GEOMETRY I

- 8.1.1 IMMERSED n -SURFACES: A DIFFERENT APPROACH TO SOME BASIC CONCEPTS
- 8.1.2 GEODETIC PARALLELISM AND SEMIGEODETIC COORDINATES
- 8.1.3 PARALLEL TRANSPORT OF A VECTOR ALONG A CURVE OF AN IMMERSED SURFACE
- 8.1.4 2-SURFACES IMMERSED IN A PSEUDOEUCLEIDEAN MINKOWSKI 3-DIM SPACE

8.2 COMPLEMENTS OF LOCAL DIFFERENTIAL GEOMETRY II

- 8.2.1 GEOMETRY OF ELEMENTARY PSEUDORIEMANNIAN MANIFOLDS
- 8.2.2 GEOMETRY OF ELEMENTARY MANIFOLDS WITH AN AFFINE CONNECTION

8.3 COMPLEMENTS OF LOCAL DIFFERENTIAL GEOMETRY III

- 8.3.1 ALGEBRA OF TENSOR FIELDS ON MANIFOLDS
- 8.3.2 DERIVATION OF TENSOR FIELDS ON ORDINARY MANIFOLDS. LIE DERIVATIVE

8.4 ELEMENTS OF INTEGRATION THEORY ON ELEMENTARY MANIFOLDS

- 8.4.1 PRELIMINARIES
- 8.4.2 THE THEOREM OF POINCARÉ-STOKES
- 8.4.3 EXAMPLES AND APPLICATIONS

8.5 EXTERIOR DIFFERENTIAL FORMS ON MANIFOLDS: COMPLEMENTS

- 8.5.1 κ -FORMS AND HODGE DUALITY
- 8.5.2 CODIFFERENTIATION
- 8.5.3 THE ∂ - $\bar{\partial}$ PROBLEM
- 8.5.4 DE RHAM'S THEOREMS

SPEC. APP.S TO CHPT. 8

- 8.A CANONICAL MODELS OF THE ELLIPTIC AND HYPERBOLIC PLANES

CHPT. 9 COMPLEMENTS OF SPECIAL RELATIVITY, GENERAL RELATIVITY

9.1 FROM GAUSS TO EINSTEIN: A HISTORICAL NOTE

- 9.1.1 TOWARDS THE NON-EUCLIDEAN GEOMETRIES
- 9.1.2 SPECIAL RELATIVITY
- 9.1.3 GENERAL RELATIVITY
- 9.1.4 RELATIVITY AND OBSERVATIONAL FACTS

9.2 ON THE GEOMETRY OF A LORENTZ MANIFOLD

- 9.2.1 PART ONE: ALGEBRAIC ASPECTS
- 9.2.2 PART TWO: ANALYTICAL ASPECTS

9.3 MATHEMATICAL FOUNDATIONS OF GENERAL RELATIVITY

- 9.3.1 PRELIMINARIES
- 9.3.2 FOUNDATIONS OF GENERAL RELATIVITY I
- 9.3.3 FOUNDATIONS OF GENERAL RELATIVITY II
- 9.3.4 FOUNDATIONS OF GENERAL RELATIVITY III
- 9.3.5 FOUNDATIONS OF GENERAL RELATIVITY IV

9.4 APPLICATIONS AND COMPLEMENTS I

- 9.4.1 GEODETIC ANALYSIS IN NON-INERTIAL FRAMES
- 9.4.2 THE TWIN PARADOX AND ITS SOLUTION ACCORDING TO MØLLER
- 9.4.3 ABOUT THE TOTAL ENERGY TENSOR
- 9.4.4 CONTINUOUS MATERIAL MEDIUM WITH "CONTIGUITY" STRESSES

9.5 APPLICATIONS AND COMPLEMENTS II

- 9.5.1 SCHWARZSCHILD'S METRICS, EXTERIOR AND INTERIOR
- 9.5.2 KILLING'S VECTORS, K-SYMMETRIES
- 9.5.3 BIRKHOFF'S THEOREM
- 9.5.4 ON THE RELATIVISTIC DYNAMICS OF A MATERIAL POINT (PRECESSION OF MERCURY, BENDING OF LIGHT, ETC.)

SPEC. APP.S TO CHPT 9

- 9.A EINSTEIN'S DEDUCTION OF THE SPECIAL LORENTZ TRANSFORMATIONS
- 9.B PARALLEL LORENTZ'S TRANSFORMATIONS FOR 4-DIM κ -TENSORS
- 9.C PARALLEL LORENTZ'S TRANSFORMATION OF THE ENERGY 2-TENSOR
- 9.D L-TRANSFORMATIONS AND L-BOOSTS
- 9.E ON THE GEODETICS ON SCHWARZSCHILD'S MANIFOLD: RADIAL AND CIRCULAR MOTIONS OF MATERIAL POINTS OR PHOTONS
- 9.F A NOTE ON THE GRAVITATIONAL WAVES
- 9.G PSEUDOHARMONIC COORDINATES

GENERAL APPENDICES

GEN APP. A ELEMENTS OF LOGIC AND SET THEORY

- A.0 PRELIMINARIES
- A.1 LOGIC AND SET THEORY: NOTATIONS, MORPHOLOGY, INTUITIVE INTERPRETATION
- A.2 ELEMENTS OF DEDUCTION THEORY
- A.3 AXIOMS AND AXIOM SCHEMATA OF THE SET THEORY, AND SOME OF THEIR CONSEQUENCES
- A.4 FOUNDATION AND FORMALIZATION OF MATHEMATICAL THEORIES
- A.5 SEMANTICS OF THE FORMAL SYSTEMS, AND MODEL THEORY (SHORT ACCOUNT)

GEN APP. B TOPOLOGY: A REASONED GLOSSARY

- B.0 PRELIMINARIES
- B.1 BASIC DEFINITIONS
- B.2 CONTINUITY, CONNECTEDNESS, COMPACTNESS, LOOPS, ETC

GEN. APP. C MEASURE STRUCTURES

- C.0 PRELIMINARIES
- C.1 INTRODUCTION
- C.2 PRE-MEASURE STRUCTURES
- C.3 ABSTRACT L-MEASURE AND J-MEASURE

GEN. APP. D INTRODUCTION TO COMPUTATIONAL SCIENCE

- D.1 APPROXIMATE SOLUTION OF EQUATIONS
- D.2 EXAMPLES AND APPLICATIONS
- D.3 ON NUMERICAL-FUNCTIONAL ANALYSIS
- D.4 CONSTRUCTIVE METHODS TO SOLVE ABSTRACT EQUATIONS (EXAMPLES)
- D.5 DYNAMICAL SYSTEMS AND DETERMINISTIC CHAOS (SHORT ACCOUNT)

GEN. APP. E A SHORT HISTORY OF CLASSICAL THERMODYNAMICS

- E.1 FIRST PRINCIPLE
- E.2 CARNOT'S THESIS AND ABSOLUTE TEMPERATURE
- E.3 SECOND PRINCIPLE, ENTROPY AND THERMODYNAMIC POTENTIALS
- E.4 CONCLUDING REMARKS (FROM THE PHYSICO-MATHEMATICAL STANDPOINT)

GEN. APP. F ELEMENTS OF MACROSCOPIC COSMOLOGY

- F.1 HISTORICAL NOTE
- F.2 EINSTEIN'S AND DE SITTER'S COSMOLOGICAL MODELS
- F.3 THE STANDARD COSMOLOGICAL MODEL
- F.4 APPLICATIONS OF THE STANDARD COSMOLOGICAL MODEL: FRIEDMANN AND LEMAITRE

GENERAL BIBLIOGRAPHY

SUBJECT INDICES (1÷4)

AUTHOR INDEX

CONTENTS

Orientativamente, le teorie fisiche “macroscopiche” si occupano di fenomeni con tempi di evoluzione tipici molto più lunghi di 10^{-14} secondi (che a loro volta corrispondono alle energie tipiche in gioco nelle reazioni chimiche). Esse abbracciano la fisica cosiddetta “classica” e quella relativistica, sia speciale che generale. La presente monografia, che ne illustra i fondamenti matematici, è divisa in quattro parti. La prima parte è dedicata alla geometria euclidea – l’unica “teoria fisica” tentativamente sviluppata su base assiomatica prima della rivoluzione scientifica (\approx XV secolo) –, in una versione equivalente a, ma diversa da, quella di Hilbert (1900), e alla geometria pseudoeuclidea di Einstein-Minkowski (1905-08); la seconda parte, alle teorie fisiche suscettibili di una formulazione variazionale che erano mature sul finire del XIX secolo; la terza parte, alla teoria della relatività generale (Einstein-Hilbert, 1915-16) e a nozioni complementari. Infine la quarta parte del libro consiste nelle sei Appendici Generali, che possono leggersi come saggi ausiliari di orientamento tecnico sugli argomenti di cui ai loro titoli.

*«A quelli che non conoscono la matematica è difficile farsi un'idea precisa della
bellezza, la profonda bellezza, della natura.»*

R. Feynman (1967)

«Non esiste matematica senza lacrime.»

(folclore)

0.0)

PRESENTAZIONE

0.0.1) CONSIDERAZIONI GENERALI E PIANO DI LAVORO

La fisica matematica ¹ che oggi diciamo “classica” ², o più precisamente “macroscopica” ³, nasce verso la metà del Seicento, e diventa una teoria asintotica – per costante di Planck tendente a zero – con la formalizzazione della meccanica quantistica, portata a sostanziale compimento nel sesto quinquennio dello scorso secolo ⁴. Bisogna tuttavia osservare che una fisica matematica sui generis, la geometria euclidea, esisteva nel Seicento da circa due millenni. Se si accetta questo punto di vista (che appare pienamente legittimo a chi scrive), il secolo XVII si deve in effetti vedere come lo spartiacque al di là del quale la tradizionale separazione tra geometria e resto della “filosofia naturale” (la φιλοσοφία φυσική dei Greci) comincia a farsi meno netta. Infatti i più elementari modelli fisico-matematici in senso stretto, i modelli meccanici, coinvolgono

¹ La fisica matematica e la fisica teorica sono spesso esageratamente accostate e tra loro confuse, pur essendo discipline notevolmente diverse. In sostanza, il fisico matematico è un *matematico* che sceglie come oggetto del suo interesse la matematica pertinente alla fisica, mentre il fisico teorico è un *fisico* il cui obiettivo, in mancanza di alternative altrettanto efficaci, è quello di rappresentare il mondo fenomenico entro convenienti cornici matematiche (o come anche si dice gergalmente, quello di “modellare” matematicamente quel mondo), e di sviluppare ed interpretare le conseguenze di quella rappresentazione. Sulle sorgenti ispirazionali della matematica ed i suoi legami con il mondo dei fenomeni, è interessante ricordare qui quanto scriveva J. von Neumann in un noto saggio del 1947 (“The Mathematician”, in “Collected Works”, 1961). «The most vitally characteristic fact about mathematics – dice von Neumann – is its quite peculiar relationship to the natural sciences, or, more generally, to *any* science which interprets experience on a higher more than on a purely descriptive level (...). Mathematical ideas originate in empirics, although the genealogy is sometimes long and obscure. But, once they are so conceived, the subject begins to live a peculiar life of its own and is better, compared to a creative one, governed by almost entirely aesthetic motivations than to anything else and, in particular, to an empirical science (...). However there is a grave danger that the subject will develop along the line of least resistance, that the stream, so far from its source, will separate into a multitude of insignificant tributaries (...). In other words, at a great distance from its empirical sources, or after much “abstract” inbreeding, a mathematical object is in danger of degeneration. At the inception the style is usually classical; when it shows signs of becoming *baroque*, then the danger signal is up (...). Whenever this stage is reached, the only remedy seems a rejuvenating return to the source: the reinjection of more or less directly empirical ideas.». Si deve anche aggiungere che non tutte le opinioni di von Neumann (il quale fu anche un grande fisico-matematico) sulle ascendenze della matematica erano condivise dai matematici a lui contemporanei, e tanto meno lo sarebbero dai matematici del ventesimo secolo.

² Nel linguaggio scientifico-esatto l’attributo “classico” usualmente contraddistingue modelli/teorie considerati a fondamento primo della disciplina in oggetto, e spesso in contrapposizione con modelli/teorie più recenti. Tuttavia il suo uso non è scevro di ambiguità. Ad esempio per un fisico relativista della prima ora era “classico” quanto ancora ignorava la relatività (speciale o generale), mentre per un fisico atomico o nucleare lo è quanto non cade nel dominio della teoria quantistica; e così via. In questo libro conveniamo di nominare come “classica” la fisica matematica macroscopica, cioè quella che può considerarsi (ragionevolmente) completa pur trascurandovi gli effetti quantistici. Essa include la relatività standard, per definizione macroscopica nel senso sopraddetto.

³ L’autore è stato tentato dal sostituire l’attributo “macroscopica” – che figura anche nel titolo del libro – con “deterministica”. Questo avrebbe tuttavia aperto certe delicate questioni epistemologiche che lo hanno dissuaso da una tale scelta.

⁴ L’equazione di Schrödinger è del 1926. Beninteso, qui con “meccanica quantistica” ci riferiamo alla sua versione prerelativistica. Il primo fondamentale passo verso il connubio con la relatività speciale, dovuto a Dirac, è del 1928.

simultaneamente sia la geometria spazio-temporale classica che la fisica, le quali si intrecciano definitivamente nella seconda legge di Newton (*Principia Mathematica*, 1687 ⁵), o “legge fondamentale della dinamica”: la geometria spazio-temporale fornendo i concetti di posizione, velocità e accelerazione, e la fisica quelli di massa e di forza.

Fin dagli albori della ricerca filosofica, partendo dalla esperienza quotidiana l'uomo aveva elaborato una rappresentazione intuitiva dello spazio e del tempo che li raffigurava come “contenitori” assoluti – cioè dotati ciascuno di sue specifiche proprietà, tra le quali la mutua e completa separazione – dei fenomeni che vi osservava. Almeno per quanto riguardava lo spazio, con Euclide (III secolo a.C.; e forse anche prima, con Talete ⁶) aveva poi preso corpo il convincimento che la conoscenza delle sue proprietà potesse essere organizzata in un sistema ipotetico-deduttivo fondato su certe “verità evidenti” e governato da certe “regole naturali”. In effetti quel convincimento si basava su un tentativo non del tutto compiuto, ma come dimostrò una storia millenaria di approfondimenti e miglioramenti, completamente perfettibile. Non è dunque improprio affermare che proprio in quella circostanza la matematica nel senso moderno del termine si affacciò prodigiosamente sulla scena umana. Quanto al tempo, esso venne sottoposto ad una analisi ragionevolmente rigorosa soltanto molto più tardi, a partire dalla nascita del calcolo differenziale, con la (ancora immatura, ma efficace) formalizzazione della cinematica da parte di Newton; cioè, appunto verso la fine del Seicento. Un po' convenzionalmente, potremo denominare il sistema di conoscenze che derivò da queste concettualizzazioni come “geometria spazio-temporale (o se si preferisce, cinematica) di Euclide-Newton”.

Questo quadro di avvio, le cui conseguenze si dimostrarono piuttosto stabili, subì un primo sconvolgimento circa cento anni fa ⁷, quando si fece luce la nozione di uno “spazio-tempo” in cui spazio e tempo si trovano fusi in un “unicum” di natura inusitata; e poco più tardi, una ulteriore evoluzione di prospettive, quando per la prima volta venne proposto che alcune proprietà dello spazio-tempo ed alcuni fenomeni fisici che vi hanno luogo potessero influenzarsi a vicenda. La maggior parte dei meccanismi che governano questa interdipendenza appaiono oggi accettabilmente compresi sulla scala spazio-temporale cosiddetta dei “fenomeni osservabili”. Questa è situata tra circa 19 ordini di grandezza (ma anche alquanto meno nei moderni acceleratori di particelle) al

⁵ Questo è l'anno di pubblicazione dei *Principia*; ma la meditazione di Newton sui fondamenti della meccanica durava, per sua testimonianza, da circa venti anni.

⁶ Probabilmente dobbiamo a Talete (Mileto, VI secolo a. C.) l'introduzione del concetto di dimostrazione matematica, chiave di volta di ogni sistema ipotetico-deduttivo. La scuola pitagorica usò poi sistematicamente tale tipo di dimostrazione come strumento per verificare la possibile validità degli asserti matematici.

⁷ In realtà esso era già stato colpito al cuore nella prima metà dell'Ottocento con la scoperta delle geometrie non-euclidee; ma gli effetti di quei mutamenti concettuali rimasero confinati all'ambito teoretico-matematico fino all'inizio del nuovo secolo.

disopra della scala di Planck ⁸ ($\approx 10^{-35}$ m (metri) o $\approx 10^{-43}$ s (secondi)) – quindi tra $\approx 10^{-16}$ m, o $\approx 10^{-24}$ s – e ovviamente, la dimensione o l'età stimate dell'universo osservabile ($\approx 10^{26}$ m, o 10^{10} anni-luce $\approx 10^{18}$ s). Le proprietà dello spazio-tempo entro queste scale, quindi sul colossale intervallo relativo di almeno $\approx 10^{16+26-24+18} = 10^{42}$, formano l'oggetto della geometria-fisica – o come potremo ormai dire, della fisica matematica – “macroscopica”. Sulla fisica matematica significativamente al disotto di $\approx 10^{-16}$ m, sino alla scala di Planck o addirittura *al disotto* di essa (quindi su fenomeni per il momento, o forse per sempre, inosservabili), si è fatto e si continua a fare un grande lavoro teorico-congetturale, che finora ha prodotto affascinanti ipotesi fisiche e non poca matematica innovativa. Intorno a questa ardua materia comincia anche ad essere disponibile una letteratura divulgativa, ovviamente di tipo metaforico (vedi nota ⁽¹⁵⁾).

La fisica matematica macroscopica *in senso stretto* si situa invece in un dominio relativo assai meno esteso; e precisamente, poiché la sua scala-limite inferiore (diciamo, temporale) è dell'ordine di 10^{-14} s (contro i precedenti $\approx 10^{-24}$ s), significativamente più piccolo di $10^{14+18} = 10^{32}$. La sopraddetta scala-limite inferiore può dirsi “quantistico-chimica”, in quanto alle reazioni chimiche si associano variazioni di energia dell'ordine dell'elettrone-volt ($1 \text{ eV} \approx 10^{-19} \text{ J}$ (joule) ⁹). Ad essa si giunge imponendo che il prodotto dell'azione di Planck ($6,63 \cdot 10^{-34} \text{ Js}$) per la frequenza ν sia $\approx 10^{-19} \text{ J}$, dal che si trae appunto $\nu^{-1} \approx 10^{-14} \text{ s}$. (Il periodo di un orologio atomico al cesio è $\approx 10^{-10} \text{ s}$; e questo consente di giungere a precisioni relative, nella misura del tempo su periodi consistentemente lunghi, dell'ordine di 10^{-13} .) I moderni acceleratori si spingono fino a $\approx 10^{12} \text{ eV} \approx 10^{-7} \text{ J}$ (fino a $\approx 7 \cdot 10^{12} \text{ eV}$ nel Large Hadron Collider (LHC) prossimo ad entrare in funzione alla data presente), e quindi a $\nu^{-1} \approx 10^{-26} \text{ s}$. Questo giustifica quanto abbiamo anticipato menzionando la scala dei fenomeni osservabili. Ovviamente ci si aspetta che una fisica matematica macroscopica in senso stretto come sopra definita debba essere *sostanzialmente incompleta*; e di fatto, innumerevoli questioni di piena pertinenza macroscopica non possono essere affrontate né tanto meno risolte al suo interno, le necessarie informazioni dovendo quindi esservi immesse in modo “artificiale” (nel senso di provenire da una scala fenomenica diversa). Questo tipo di incompletezza non può del resto non affliggere qualunque fisica matematica artificialmente confinata entro un intervallo relativo di scale (i cui estremi sono sempre da pensarsi come asintoticamente lontani) significativamente più piccolo del più grande intervallo relativo oggi ragionevolmente concepibile

⁸ Nel 1899 M. Planck propose che si costruissero unità “naturali” di massa, lunghezza e tempo sulla base delle tre più fondamentali “costanti di natura”, e cioè: di gravitazione G , di velocità della luce c , e di azione (di Planck) h . Il risultato è: m_{pl} (“massa di Planck”) =: $(hc/G)^{1/2} = 5,56 \cdot 10^{-7} \text{ kg}$ (kilogrammi); l_{pl} (“lunghezza di Planck”) =: $(Gh/c^3)^{1/2} = 4,13 \cdot 10^{-35} \text{ m}$; t_{pl} (“tempo di Planck”) =: $(Gh/c^5)^{1/2} = 1,38 \cdot 10^{-43} \text{ s}$. Associata a t_{pl} è l’“energia di Planck” $E_{\text{pl}} = h/t_{\text{pl}} = m_{\text{pl}}c^2 = 4,80 \cdot 10^9 \text{ J}$ (joule), e quindi la “temperatura di Planck” T_{pl} , pari a E_{pl} espressa in unità gradi-kelvin, $T_{\text{pl}} = 3,5 \cdot 10^{32} \text{ K}$ (kelvin).

⁹ Ricordiamo che la carica e dell'elettrone è ca. $1,60 \cdot 10^{-19} \text{ C}$ (coulomb).

($\approx 10^{35+26=61}$; o anche più grande se si scende al *disotto* della scala di Planck); ma come è noto, una fisica matematica che abbracci la *totalità* di questo intervallo, o come oggi si dice una “teoria del tutto”, è ancora lontana da un assetto accettabile allo stato attuale delle conoscenze.

Il modello di spazio-tempo oggi corrente nella fisica matematica macroscopica in senso stretto, o modello dello “spazio-tempo di Riemann-Einstein”¹⁰, che denoteremo brevemente come modello (G) (G come “Generale”), è quello di una varietà 4-dimensionale pseudoriemanniana – possibilmente orientata – con segnatura lorentziana e curvatura legata alla densità di materia-energia nella varietà stessa. Nel limite in cui il rapporto tra l’energia potenziale di gravità per unità di massa φ e il quadrato della velocità della luce c^2 è trascurabile rispetto a 1, $\varphi/c^2 \ll 1$, il modello (G) degenera nel modello (S) (S come “Speciale”) dello “spazio-tempo pseudoeuclideo di Minkowski-Einstein” (una varietà 4-dimensionale anch’essa con segnatura lorentziana ma curvatura nulla); e nel limite in cui $v^2/c^2 \ll 1$, v essendo una velocità tipica, il modello (S) degenera a sua volta nel modello (C) (C come “Classico”) dello “spazio-tempo di Euclide-Newton”, prodotto cartesiano (\times) degli spazi euclidei E^3 e E ¹¹ (ad esempio presi in quest’ordine).

La prima degenerazione (G) \rightarrow (S) pone in evidenza il carattere asintotico (per $\varphi/c^2 \rightarrow 0$) della separazione della cinematica di Einstein-Minkowski dalla fisica in senso stretto¹²; mentre l’analoga degenerazione (S) \rightarrow (C) mostra lo stesso carattere asintotico (per $v^2/c^2 \rightarrow 0$) della separazione dello spazio dal tempo, ovvero della difformità tra la cinematica di Newton-Euclide e quella di Minkowski-Einstein. È allora chiaro che una discussione didattica dei fondamenti della fisica-matematica “macroscopica in senso stretto” – o come diremo in breve nel seguito, semplicemente “macroscopica” – deve partire dalla geometria euclidea (E) (vero e proprio archetipo, per quanto imperfetto nella sua accezione storica, di teoria fisico-matematica), ed ha a suo coronamento quella del modello (G).

Lo svolgimento, opportunamente sintetizzato, di questo programma è grosso modo l’oggetto del presente lavoro. Esso implica preconcoscenze matematiche e fisiche progressivamente più sofisticate muovendo dalla geometria euclidea (E), che in linea di principio si può descrivere quasi in assenza di preconcoscenze di qualsiasi tipo, fino alla relatività generale (G), che comporta invece una laboriosa propedeutica di livello superiore. Il libro è diviso in tre parti: la prima (Capp.

¹⁰ Per brevità, qui e altrove riferiamo *simbolicamente* a Riemann il contributo matematico alla relatività generale. Più o meno lo stesso varrà più sotto per quello di Minkowski alla relatività speciale.

¹¹ Con “spazio euclideo $E^{n \geq 1}$ ” intendiamo qui la potenza cartesiana n -ma della retta reale \mathbf{R} definita a meno della sua origine (cioè considerata come spazio affine), e dotata della distanza $d(x,y) =: [\sum_{i=1}^n (y^i - x^i)^2]^{1/2} = d(y,x)$ per generici $x \equiv x^{1 \leq i \leq n}$ e $y \equiv y^{1 \leq i \leq n}$ di \mathbf{R}^n (ossia della cosiddetta “metrica pitagorica”). Questo è possibile perché la scelta dell’origine non influenza la distanza pitagorica tra x e y . (Del resto nemmeno la scelta dell’orientamento la influenza.)

¹² La fisica è infatti estranea a tale cinematica se non contiamo l’emissione, propagazione, riflessione ecc. di impulsi luminosi, da considerare come oggetti *primitivi*. È precisamente in questo spirito che è concepita la prima parte del nostro libro.

1-2) prevalentemente dedicata all'analisi dei modelli (E), (C) e (S), la seconda (Capp. 3-7) all'illustrazione di strumenti matematici dei quali si è ritenuto opportuno attrezzare il lettore “entro il contesto”, e la terza (Capp. 8-9) a questioni complementari di geometria differenziale e all'analisi del modello (G). È comunque da rimarcare che ogni presunta separazione tra i cosiddetti “contenuti” di una teoria fisico-matematica e i cosiddetti “strumenti” necessari alla loro comprensione/assimilazione è in buona misura artificiale, e che non esistono criteri generali cui ispirarsi, se non quelli della tradizione, per risolvere nel modo migliore i problemi di organizzazione del materiale che nascono da tale situazione (torneremo tra breve su questo punto). Ciò è del resto ben comprensibile, e discende dalla natura *intrinsecamente ibrida* della nostra disciplina.

Sottolineiamo infine che, pur essendo ispirata alla generalità, la nostra esposizione è per lo più strumentale dal punto di vista matematico. L'obiettivo principale è cioè quello di offrire al lettore-studioso una selezione ragionevole della matematica con cui poter presuntivamente “viaggiare attraverso” la fisica teorica macroscopica contemporanea; beninteso, al di là di un ragguardevole insieme di conoscenze considerate come abbastanza elementari e date per note. Non a caso secondo chi la propone, una parte significativa di questa selezione risulta avere carattere geometrico, e più specificamente geometrico-astratto (nel senso della teoria delle varietà differenziabili metriche, o anche più generali).

La vastità del tema ha imposto un numero di scelte: in esso confluiscono infatti così numerosi ed eterogenei capitoli della matematica e della fisica macroscopica da rendere innanzitutto necessaria la ricerca di una conveniente formula espositiva. A complicare questo stato di cose si aggiunge il fatto che, all'incirca a partire dagli ultimi decenni dell'Ottocento, la “matematica della fisica-matematica in generale” (per così esprimerci) ha avuto sviluppi tanto cospicui e in certo senso autonomi, che è ormai norma porsi il problema di quanta sua parte sia obiettivamente indispensabile usare, *a sostanziale parità di contenuti fisici*, in una esposizione didattica del particolare argomento in oggetto.¹³

Tornando al tema del libro, ai fondamenti matematici della fisica macroscopica, è ragionevole, da parte di un autore singolo che non voglia o possa dedicare all'impresa un tempo praticamente illimitato, rinunciare alla linea del trattato autosufficiente – cioè autonomo a meno di preconcoscenze minimali. È dunque necessario presupporre nel lettore una preparazione specifica di livello post-universitario ed un consistente interesse verso gli aspetti fondativi delle scienze esatte in

¹³ Per fare un esempio significativo, si pensi a due classici della meccanica razionale quali sono il trattato in tre volumi di T. Levi-Civita e U. Amaldi (Zanichelli, 1923-26-27) e quello di R. Abraham e J. Marsden (Benjamin, 1^a ed. 1967, 2^a ed. 1978), e alla enorme distanza metodologica che li separa. Benché buona parte di questa difformità abbia una sua ragione di essere – la transizione dal punto di vista locale a quello globale, e tutto quanto ciò comporta –, resta il fatto che la *particolare fisica della quale in esse ci si occupa* sia rimasta più o meno la stessa durante i quarant'anni che intercorrono tra le due opere. Il passaggio “locale” → “globale” ispira l'intero corpo della meccanica analitica su varietà simplettiche, ormai giunta ad un livello di sviluppo molto sofisticato.

generale; e su questa base, selezionare una equilibrata gerarchia di argomenti sui quali concentrarsi. È anche ovvio che le scelte operate in proposito possano non essere condivise, e l'autore ne assume l'esclusiva responsabilità.

Una opzione di grande importanza riguarda lo “stile” espositivo. Ad un estremo della scala, vi è la possibilità di ridurlo a quell'articolazione “Definizione-Teorema-Dimostrazione” (diciamo, DTD) tipica della maggior parte dei testi della fisica matematica contemporanea. Essa comporta la necessità, o l'opportunità, di minimizzare i molti aspetti della materia che esulano dalla sua stretta trattazione tecnica; ma questi non sono necessariamente accessori nell'economia di un discorso che voglia includere approfondimenti abbastanza flessibili e sfaccettati. All'altro estremo, abbiamo invece quel modo di presentare le cose che si potrebbe definire “divulgativo-ma-elevato” – in concreto, e non risparmiandoci una certa dose di sarcasmo, “quasi-senza-formule”¹⁴ -ma-incomprensibile” (ai non-addetti ai lavori) – che da qualche tempo sembra riscuotere consenso di editori e di pubblico. Opere didattiche ispirate a tale scelta sembrano postulare quella separabilità tra contenuti e strumenti cui si accennava più sopra, come se si trattasse di realtà autonome e distinte. Senza dubbio attraente agli occhi dei lettori meno preparati, questa idea si rivela di solito illusoria: perché quei contenuti e quegli strumenti sono in buona parte la stessa cosa, e una loro separazione non può essere che assai limitata e prudente. In conclusione, e come anche suggeriscono l'esperienza e il senso comune, un vero successo dello stile divulgativo-elevato, in campo fisico-matematico o fisico-teorico, presuppone qualità ed attitudini alquanto eccezionali da entrambe le parti della cattedra: un totale dominio della materia e una buona dose di coraggio (o di temerarietà, o di masochismo) da parte del docente, ed un livello di maturità, oltre che di impegno, molto spinto da parte dello studioso-discente. Un ovvio nonsenso nasce inoltre dal possibile simultaneo verificarsi di queste due condizioni: perché l'atto didattico tende allora a trasformarsi in un più o meno gratificante “intrattenimento tra (quasi) pari”, perdendo gran parte della sua connotazione naturale. Se poi – come assai più spesso succede – l'uno e/o l'altro dei presupposti viene a mancare, il preteso discorso divulgativo-elevato, su una piccola o grande porzione di scienza esatta, rivela infallibilmente la sua natura velleitaria e didatticamente improduttiva.^{15, 16}

¹⁴ Non vorremmo che questa espressione fosse presa troppo alla lettera: la scarsa presenza di formule – in un testo fisico-matematico – è infatti qui evocata ironicamente come equivalente di “a rischio di insufficiente chiarezza e rigore logico”.

¹⁵ Esiste (almeno) un terzo approccio alla didattica delle scienze esatte, che potremmo nominare come “metaforico”. Descrivere “per metafore” (letteralmente, “per trasferimenti”) ad un ipotetico discente l'oggetto O significa indicargli altri oggetti O', O'', ... a lui presuntivamente più familiari, ed in rapporto di sufficiente analogia con O sotto gli aspetti di specifico interesse. Si può anzi affermare che il linguaggio metaforico sia l'unico praticabile per parlare di scienze esatte rinunciando alle formule; ma è anche chiaro che una metafora, anche se talvolta non priva di fascino, è una sorta di presa in giro. Insomma, tutto sembra suggerire che per evitare le metafore sia quasi sempre necessario digerire una quantità ragguardevole di matematica. Nessuno conosce con certezza la ragione di questo fatto; ma è così, e non ci sono alternative efficaci. Alla didattica metaforica – nonché all'altrettanto elusiva tecnica espositiva con la quale si sposta

La formula didattica da noi adottata può considerarsi un ambizioso compromesso tra gli estremi più sopra approssimativamente descritti (stile trattatistico “quasi” autosufficiente vs. stile divulgativo-elevato), *fortemente spostato verso il primo di essi*. Avendo provveduto ad una meditata selezione degli argomenti da assumere come irrinunciabili, si è cioè optato per una esposizione un po’ minimalista di una parte delle questioni tecniche. Vale a dire, mentre quasi tutte le nozioni introdotte sono definite con precisione, ed i teoremi importanti e/o impegnativi sono essenzialmente dimostrati (oppure, in alcuni casi, esplicitamente rinviati a tal scopo a trattazioni specializzate), le dimostrazioni a nostro giudizio abbastanza semplici sono spesso rimesse alla buona volontà o alla fiducia del lettore, segnalandogli la circostanza con accorgimenti opportuni. Il risultato è un testo che in certe sue parti sfiora il carattere di un “manuel en problèmes”, e può quindi apparire troppo faticoso ad una lettura non adeguatamente impegnata. Esso è tuttavia leale (o almeno tale si propone di essere), e riesce significativamente più compatto di quanto sarebbe possibile con una esposizione tradizionale, anche nello stile DTD.

Sull’altro piatto della bilancia, il libro offre un’ampia gamma di informazioni complementari: richiami a nozioni/concetti che il lettore può aver dimenticato o ignorare, incursioni in settori conoscitivi limitrofi a quello di principale interesse, suggerimenti di dimostrazioni, spunti critici, notizie storiche (in verità alquanto ridotte, con eccezioni nella Sez. 9.1 e nelle App. Gen. E e F), approcci alternativi alla teoria o frammento di teoria di cui si tratta, commenti di varia natura ed invenzione. In pratica, ciò ha comportato che il testo corra spesso su due livelli paralleli: quello principale, e quello insolitamente cospicuo delle note a piè pagina. Il prezzo pagato per queste scelte è un grado di ordine e sistematicità che non è certo il più desiderabile, e quindi una probabile inadeguatezza a fronte di semplici *consultazioni*. Del resto, il libro non è stato affatto concepito come un manuale: nelle intenzioni dell’autore, esso è destinato ad essere letto/studiato piuttosto che consultato. (I quattro Glossari, che sono abbastanza dettagliati, potranno forse offrire un aiuto al

l’attenzione del lettore dagli *oggetti* delle scienze esatte ai suoi *soggetti* –, si impronta la quasi totalità della letteratura scientifica-esatta di tipo divulgativo. Nonostante le sue inaccettabili insufficienze, la recente crescita di questo tipo di letteratura deve tuttavia considerarsi un evento globalmente positivo: forse essa costituisce un primo passo verso un ormai indispensabile ri-orientamento in senso scientifico della cultura di base, specialmente nel nostro paese.

¹⁶ Un recentissimo e cospicuo esempio di fisica matematica fondamentale praticamente onnicomprensiva ispirata allo stile divulgativo-elevato è quello del per certi versi ammirevole “The Road to Reality” (Jonatan Cape, 2004, tr. ital. Rizzoli, 2005, 1114 pp) di R. Penrose. È tuttavia legittimo dubitare che quanti affronteranno questo libro attratti dalla sua accattivante veste colloquiale, ma essendo addirittura incapaci «di padroneggiare la manipolazione delle frazioni», riusciranno poi ad assimilare almeno una parte, ad avere «una vaga percezione», dell’ardua lezione che esso propone; né l’«inguaribile ottimismo» dell’autore fuga significativamente i nostri dubbi in proposito. È inutile girare intorno ad una *verità didattica* di fondo: come si diceva più sopra, la “via del re” verso le scienze esatte è quella che passa attraverso la matematica. Essa è dunque in dura salita; ma a conti fatti anche la più breve e sicura. Per concludere questo tipo di considerazioni, e in qualche modo riequilibrare i termini del problema “divulgazione sì” vs. “divulgazione no” in campo fisico-matematico e matematico, ci piace ricordare la severa opinione di A. Weil, secondo la quale sarebbe impossibile fare opera di divulgazione matematica. «È completamente inutile – ammoniva Weil – parlare di matematica a chi non è matematico» (mettendo anche in conto che Weil definiva “matematico” «soltanto chi ha scoperto almeno un teorema importante»). Pur sentendoci inclini a qualche compromesso in materia, proviamo molta simpatia per Weil e per la sua coraggiosa intransigenza.

lettore frettoloso.) Infine sei Appendici Generali, in qualche caso nulla di più che glossari ragionati, contengono integrazioni “a portata di mano” al testo principale. Entro certi limiti, esse possono anche essere lette autonomamente, come piccoli saggi o sintesi di orientamento tecnico sugli argomenti specificati nei loro titoli.

Soprattutto nella prima e nella terza parte del libro, il lettore troverà quindi una illustrazione delle geometrie euclidea, pseudoeuclidea, riemanniana e pseudoriemanniana – le ultime due riferite sia a varietà immerse in uno spazio euclideo, e rispettivamente pseudoeuclideo, che a varietà “astratte”, ossia non immerse in alcuno spazio metrico affine; nonché, naturalmente, dello stretto e reciproco legame tra queste matematiche e i modelli formalizzati di Euclide-Newton (geometria e meccanica classica, modello (C)), di Minkowski-Einstein (geometria e fisica relativistica speciale, modello (S)) e infine di Riemann-Einstein (geometria e fisica relativistica generale, modello (G)). Sebbene esuli dallo stretto contesto della teoria della relatività, si sono inoltre fornite informazioni sulla geometria di varietà più generali – prive di metrica ma dotate di una connessione – di quelle pseudoriemanniane. Vi sono infine argomenti, come ad esempio quello dei gruppi topologici o lisci, e specificamente dei gruppi di Lie, dell’algebra di Clifford, delle varietà di Kahler, ecc., che pur non essendo parte *indispensabile* dello strumentario matematico storico della fisica teorica macroscopica, restano di sostanziale pertinenza fisico-geometrica (e fisico-matematica), e meritano o meriterebbero la conveniente attenzione.

Senza dubbio, una componente centrale dell’oggetto del nostro libro è in quella epocale transizione dalla geometria euclidea alle geometrie non-euclidee – in fattispecie alla geometria pseudoeuclidea della relatività speciale (o “ristretta”) e a quella pseudoriemanniana della relatività generale (entrambe 4-dimensionali, ma facilmente generalizzabili ad un numero arbitrario di dimensioni spaziali) – che costituisce una delle più straordinarie avventure nella storia delle scienze esatte, nonché una delle più alte conquiste intellettuali di ogni tempo. Preceduta da un lungo lavoro preparatorio, la grande vicenda si concentra in nove/dieci decenni a partire dal 1830 circa, vero e proprio “sæculum mirabile” della geometria e della fisica matematica. Essa annovera tra i suoi protagonisti (qui elencati per data di nascita) matematici e/o fisici-matematici del calibro di Gauss, Grassmann, Riemann, Beltrami, Lie, Klein, Ricci-Curbastro, Poincaré, Bianchi, Hilbert, Minkowski, É. Cartan, Levi-Civita, Weyl, e, giusto per fermarci, de Rham; oltre naturalmente alle grandi figure di Maxwell, di Lorentz e soprattutto di Einstein, fisici teorici e (soprattutto l’ultimo) epistemologi. L’esperienza cruciale, che segna il passaggio dalla fisica spazio-temporale “separatamente” 4-dimensionale (o per così dire, “(3+1)-dimensionale”) di Euclide-Newton a quella “congiuntamente” 4-dimensionale di Minkowski-Einstein, e una decina di anni appresso di

Riemann-Einstein, è soltanto del 1887.¹⁷ Di questa drammatica evoluzione della conoscenza del mondo fisico macroscopico l'autore ha tentato di offrire una specie di sommario storico nella Sez. 9.1 che introduce all'ultimo capitolo, e che è concepito in modo da essere relativamente indipendente dai contenuti che lo precedono.

Il nucleo della seconda parte del libro, in corrispondenza con i Capp. 5 ÷ 7, contiene invece argomenti di sapore francamente classico; per quanto riesca approssimativo e semplicistico nominarne il carattere come quello di “strumenti” (come si è fatto) piuttosto che di “contenuti”. Ad ogni modo questa classificazione non ha alcuna importanza: ciò che conta è aver dedicato la conveniente attenzione a quegli universi dell'analisi e della fisica matematica che sono ad esempio il calcolo tensoriale o quello delle variazioni, la teoria delle equazioni differenziali (sia ordinarie che alle derivate parziali), la teoria delle varietà differenziabili, .. e via dicendo – beninteso limitatamente ai loro aspetti fondamentali ed alla illustrazione esemplificativa di qualche applicazione.

Ancora un commento merita l'opportunità di presuntivamente ricostruire ed illustrare i processi induttivi sottostanti alla messa a punto dei modelli fisico-matematici, e in particolare dei modelli classico e relativistico dello spazio-tempo (senza e con gravità). Si tratta evidentemente di un problema didattico *di natura generale*, che si pone nella descrizione di qualunque teoria fenomenica matematizzata. Secondo il punto di vista fisico-matematico, questa materia è confinata all'interesse storico e psicologico, e dovrebbe quindi ricevere attenzione soltanto da parte degli storici della scienza e degli psicologi dell'invenzione scientifica: un'idea sostanzialmente corretta, qui accolta quasi senza eccezioni.

Indubbiamente, esiste una specie di filo rosso che partendo dalla realtà osservata, ed attraversando le segrete regioni dell'intuizione, porta alla possibile identificazione delle basi di una teoria fisica matematizzata (o teoria fisico-matematica); cioè, delle definizioni dei suoi termini, relazioni ed assiomi¹⁸. Altrettanto indubbio è che il tentativo di ripercorrere quel filo possa avere un ruolo ai fini di una migliore comprensione. Tuttavia è raro, a nostro avviso, che questo possibile aspetto della trattazione sia veramente efficace dal punto di vista didattico. Infatti il momento induttivo dello sviluppo di una teoria fisico-matematica ha essenzialmente la natura di una selezione tra innumerevoli scelte alternative, guidata dal sondaggio mentale delle loro implicazioni e dal finale confronto di queste con la realtà osservata; attività sfortunatamente quasi sempre sepolta nel segreto di chi l'ha effettuata, e quindi di ricostruzione assai problematica. In conclusione, questo tipo di (presunta) informazione ci appare di dubbia utilità didattica, quando non francamente

¹⁷ A.A. Michelson, E.W. Morley, Amer. Journ. Sci. **34** (1887). L'interesse e l'attività di Michelson attorno al problema risalgono tuttavia a qualche anno prima.

¹⁸ Le regole di inferenza, in una teoria fisico-matematica, sono in pratica sempre quelle (cosiddette) “naturali”.

fuorviante, se il fine principale è quello di una descrizione per quanto possibile *oggettiva ed economica* della teoria stessa.

0.0.2) I CONTENUTI CAPITOLO PER CAPITOLO: SINTESI E COMMENTI

Prescindendo dall'Introduzione 0.1, intesa a lumeggiare aspetti molto generali del tema che sarà svolto nei capitoli successivi, come già si diceva il libro è diviso in tre parti (Parte I, Capp. 1÷2, Parte II, Capp. 3÷7, Parte III, Capp. 8÷9), più una Parte IV contenente le Appendici Generali. Entro certi limiti, il discorso segue una traccia “geometrica”. Questo è ben evidente all'inizio, con il Cap. 1 interamente dedicato alla geometria euclidea 3-dimensionale in formulazione sintetico-assiomatica. Tale carattere assiomatico è meno esplicito nel Cap. 2, che ha per oggetto la geometria pseudoeuclidea 4-dimensionale della relatività speciale; e ciò, in quanto è apparso innaturale introdurre in tal modo questa geometria iperbolica, della quale la comune esperienza quotidiana – a differenza di quanto accade per quella euclidea – non ci suggerisce nulla di immediatamente evidente. La seconda parte del libro ha natura più ibrida, pur restando parzialmente ispirata alla geometria. Questo è particolarmente palese nel Cap. 4, che tratta della geometria delle varietà astratte, e nel Cap. 5 in cui ci si comincia ad occupare di integrazione, cioè dell'aspetto meno astratto della teoria della misura. I successivi Capp. 6 ÷ 7 tornano allo spazio-tempo 4-dimensionale euclideo-newtoniano come contenitore naturale della fisica matematica macroscopica, concentrandosi sulla sua formulazione variazionale; ed è appena il caso di ricordare le molte ascendenze geometriche del calcolo delle variazioni. La terza parte riprende infine la linea geometrico-fisica, almeno nella misura in cui la relatività generale (che è stata di recente felicemente rinominata come “geometrodinamica”) è una teoria puramente geometrica. Segue una breve illustrazione dei singoli capitoli.

§1. Dopo una introduzione sostanzialmente informale (Sez.1.1), il primo capitolo sviluppa una assiomatizzazione di tipo metrico, cioè fondata sulla nozione primitiva di distanza (secondo il classico suggerimento di B. Riemann), della geometria euclidea sintetica (Sez. 1.2); di fatto, il più venerabile esempio di teoria fisico-matematica. Il modello al quale si approda in modo naturale è quello di uno spazio isomorfo a \mathbf{R}^3 -affine, metrizzato con la metrica pitagorica e possibilmente orientato, ossia allo spazio euclideo 3-dimensionale E^3 (o \hat{E}^3 se orientato).¹⁹ Lo spazio sintetico assiomatico orientato \hat{H}_3 e lo spazio euclideo orientato empirico risultano così \aleph_1 -categorici, nell'usuale senso che essi ammettono \hat{E}^3 come modello normale di cardinalità \aleph_1 , e che il modello

¹⁹ Sono anche offerti spunti sufficienti ad estendere la trattazione all'analogo caso $(n \neq 3)$ -dimensionale.

è unico a meno di isomorfismi. Di tale isomorfismo tra \hat{H}_3 e \hat{E}^3 si tratta nella Sez. 1.3. L'ultima Sez. 1.4 si occupa infine di questioni di estensione (o misura²⁰) e di orientamento. Due appendici speciali, dedicate alle relazioni d'ordine e rispettivamente alle funzioni Cos e Sin (da non confondere con cos e sin), concludono il primo capitolo. Da una parte l'estrema vastità della materia e dall'altra la complessità del problema dei fondamenti danno a questo capitolo un carattere relativamente compendioso, nonostante le sue non trascurabili dimensioni. §

§2. Alla geometria pseudoeuclidea si passa nel Cap. 2, con l'introduzione della lunghezza römeriana come quarta coordinata. Per ovvie ragioni logiche e storiche, alle tre sezioni strettamente relativistico-speciali 2.2 ÷ 2.4 si è premessa una breve rassegna dei fondamenti della cinematica e della meccanica newtoniane (Sez. 2.1). Il supporto matematico alla cinematica-dinamica relativistica-speciale è fornito dall'algebra di Minkowski sviluppata nelle Sezz. 2.3 e 2.4. Questa e la geometria associata vengono illustrate nella misura necessaria alla finalità principale, che è quella di descrivere la fisica che le ispira mantenendo per quanto possibile in primo piano il contatto tra nozioni formali e corrispondenti operazioni e verità empiriche. Ricordiamo qui come l'intera teoria della relatività speciale, rivoluzionaria sotto il profilo fisico e filosofico, fu e resta descrivibile, nei suoi contenuti di base, mediante un apparato matematico assolutamente elementare: in pratica, le quattro operazioni più l'estrazione di radice quadrata e (ma si può rinunciare senza serie conseguenze) le regole leibniziane di derivazione. Il capitolo prosegue con una Sez. 2.5 dedicata ai fondamenti della teoria elettromagnetica classica (vera e propria "madrina" della relatività speciale), e si conclude con nove appendici speciali: la prima a complemento della Sez. 2.1, la seconda della Sez. 2.2, la terza dedicata ad approfondimenti sulle trasformazioni di Lorentz, la quarta alle formule di trasformazione di Lorentz "parallela", la quinta alle leggi di conservazione relativistico-speciali per sistemi di punti materiali, la sesta ad una induzione alternativa, su base puramente meccanica, del concetto di "massa relativistica di moto", la settima ai cosiddetti punti di vista lagrangiano ed euleriano nella descrizione di un continuo passibile di spostamenti, l'ottava alla dinamica classica e relativistica di un continuo materiale, e l'ultima ai paradossi implicati dalla possibile esistenza di segnali superluminali. Con il secondo capitolo si conclude la prima parte del libro. §

Nella seconda parte si cominciano a predisporre gli strumenti matematici di diretta pertinenza alla fisica matematica macroscopica. (Come si diceva, e come è ovvio, essi escludono gran parte dei contenuti dei corsi istituzionali di Algebra, Analisi, Geometria, ... ecc., dei quali il lettore si presuppone già edotto.) L'orizzonte matematico si allarga con ciò considerevolmente, ponendosi spesso al di là di quanto un laureato in fisica (e talvolta anche in matematica) ha appreso dai suoi corsi universitari. Procedendo per il possibile secondo un ordine logico, nomineremo:

²⁰ Alla Teoria della Misura è dedicata l'App. Gen. C, non limitata peraltro alla misura dei soli insiemi di \mathbf{R}^n .

(1) l'algebra tensoriale in spazi (euclidei e) pseudoeuclidei di dimensione finita;²¹ (2) la geometria differenziale di varietà pseudoriemanniane immerse o incastonate (“embedded”) in uno spazio pseudoeuclideo di dimensione convenientemente maggiore, e l'analisi tensoriale, differenziale e integrale, su tali varietà; (3) la teoria delle varietà astratte r -differenziabili (in particolare per $r = \infty$, o “lisce”), con e senza metrica²² – nel primo caso generalmente indefinita, e in particolare con segnatura lorentziana –, e la relativa analisi tensoriale; (4) all'interno dell'analisi su varietà differenziabili, in particolare, la teoria delle forme differenziali esterne (completamente antisimmetriche, o “alterne”) e la connessa algebra di Grassmann, i teoremi di Poincaré e di Frobenius, la teoria dell'integrazione su varietà, il teorema cosiddetto “generalizzato di Stokes”, ecc.; (5) il calcolo delle variazioni, che provvede potenti strumenti logici all'euristica e all'interpretazione di una vasta classe di modelli fisico-matematici, compreso il modello relativistico-generale (azione di Hilbert); (6) i legami tra il calcolo delle variazioni e la geometria differenziale classica e relativistica; (7) la teoria dei sistemi differenzialparziali (SDP) in generale (ma in particolare di quelli quasi-lineari iperbolici del secondo ordine, con riferimento al SDP gravitazionale di Einstein-Hilbert (EH)) e connesso problema di Cauchy; (8) il calcolo di soluzioni esatte o approssimate di questo SDP (perché le equazioni tacciono, mentre le soluzioni parlano); (9) questioni di analisi tensoriale *globale* su varietà, astratte e non, generalmente pseudoriemanniane, o anche soltanto affinemente connesse, ... e via elencando. Di questo rispettabile arsenale, qui passato in rassegna molto sommariamente, una scelta di argomenti di maggior rilievo ai nostri fini è presentata nella seconda e nella terza parte del libro.

§3. Il terzo capitolo si apre con una prima Sez. 3.1 dedicata all'algebra dei κ -tensori (con κ intero non-negativo) in spazi pseudoeuclidei, passando attraverso quella ad essa isomorfa delle forme κ -lineari. Segue l'illustrazione di un insieme un po' disorganico di applicazioni di interesse fisico-matematico immediato (Sez. 3.2). Le Sezz. 3.3 ÷ 3.4 introducono alla analisi tensoriale in varietà pseudoriemanniane immerse (o incastonate) in un conveniente spazio pseudoeuclideo; mentre la quinta Sez. 3.5 offre un sunto di quella teoria della curvatura che così grande importanza concettuale riveste per una teoria generale dello spazio-tempo – nonché delle curvature di ordine superiore al secondo di *curve* congruentemente differenziabili, ma limitandosi a curve immerse in uno spazio *euclideo*. Chiudono il terzo capitolo un paio di brevi appendici speciali. §

²¹ Tutti gli spazi-base (e tutte le varietà “pure”) presi in considerazione sono da supporre di dimensione finita; e ciò, sebbene il passaggio a varietà di Banach non comporti speciali difficoltà, apparendo del tutto naturale a molti matematici contemporanei.

²² Naturalmente la prescrizione di una metrica è essenziale in geometria differenziale relativistica, ma è importante mostrare come una ragionevole geometria differenziale possa svilupparsi, su una data varietà abbastanza regolare, anche in sua assenza, introducendovi una conveniente connessione (o addirittura anche prescindendo da quest'ultima).

§4. Con il Cap. 4 si apre l'importante argomento delle varietà *astratte* n -dimensionali, topologiche e differenziabili, trattate come insiemi localmente idempotenti a \mathbf{R}^n (possibilmente orientato) e per il resto generici. (Quest'ultima scelta si discosta da quella della maggior parte della trattatistica). Una prima Sez. 4.1 fornisce le nozioni di base, toccando anche alcune questioni più generali, come ad esempio quella delle varietà bordate o quella delle varietà dimensionalmente "non-pure". La conoscenza di alcune nozioni/teoremi fondamentali dell'analisi delle funzioni di più variabili reali (ad esempio del teorema della funzione implicita), come pure delle numerose nozioni topologiche coinvolte nella teoria, è qui essenzialmente presupposta. La seconda Sez. 4.2 è dedicata ai fondamenti dell'analisi differenziale su (o tra) varietà astratte congruentemente differenziabili, presentata innanzitutto mediante il cosiddetto "metodo delle carte" (ma sostanzialmente discussi sono anche il "metodo algebrico" e quello "dei germi di curve", mostrandone l'equivalenza al primo). Alla fine di questa sezione, il lettore dovrebbe padroneggiare gli importanti concetti di spazio tangente e cotangente, di base e cobase canoniche, e del differenziale (o mappa tangenziale) di un'applicazione di una varietà in un'altra, in particolare di (un germe di) funzione reale sulla varietà, con riferimento ad un suo punto generico prefissato. Un sommario delle algebre, tra loro equivalenti, delle $\langle a, b \rangle$ -forme multilineari (con a e b interi non negativi) e dei corrispondenti $\langle a, b \rangle$ -tensori, ora in generici spazi n -dimensionali lineari (su \mathbf{R}), occupa la prima parte della Sez. 4.3. Dopo di ciò il terreno è pronto per passare alle nozioni di campo $\langle a, b \rangle$ -tensoriale su una varietà n -dimensionale astratta (in generale senza metrica); e avendovi assegnata una connessione, alla relativa analisi (locale) differenzial-tensoriale. Di quest'ultima si occupa il resto della Sez. 4.3, in cui si ripercorrono alcune importanti tappe dell'analoga analisi su varietà immerse in spazi pseudoeuclidei, ormai generalmente in presenza di torsione. La Sez. 4.3 si chiude infine con l'esame dei campi $\langle a, b \rangle$ -tensoriali cosiddetti "relativi" e della loro analisi differenziale. La successiva Sez. 4.4 è dedicata all'algebra delle forme multilineari completamente simmetriche e di quelle (ben più familiari al fisico teorico) completamente antisimmetriche. L'algebra di queste ultime è propedeutica alla analisi differenziale esterna su varietà e ai più importanti teoremi integrali ad essa connessi, temi trattati nell'ultima Sez. 4.5 del capitolo. Superfluo aggiungere che, oltre alla sua importanza intrinseca, il riferimento a varietà differenziabili cosiddette "non-elementari" diventa obbligato quando l'attenzione si sposti dai problemi locali a quelli globali. L'App. Spec. 4.A estende le classiche formule di Frenet-Serret ad una generica varietà pseudoriemanniana, e la 4.B introduce alle geometrie a connessione affine, accennando poi alla teoria unitaria di Weyl. §

§5. Come dicevamo, il quinto capitolo fa in certo senso un passo indietro, occupandosi di questioni che sono alla base di gran parte della fisica matematica classica. Esso si apre con la vasta tematica

dell'integrazione, molto legata ai contenuti dell'App. Gen. C. Attraverso le nozioni di spazio carico e normalmente carico, la prima Sez. 5.1 del capitolo offre una linea di accesso concettuale all'integrazione più diretta di quella praticabile a partire dalla usuale teoria della misura. In particolare si possono così agevolmente valutare le misure di Jordan (J-misure) di sottoinsiemi notevoli di \mathbf{R}^n come i $(k \leq n)$ -blocchi, i k -simplessi e le k -palle. L'ultima parte della sezione tratta dell'integrazione su sottovarietà di \mathbf{R}^n . La successiva Sez. 5.2 illustra, ovviamente in modo non sistematico, quel fondamentale rapporto tra integrazione e differenziazione che trae origine dal teorema di Gauss-Ostrogradskij, e che costituisce uno dei capisaldi analitici della fisica matematica. L'argomento è vastissimo, e se ne è quindi presentata più che altro una esemplificazione significativa dal punto di vista dei fondamenti. Nonostante qualche perplessità sulla opportunità di offrire qui delle informazioni generali, ma non futili, sull'universo delle equazioni differenziali (o loro sistemi) di interesse fisico-matematico, l'autore si è risolto a trattare del problema ai valori iniziali (problema di Cauchy), sia normale che generalizzato per sistemi differenzialparziali (SDP) di ordine arbitrario, e della teoria di un'unica equazione differenzialparziale (EDP) del 1° ordine (la "madre di tutte le EDP"): argomenti classici e di primaria importanza dal punto di vista dei fondamenti della fisica matematica. Di essi si leggerà nella Sez. 5.3 e rispettivamente nella Sez. 5.4. Un'appendice speciale è infine dedicata ai problemi di Dirichlet e di Neumann interni ed esterni, alla luce della teoria delle equazioni integrali di Fredholm. §

§6. Passando al sesto capitolo, le Sezz. 6.1 e 6.2 sono interamente dedicate al calcolo delle variazioni (CDV) unidimensionale (o "a integrale singolo") e ai suoi aspetti/applicazioni più importanti. Anche in questo caso l'argomento è praticamente sterminato, e si è resa necessaria una energica selezione per contenere il discorso entro limiti ragionevoli. Successivamente, la Sez. 6.3 presenta una sintesi della meccanica analitica classica nella sua accezione locale (e in particolare della teoria delle trasformazioni canoniche). La Sez. 6.4 illustra appunto alcuni aspetti ed applicazioni di questa meccanica; tra i primi, soprattutto l'equazione di Hamilton-Jacobi, (esempio canonico della possibile equivalenza tra una EDP ed un sistema differenziale ordinario (SDO)), e tra le seconde quelle all'astronomia (o cosiddetta "meccanica celeste"). Infine l'App. Spec. 6.A offre una rivisitazione della teoria delle coniche *in rappresentazione polare*, mentre la 6.B dà brevi informazioni complementari sulla (formulazione variazionale della) dinamica relativistica del punto. §

§7. Con il settimo capitolo si passa al CDV multidimensionale (o "a integrale multiplo"), o "teoria dei campi (classici)" nel linguaggio dei fisici. Va peraltro ricordato che il legame tra questo CDV multidimensionale e la teoria generale dei SDP/EDP è ormai così intimo da rendere irrimediabilmente ingenuo qualsiasi tentativo di trattazione sistematica dell'argomento, soprattutto

in un libro come il presente. La Sez. 7.1 pone le basi concettuali della teoria, illustrandone alcune importanti applicazioni fisico-matematiche. La Sez. 7.2 torna invece a problemi variazionali più generali, soffermandosi principalmente sulla teoria di Hamilton-Jacobi, ma toccando anche concetti come quello dell'integrale indipendente di Hilbert o della funzione di eccesso di Weierstrass (S.sez. 7.2.1). Di seguito, la S.sez. 7.2.2 è dedicata ad un esame critico del significato e dell'uso di quel teorema di Noether che è così popolare tra i fisici teorici. Infine la Sez. 7.3 tratta dei problemi variazionali cosiddetti "omogenei", soprattutto unidimensionali. Conclude il capitolo una appendice speciale che introduce all'importante problema della propagazione di fronti di discontinuità di soluzioni di sistemi differenziali parziali quasi-lineari. §

Termina così anche la seconda parte del libro.

§8. I capitoli 2, 3 e 4 contengono abbondante materiale preparatorio all'oggetto specifico della terza parte, i fondamenti matematici della teoria relativistica. Tenuto conto del disegno complessivo del libro e del suo carattere generale, questo ha consentito di concludere nello spazio relativamente breve di due capitoli la nostra trattazione dell'argomento. L'ottavo capitolo continua dunque l'illustrazione di problemi geometrico-differenziali propedeutici – con vario grado di irrinunciabilità – allo studio dello spazio-tempo einsteiniano. Dopo una rivisitazione di alcuni concetti di base lungo una linea logica alternativa, la Sez. 8.1 si occupa di certe questioni complementari – ad esempio del parallelismo geodetico e delle coordinate semigeodetiche – per concludere con i classici teoremi di Gauss "del triangolo geodetico" e di Levi-Civita "del trasporto parallelo". (Un'ultima S.sez 8.1.4 tratta anche delle 2-superfici immerse in un 3-spazio minkowskiano.) La successiva Sez. 8.2 illustra altre questioni complementari, e in particolare quella della transizione dalla geometria pseudoriemanniana a quella "a connessione affine", concludendo con l'analisi dell'operatore differenziale del 2° ordine cosiddetto "commutatore assoluto" in varietà a connessione generalmente non-simmetrica. Dopo un esame generale dell'algebra dei campi tensoriali su varietà, la Sez. 8.3 passa alla importante nozione di derivata di Lie – secondo un dato campo vettoriale – di un campo tensoriale in una varietà regolare ma priva di connessione (varietà "ordinaria"), per estenderla infine ai campi di tensori relativi e di forme differenziali esterne. Con la Sez. 8.4 si affronta il problema dell'integrazione su domini di varietà orientate, arrivando alla dimostrazione del teorema di Poincaré-Stokes (la dualità tra forme e catene non può tuttavia essere messa nella sua piena luce in assenza del duale dell'operatore di "bordificazione" (su catene), come previsto dalla teoria omologica assiomatica). Stiamo qui penetrando, per quanto ci sembra, nella zona grigia di un interesse marginale da parte del fisico-matematico "macroscopico" medio, almeno alla data presente. Ciononostante, si è ritenuto vi fosse ancora spazio per qualche informazione sulle forme differenziali esterne in varietà pseudoriemanniane (Sez. 8.5), introducendo il concetto della

loro “codifferenziazione” (via pull-back alla Hodge), ma limitandosi al caso elementare di \mathbf{R}^n e coordinate ortonormali. Si può così provare che l’operatore differenziale del 2° ordine cosiddetto di Laplace-Beltrami (su forme) coincide, in tali condizioni, con l’operatore di Laplace standard. Si passa quindi a quello che abbiamo detto “problema ∂ - δ ”, naturale generalizzazione del problema “rot-div” in \mathbf{R}^3 (per un campo vettoriale), vedi S.sez. 5.2.3. In forza della loro importanza, un’ultima S.sez 8.5.4 è dedicata a rudimenti di topologia algebrica in termini di semplici, e ai due teoremi di de Rham (enunciati senza dimostrazione). Il capitolo si conclude con una appendice speciale sui modelli canonici del piano ellittico (di Riemann-Klein) e del piano iperbolico (di Poincaré). §

§9. Entreremo in qualche maggior dettaglio sui contenuti di questo ultimo lungo capitolo. Della prima sezione 9.1 abbiamo già detto. Il capitolo prosegue (Sez. 9.2) con un approfondimento della geometria p.euclidea/p.riemanniana, sia in chiave algebrica (S.sez. 9.2.1) che analitica (S.sez. 9.2.2). In particolare, la S.sez. 9.2.1 offre la dimostrazione di alcuni teoremi soltanto enunciati nel Cap. 2: quello dell’esistenza di una base B-ortogonale, quello cosiddetto “del rango” e quello di Sylvester. Descritti sono anche il protocollo di Gauss e il metodo di Jacobi per la costruzione di una base canonica. Tra i corollari, figura il criterio di Sylvester sulla definitezza positiva di una forma quadratica. La S.sez. 9.2.2 è invece dedicata ad un insieme di problemi analitici su una varietà lorentziana 4-dim. Oggetto di questi approfondimenti sono innanzitutto le equazioni geodetiche; segue quindi l’esame delle nozioni di tetraede p.ortonormale, di matrice lorentziana, di trasporto alla Fermi-Walker, ecc. Con la Sez. 9.3 siamo finalmente al cuore del capitolo, cioè ai fondamenti matematici della relatività generale. Questi sono riassunti negli assiomi \mathbf{A}_1 (S.sez. 9.3.4, proprietà dello spazio-tempo 4L), \mathbf{A}_2 e \mathbf{A}_{2bis} (S.sez. 9.3.5, assiomi della geodetica) e \mathbf{A}_3 (S.sez. 9.3.5, equazioni geometrodinamiche di Einstein-Hilbert (EH)). La sezione contiene anche il calcolo esatto del tensore einsteiniano $E_{(2)}$ statico, che viene poi utilizzato per determinare, via principio di corrispondenza, la costante di proporzionalità nelle equazioni di campo in termini di costante di Newton-Cavendish. Le successive sezioni 9.4 e 9.5 sono interamente dedicate alle applicazioni fondamentali della teoria relativistica generale – che in certi casi hanno permesso le prime storiche verifiche della sua validità – e ad alcuni importanti complementi. La Sez. 9.4 parte con l’analisi geodetica di certe varietà non-piatte (S.Sez. 9.4.1), applicandola in particolare al cosiddetto “paradosso dei gemelli” (S.sez. 9.4.2). Si passa poi, per così dire, “dall’altra parte” delle equazioni di EH (S.sez. 9.4.3 e 9.4.4). Il problema è qui quello della determinazione del tensore energetico in casi concreti e fisicamente significativi; in pratica, nel caso del continuo materiale con sforzi “di contiguità”, solido o fluido (vedi anche App. Sp. 9.C). La Sez. 9.5 inizia allo studio delle soluzioni delle equazioni di EH, e in particolare a quelle a simmetria sferica, giungendo così alle metriche di Schwarzschild esterna (nel vuoto) ed interna (in un fluido) (S.sez. 9.5.1). Dopo di questo, si viene

alle metriche “formalmente invarianti”, e quindi alla nozione di campo vettoriale di Killing (S.sez. 9.5.2). Nella stessa sottosezione si discute anche la versione di base del teorema di Birkhoff, che offre una più convincente giustificazione delle soluzioni di Schwarzschild. Con la S.sez. 9.5.3 si affronta la dinamica relativistica generale del punto materiale e (come caso-limite) del fotone, e le sue più classiche applicazioni: lo studio della precessione dell’orbita di Mercurio e rispettivamente quello dell’incurvamento dei raggi luminosi che rasentano la superficie solare. Completano il capitolo sette appendici speciali. La prima ha essenzialmente interesse storico; la seconda è una raccolta di formule di trasformazione di Lorentz che il lettore volenteroso potrà controllare a titolo di esercizio; la terza (già segnalata) è dedicata al tensore degli sforzi in approssimazione relativistica speciale; la quarta illustra certi complementi sul gruppo delle trasformazioni di Lorentz parallele (o “L-boosts”), conducendo alla precessione di Thomas e alle trasformazioni di Møller; la quinta verte sulle geodetiche della varietà di Schwarzschild esterna, con applicazione ai moti radiali e circolari; la sesta riprende la trattazione generale dei fronti di discontinuità delle soluzioni di sistemi di 2° ordine iperbolici (vedi App. Sp. 7.A) applicandola alle onde di gravità; e infine l’ultima approfondisce la questione del carattere sottodeterminante delle equazioni di EH con l’introduzione delle coordinate pseudoarmoniche.

§10. Diamo infine i titoli, ai quali qui ci limitiamo, delle sei appendici generali: “Nozioni elementari di logica e di teoria degli insiemi” (App. Gen. A); “Glossario ragionato di topologia” (App. Gen. B); “Strutture di misura” (App. Gen. C); “Introduzione alla scienza computazionale” (App. Gen. D); “Sui fondamenti della termodinamica classica” (App. Gen. E); “Elementi di cosmologia teorica macroscopica” (App. Gen. F).

Completano l’opera: 1) la Bibliografia Generale; 2) una lista degli acronimi usati; 3) i quattro Glossari (delle parti 1^a, 2^a, 3^a e delle appendici generali), e 4) un elenco dei nomi notevoli con le relative cronologie.

0.0.3) “ISTRUZIONI PER L’USO”

La maggior parte dei termini/espressioni scritti in **grassetto** al loro primo apparire sono elencati nei glossari insieme con la corrispondente sottosezione. Si tratta di solito di definizioni in senso stretto (ma talvolta anche “ragionevolmente” tali), oppure di espressioni con cui si introducono nozioni importanti nel contesto. Si è ritenuto preferibile redigere quattro glossari separati, relativi rispettivamente alle parti prima, seconda, terza, e all’insieme delle appendici generali, piuttosto che uno soltanto. Si è anche fatto un uso abbastanza liberale di acronimi (alcuni

del tutto standard, altri inconsueti e ad hoc), il cui elenco alfabetico è fornito alla fine del libro. Nella sola Sez. 1.1, le parole “formali” e tuttavia dotate di significato nel linguaggio naturale sono racchiuse tra $\{ \}$ al fine di iniziare il lettore all’uso di parole di uso comune in senso tecnico-ingenuo (o in generale semplicemente suggestive), ma prive di significato formalmente definito. All’opposto, nella Introduzione 0.1 e nella Nota Storica 9.1 la terminologia è usata con tutta la conveniente libertà.

Le “ ” hanno le usuali e flessibili finalità della fraseologia scritta. Con l’eccezione dei titoli di memorie o monografie/trattati (che sono trascritti tra “ ”), le citazioni letterali (a meno della traduzione) sono sempre riportate tra « ». Rinunciando ad altre più abituali convenzioni, le stesse « » racchiudono poi gli enunciati dei teoremi, lemmi, corollari (denominati $T_1, \dots, L_1, \dots, C_1, \dots$, ecc. oppure $T1, \dots, L1, \dots, C1, \dots$, ecc.) in modo che molti di essi possono essere inseriti nel testo risparmiando spazio. Il segno # posto dopo un punto fermo è segno di conclusione, tipicamente di una dimostrazione. Argomenti collegati da analogie, o facenti parte di una lista, sono spesso aperti da un § numerato progressivamente (a partire da §1 all’interno della stessa sottosezione), e chiusi da un altro §.

Il segno “=” significa “uguale per definizione”, mentre “ \equiv ”, oltre all’usuale significato di “uguale identicamente” (rispetto a qualche variabile), è anche usato, nel linguaggio corrente, come segno di equivalenza logica. Le parentesi quadre [...] fraseologiche hanno significato di congiunzione (“e”, cioè “et”), ovvero di alternanza (“o”, cioè “aut”), ovvero di congiunzione/alternanza (“e/o”, cioè “vel”); vale a dire, “[...]” sta a seconda dei casi per “et (aut, vel) rispettivamente ...”. Il carattere *corsivo* ha sempre il ruolo di enfattizzatore.

Il riferimento alle sezioni e alle sottosezioni dei vari capitoli è “x.y” e risp. “x.y.z”, dove “x” è il numero del capitolo, “y” quello della sezione e “z” quello della sottosezione. Il riferimento alle formule è “(x.y.z, u)” [“(x.y.z, (u,v,w, ..))”] dove “u” è il numero d’ordine della formula richiamata [dove “(u,v,w, ..)” sono i numeri d’ordine delle formule richiamate], e “x.y.z” quello della sottosezione; ma “x.y.z” è ovviamente omissa, in quelle combinazioni, all’interno della sottosezione corrente.

La numerazione delle formule e dei teoremi riparte da (1) all’inizio di ogni sottosezione. In qualche caso, a questo scopo si è preferito usare un riferimento non numerico (come (°), (*), (†), ecc.), intendendosi che *il rinvio è valido soltanto all’interno del capoverso corrente*. La numerazione delle note a piè pagina riparte da (¹) all’inizio di ogni sezione, insieme di appendici speciali o appendice generale. Il riferimento alle appendici generali poste alla fine del libro è nella forma App. Gen. A, App. Gen. B, .. ecc.; quello alle appendici speciali poste alla fine del Capitolo

x, nella forma App. Spec. x.A, App. Spec. x.B ... ecc. Il rinvio alla Bibliografia Generale è “(vedi Bibl. Gen. x)”, dove x indica la sezione della Bibliografia Generale.

Alla loro *prima* occorrenza nel testo – note a piè pagina comprese, ma esclusa questa Presentazione –, ai nomi degli studiosi contestualmente più importanti seguono tra parentesi il (solo) primo nome, la cronologia, e in alcuni casi, i luoghi di nascita/morte. Fanno eccezione le prime occorrenze di tipo attributivo, come ad es. in “scala di Planck” o “disuguaglianza di Schwarz”. Un elenco di tutti gli autori citati, cronologie e luoghi di nascita/morte compresi, si trova alla fine del libro. Per quanto siano semplici atti di diligenza redazionale, questi strumenti possono essere utili ai fini di un inquadramento storico minimale di ciò di cui si ragiona.

Veniamo ora alle notazioni matematiche. In linea di massima, abbiamo cercato di suggerire al lettore un buon grado di flessibilità piuttosto che bloccarlo su convenzioni fisse, fornendogli comunque indicazioni sufficienti ad evitare equivoci. Per cominciare, abbiamo quasi sempre rinunciato al tradizionale carattere grassetto per denotare vettori, o ad altri caratteri speciali per denotare ($\kappa > 1$)-tensori. Le componenti vettoriali o tensoriali si distinguono subito perché affette da indici, soprascritti o sottoscritti destri. Quanto al modulo di un vettore v , esso è indicato con $|v|$ quando si presenta realmente come tale; ma se è quadrato, abbiamo usato v^2 in luogo di $|v|^2$. Se poi vogliamo riferirci ad un ($\kappa > 1$)-tensore come ad un tutto, spesso apponiamo un sottoscritto destro (κ) alla lettera usata per le sue componenti; oppure, talvolta, scriviamo quella lettera, affetta dai suoi indici, tra $\{ \}$ (o tra le $\langle \rangle$ di insieme ordinato). Indici latini sono intesi variare da 1 a n (tipicamente, la dimensione della varietà di riferimento), o nel caso specifico dello spazio-tempo, da 1 a 4; indici greci, invece, sono intesi variare da 1 a $n - 1$ (nel caso dello spazio-tempo, da 1 a 3). Una riga o colonna di oggetti indicizzati da 1 a n , diciamo a_i , viene di solito denotata con $\{a_i\}_{i=1 \rightarrow n}$ o $\{a_i\}_{i=1, \dots, n}$, o anche soltanto $\{a_i\}$, o addirittura $\{a.\}$. Similmente nel caso di oggetti a due (o più) indici, ad esempio elementi a_{ij} di una matrice, la matrice come un tutto si denoterà con $\{a_{ij}\}_{i,j=1 \rightarrow n}$ o come $\{a_{ij}\}$ o come $\{a..\}$ (o anche, come appena detto, con $\langle \rangle$ in luogo di $\{ \}$). Una ($n \times n$)-matrice diagonale di elementi a_{11}, \dots, a_{nn} è talvolta denotata con $\text{diag}(a_{11}, \dots, a_{nn})$; con la convenzione da noi adottata per la relativa metrica, la matrice di Minkowski (spazio lorentziano) viene dunque scritta $\text{diag}(1,1,1,-1)$. Se un indice è ripetuto in alto e in basso, come in $()_i^i$, si sottintende la somma sull'insieme dei suoi valori (\equiv convenzione di somma alla Einstein); se tuttavia, nel caso di una varietà piatta, un indice è ripetuto solo in basso o solo in alto, si sottintende la moltiplicazione per la relativa matrice metrica; per cui, ad esempio in uno spazio lorentziano, $()_{ii} = ()_{11} + ()_{22} + ()_{33} - ()_{44}$. La riga o colonna delle coordinate $\{x_i\}_{i=1 \rightarrow 4}$ viene preferibilmente denotata con x se è argomento di una funzione (anche tensoriale). In certi casi, l'argomento x è denotato come (\mathbf{x}, x_4) ,

dove si sottintende che \mathbf{x} rappresenti $\{x_i\}_{i=1 \div 3}$. La coordinata x_4 è sempre intesa in senso römeriano come ct (o altrettanto frequentemente come ict); abbiamo cioè evitato di usare la celerità della luce c come unità, allo scopo di facilitare i controlli dimensionali delle formule.

Le derivate parziali della funzione f sono denotate con la scrittura esplicita $\partial f / \partial x^i$, o anche $\partial_i f$, e soltanto raramente con una semplice virgola, come $f_{,i}$. Le derivate covarianti [controvarianti] di uno scalare (o in generale di un $(\kappa \geq 1)$ -tensore) sono invece scritte come ${}_{/i} [{}^{/i}]$, piuttosto che ${}_{;i} [{}^{;i}]$ secondo la consuetudine più comune. Il prodotto scalare di due vettori è indicato con un \cdot , e quello vettore 3-dimensionale con \times . Il gradiente di uno scalare f è scritto come ∇f , la divergenza di un vettore v come $\nabla \cdot v$ (quindi l'operatore di Laplace come $\nabla \cdot \nabla = \nabla^2$, ma talvolta anche come Δ) e il suo rotore come $\nabla \times v$. L'operatore di d'Alembert crea un piccolo problema con il suo segno: abbiamo usualmente preferito definirlo con la derivata temporale 2^a a sottrarre. Se non altrimenti specificato, le equazioni si intendono valere in un aperto della varietà di interesse (tipicamente nel dominio di una sua carta), o più di rado, in tutta la varietà.

Il testo è stato composto con Microsoft Word 2000 SR-1 Professional, con successiva elaborazione di files PDF mediante Acrobat Distiller. Delle molte insufficienze di Word 2000 ai fini della stesura di un testo scientifico ricco di formule matematiche, rispetto ai competitori specializzati, forse le più fastidiose sono l'impossibilità di creare sottoscritti e/o soprascritti di secondo livello, o quella di non poter allineare verticalmente sottoscritti e soprascritti (o più in generale di non poter sovrapporre simboli, ad es. un \wedge esattamente sopra un simbolo arbitrario). Si è cercato di ovviare alla prima insufficienza con alcuni espedienti di fortuna (ad es. realizzando il sottoscritto/soprascritto indicizzato " i_1 " come " $i1$ ", o "A elevato alla n^k " come " A^{n^*k} "), nella confidenza che ciò non dia luogo a interpretazioni errate. Quanto alla seconda, certamente meno seria, si sono lasciate le cose come stavano, con l'avvertenza di non dare significato all'apparente ordinamento di sottoscritti/soprascritti che avrebbero dovuto/potuto essere allineati verticalmente. In compenso, un word processor come Word 2000 permette un immediato e flessibile controllo di quanto si scrive o si modifica (quasi come con l'uso di carta, gomma e matita), e questa è la ragione per cui abbiamo optato per esso.

0.0.4) RINGRAZIAMENTI

In primo luogo, desidero ringraziare l'Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali di Frascati – segnatamente nella persona del Direttore S. Bertolucci e dei Direttori che gli sono succeduti, M. Calvetti e U. Dosselli –, per l'ospitalità e l'accesso alla documentazione che mi sono stati offerti. In secondo luogo, il mio grazie va a quanti tra i colleghi (INFN–LNF e non) hanno mostrato interesse per questo lavoro; e innanzitutto a D. Babusci – cui devo anche l'iniziale suggerimento di cimentarmici –, a G. Bellettini, S. Bellucci, G. Capon, G. Isidori, B. Robuch, A. Sestero e soprattutto A. Tenore, che ne hanno volenterosamente letto stralci più o meno importanti. L'ultimo e non meno caldo ringraziamento va a mia moglie Renate, per avermi sopportato con l'amabilità di sempre durante la mia lunga fatica.

C. Lo Surdo

0.1)

ALCUNE RIFLESSIONI GENERALI SULLE TEORIE FISICO-MATEMATICHE ¹

0.1.1) PREMESSA

La creazione di una “teoria fisico-matematica” (TFM) è un processo molto complicato, impossibile da descrivere in termini generali e con accettabile precisione senza richiamarsi a concetti tecnicamente sofisticati ed usare un adeguato linguaggio. Per un verso solitaria e indecifrabile in certi suoi passaggi-chiave, e per l’altro legata alla collaborazione di una pluralità di individui con competenze e ruoli diversi (quindi inevitabilmente pubblica), una tale impresa articolata e improgrammabile, allo stesso tempo esplorativa e inventiva, lascia nondimeno distinguere al suo interno due fondamentali e più o meno concomitanti tipi di attività che potremo dire di **modellazione** e rispettivamente di **matematizzazione**. Con la prima, rivolgiamo la nostra attenzione alla realtà dei fenomeni fisici con l’intento di produrne una rappresentazione idealizzata e in qualche modo organizzata, o “modello” ² ; mentre con la seconda cerchiamo di sviluppare quella rappresentazione in modo da conferirle una specifica capacità di “prevedere” (nel senso dell’implicazione “se ... allora”).

Non è facile descrivere in modo più soddisfacente, ma in poche parole, i due tipi di attività nominate. La prima è di natura **osservativa-induttiva**, e consiste nella raccolta, valutazione, classificazione, interpretazione, analisi correlazionale, .. ecc., dei dati fenomenici di interesse. La seconda è di natura **assiomatica-deduttiva**, e si esplica in una peculiare “organizzazione astratta” di quei dati. In essenza, potremmo dire che una TFM (o correntemente, una **scienza esatta**, o una

¹ I contenuti di questa Introduzione 0.1 sono essenzialmente informali: le definizioni in senso stretto sono rare, e buona parte della terminologia è usata con libertà. Il carattere **grassetto** (vedi Presentazione 0.0.3) serve quindi, per lo più, soltanto ad evidenziare espressioni/nozioni di particolare interesse contestuale, all’atto della loro introduzione. Il carattere *corsivo* è invece usato come evidenziatore, come nel resto del libro.

² È il caso di ricordare che i logici-matematici usano questo termine invertendone il ruolo. Essi chiamano infatti “modello” una realtà di cose, anche soltanto pensate, che sia immagine dell’oggetto principale della loro attenzione, cioè del sistema formale che mima le teorie matematiche di interesse, fornendo di questo una possibile “interpretazione”. In altre parole, i modelli dei logici-matematici – che si potrebbero dire “semantici” – sono immagini dei loro “segni” in un mondo di *cose* (anche astratte). Invece i modelli dei fisici – che si potrebbero dire “teorici” o “matematici” – sono immagini di ambiti della realtà fenomenica in un mondo di simboli. Qui abbiamo usato i termini “segno” e “simbolo” dando al secondo, a differenza del primo e alla luce dell’etimo (σῶμβολον ≡ segno di riconoscimento), una qualche carica significativa. Si noti anche come nei due casi (modelli semantici vs. modelli teorici) si abbiano flussi interpretativi di verso opposto. Ad evitare equivoci, ed anche per brevità, scriveremo “Modelli” per i modelli semantici (salvo che nella App. Gen. A) e “modelli” per i modelli teorici.

scienza in senso stretto) è il risultato della somma sinergica di una modellazione specifica e di una corrispondente matematizzazione specifica.

In realtà l'elaborazione di una TFM è un'impresa alquanto più articolata dell'immagine schematica che ne abbiamo data più sopra. Per limitarci a uno degli aspetti più importanti di questa maggiore complessità, oltre ai menzionati “momento modellizzante” e “momento matematizzante”, di norma a quella elaborazione si associa un “momento tecnologico”, che interagisce con quello strettamente osservativo fornendogli nuovi e più potenti strumenti diagnostici. Ai nostri giorni, questi strumenti raggiungono talvolta un tale grado di sofisticazione, complessità e costo, che la loro stessa realizzazione diventa un obiettivo di per sé.³

Per quanto approssimativa, la precedente definizione di scienza esatta resta comunque sostanzialmente valida. Le molte complicazioni che le sono connesse sono giustificate dal tipo di conoscenza che le TFM mirano a conseguire, e che non è soltanto il “sapere” intorno al mondo dei fenomeni (per quanto questi possano essere difficili da osservare), ma piuttosto il “sapere esplicativo”, cioè il sapere inteso a *capire* (nell'accezione che tenteremo di descrivere) perché i fenomeni si manifestano come si manifestano. Tutto ci fa allora convinti che il sapere esplicativo sia il solo atto a conferirci una capacità predittiva affidabile.

La capacità predittiva orienta la “freccia di valore” all'interno della conoscenza scientifica esatta: una TFM è considerata più valida di una TFM “concorrente” (\equiv avente per oggetto la stessa classe di fenomeni) se e solo se fornisce una capacità predittiva maggiore, alla luce dell'esperienza, nei confronti di quel comune oggetto. Esiste dunque un criterio di valutazione uniforme applicabile alle TFM concorrenti; e in base ad esso, la TFM valutata come predittivamente meno potente, con la più alta certezza disponibile al momento, viene eliminata in favore della TFM più potente.

A questa *uniformità* del criterio di valutazione delle TFM concorrenti si contrappone invece la natura tipicamente *esotica* dei modelli teorici da cui muovono le singole TFM. Non si fa infatti molta strada, nella ricerca di un modello teorico efficace, ispirandosi al cosiddetto (e confortevole) “senso comune”: quasi sempre bisogna spingersi al di là (o contro) di esso per realmente avanzare nella conoscenza scientifica fondamentale. In certi casi, poi, i modelli da cui si sviluppano le TFM sono talmente contro-intuitivi da rendere la loro comprensione ed assimilazione profonda – dopo essere stati formulati da altri, e quindi essere diventati oggetti di apprendimento – praticamente irraggiungibile da parte dell'outsider “normalmente dotato”. Gli esempi più comunemente citati, a questo proposito, sono quello della teoria della relatività e soprattutto quello della teoria quantistica

³ Sarebbe tuttavia sbagliato vedere in questo scambio di ruoli – dopotutto, abbiamo un mezzo che diventa un fine – un fenomeno degenerativo; in primis, perché esso è conseguenza inevitabile della natura “a molte facce” della moderna impresa di ricerca scientifica-esatta, e poi perché il momento tecnologico può avere (e spesso ha di fatto) ricadute sorprendenti su un piano allargato di finalità. Lo sviluppo di Internet, nato da esigenze di comunicazione all'interno del CERN, è un esempio eloquente di questa possibilità.

(nella sua interpretazione ortodossa cosiddetta “di Copenhagen”). Né bisogna credere che il carattere esotico dei modelli fondamentali, da Galileo in poi, sia una novità dei nostri giorni: basti pensare alla teoria newtoniana della gravitazione, che oltre tre secoli fa postulava l’esistenza di una forza trasmessa istantaneamente da un corpo all’altro senza che un mezzo materiale fosse interposto tra essi, e all’impatto di un’idea del genere sulle conoscenze (o sui pregiudizi) dell’epoca.⁴

Venendo meno (o restando in secondo piano) il momento matematizzante, si recede dalle scienze esatte alla scienza tout court, che delle prime è una versione variamente meno affidabile, perché meno (o *assai* meno) dotata di capacità predittiva; e questo, non contando il frequentissimo uso improprio che del termine “scienza” si fa nel parlare corrente (torneremo più avanti su questo punto). Della “scienza senza altri attributi” esiste una gran varietà di definizioni informali: in ogni enciclopedia, o in ogni manuale di filosofia, se ne trova una.⁵ Quello di definire la scienza (o la conoscenza) è del resto un problema filosofico antichissimo, già presente in Platone (427–347 a.C.) e in Aristotele (384–322 a.C.); e poi, nelle “teorie del metodo” di Bacone (1561–1626), in Galileo (1564–1642) e in Cartesio (1596–1650), per approdare all’indagine gnoseologica di Kant (1724–1804), e infine alla moderna epistemologia. Tuttavia, è anche legittimo chiedersi quale possa essere il fine pratico di una tale ricerca, tenuto conto che essa e i suoi esiti interessano poco la maggioranza degli “scienziati”⁶; più o meno per le stesse ragioni per cui poco interessa o interesserebbe ai musicisti una discussione filosofica sulla natura specifica della musica.⁷

Torniamo dunque alle scienze in senso stretto, a partire da quelle più antiche e consolidate, quali sono la geometria euclidea, la meccanica classica “locale” dei sistemi discreti o continui, l’elettromagnetismo di Maxwell, la relatività speciale, .. ecc., e chiediamoci, questa volta con sufficiente legittimità: è oggi possibile dare di queste scienze “matematizzate” una soddisfacente definizione univoca e completa, che le collochi in un comune quadro generale? Questo problema di

⁴ Non affermiamo qui che la rappresentazione teorica dell’oggetto fisico considerato sia necessariamente unica; ma a fronte di rappresentazioni *predittivamente equivalenti* dello stesso oggetto, la scelta privilegia sempre la *più semplice*: «L’esperienza ci lascia una libera scelta, ma la guida aiutandoci a discernere il cammino più comodo» (H. Poincaré, “La Science et l’Hypothèse”, 1903). Naturalmente, attribuita ad una rappresentazione teorica, la qualifica di “semplice” è in parte opinabile; e ciò può dar luogo all’idea che alcuni dei principi che sono alla base della elaborazione di una TFM abbiano la natura di convenzioni. Tra i grandi fisici-matematici di ogni tempo, Poincaré (Henri, Nancy Fr. 1854, Parigi 1912) fu uno dei più inclini a questa visione convenzionalistica, entro ovvi e assai ristretti limiti, della scienza esatta.

⁵ Riportiamo ad esempio la definizione informale che si trova alla voce “science” della Encyclopædia Britannica (Micropædia, 15th ed., repr. 1997): «Any of various intellectual activities concerned with the physical world and its phenomena, and entailing unbiased observations and systematic experimentation. In general, a science involves a pursuit of knowledge covering general truths or the operations of fundamental laws.»

⁶ Come forse non tutti sanno, il termine “scenziato” (nella sua versione inglese di “scientist”) è stato coniato da W. Whewell (1794–1866), che nel 1841 divenne direttore del Trinity College di Cambridge. Prima non esisteva, perché non veniva attribuita alcuna specifica qualifica a chi si occupava di scienza a tempo pieno. Sembra anche che l’idea della nuova parola fu suggerita dall’analogia di “science → scientist” con “art → artist”.

⁷ Della definizione dell’attività “scientifica” si occupa, come è naturale, la Filosofia della Scienza. A proposito della supponenza di molti scenziati verso questo tipo di problemi, viene alla mente il sarcastico ma superficiale aforisma di Feynman (Richard, 1918 - 1988), secondo il quale «la filosofia della scienza serve agli scenziati quanto l’ornitologia serve agli uccelli».

una precisa definizione delle scienze esatte, al quale per semplicità ci riferiremo nel seguito come al “problema epistemologico”, sembra abbastanza ben posto da meritare la piena attenzione del fisico-matematico; ma se lo si affronta in modo serio, immediatamente se ne percepisce la grande difficoltà. Come potremmo infatti definire con precisione un oggetto complesso quale è una scienza esatta senza conoscerne a fondo la natura?

Man mano che ci si addentra nel problema, appare ragionevole che allo scopo occorra anche una certa preparazione logico-matematica di base, quale non può essere senz’altro presupposta nel nostro lettore medio – e nemmeno in quel lettore “adulto”, dal punto di vista fisico-matematico, che abbiamo delineato nella Presentazione (vedi S.sez. 0.0.1). Proseguiremo dunque la presente Introduzione nel modo quanto più possibile informale. Gli approfondimenti che proporremo non potranno quindi essere considerati come qualcosa di più che un abbozzo di trattazione sistematica; ma per fortuna, i veri temi di questo libro sono altri. Sulla logica matematica di base, il lettore che ne avesse bisogno troverà alcune informazioni nella App. Gen. A.

0.1.2 RAGIONAMENTO ASSIOMATICO-DEDUTTIVO E RAGIONAMENTO INDUTTIVO

Da una teoria fisico-matematica (TFM), lo studioso dei fenomeni naturali, il “fisico” nell’accezione dell’etimo, si aspetta che attribuendo ai suoi termini i loro significati intuitivi nel linguaggio naturale, e analogamente traducendo i suoi enunciati, si ottengano i corrispondenti oggetti empirici, operativamente ben definiti; e in particolare riferendosi ai suoi assiomi, e ai teoremi che da essi conseguono attraverso le regole deduttive, si ottengano verità empiriche verificate o verificabili entro i limiti delle approssimazioni osservative.

Così come è stata informalmente descritta, una TFM risulta **affidabile** (o presuntivamente corretta) in quanto ogni suo teorema, letto nell’interpretazione naturale, sia convalidato dall’esperienza; e complementariamente, risulta **adeguata** (o presuntivamente completa), rispetto al dominio empirico preso in esame, in quanto ad ogni verità osservata in quel dominio si possa far corrispondere un teorema nella interpretazione naturale. È tuttavia chiaro come tale presunta correttezza/completezza di una TFM risulti continuamente esposta alla prospettiva di essere invalidata dal raffinarsi delle nostre capacità osservative-induttive, e in secondo ordine anche assiomatiche-deduttive. Se questo si verifica, anche in un solo caso o classe di casi ragionevolmente accertati, la TFM dovrà essere respinta, e se possibile sostituita con una TFM diversa, confacentesi alla nuova situazione.

In pratica, la possibilità che una TFM consolidata diventi *del tutto* inutilizzabile a seguito del riconoscimento di una sua qualche mancanza di correttezza o completezza è poco probabile: come la storia delle scienze esatte ci insegna, in alcuni casi fondamentali sono bastate *modifiche* – di fatto generalizzazioni, per quanto sostanzialmente non banali –, alla luce delle quali la vecchia TFM risulta compatibile con la nuova in quanto identica ad un particolare “caso-limite” di quest’ultima, continuando ad essere praticamente corretta/completa in condizioni prossime alle condizioni-limite. In altre parole, si è autorizzati a pensare che spesso la nuova e più potente TFM *raffini la vecchia piuttosto che negarla radicalmente*. Questo fatto è talvolta proposto come **principio del carattere cumulativo** delle teorie scientifiche esatte. Ma in generale, al di fuori dei casi di una sua possibile “correzione locale”, la emendabilità di una teoria scientifica invalidata deve essere di massima esclusa. Se infatti un teorema della TFM in questione è riconosciuto come empiricamente falso, o ad una verità empirica non corrisponde un suo teorema (o addirittura esiste un teorema che la nega), di quella TFM occorrerà modificare gli assiomi e (assai più raramente) le regole deduttive. Ora un’operazione di questo tipo avrebbe conseguenze potenzialmente devastanti, e di norma insondabili, sull’identità della TFM in questione; e questo significa appunto che essa dovrà essere respinta in toto, e (auspicabilmente) sostituita con una TFM diversa.

Precisiamo anche – per quanto qui possibile – cosa intendiamo per “correzione locale” di una TFM. Supponiamo che un assioma A della TFM invalidata possa essere sostituito da un corrispondente assioma $A'(c)$ contenente certi parametri c appartenenti ad uno spazio metrico, in modo tale che (i) la TFM con $A'(c \rightarrow c_0)$ in luogo di A sia equivalente alla TFM originale, ad esempio perché $A'(c \rightarrow c_0) \equiv A$, e che (ii) sostituendo A con $A'(c=c^*)$, ove c^* e c_0 sono distinti ma convenientemente vicini nella scala dei valori di interesse, essa risulti corretta/completa alla luce di più precise verità empiriche. In questo caso la vecchia TFM diventa un “limite” (per $c \rightarrow c_0$) della nuova, ed è dunque praticamente corretta “per c abbastanza vicino a c_0 ”.⁸ La fisica matematica ha conosciuto più d’una situazione del tipo descritto (e altre probabilmente ne conoscerà): ad esempio, in occasione della transizione dall’ottica geometrica a quella ondulatoria, oppure dalla dinamica classica a quella relativistica speciale, o ancora dalla meccanica newtoniana a quella quantistica, .. e via dicendo. Qui le teorie di partenza sono tutte limiti (per lunghezza d’onda tendente a zero in rapporto alle lunghezze tipiche, rispettivamente per velocità tipiche tendenti a zero in rapporto alla celerità della luce, rispettivamente per azioni tipiche tendenti a infinito in rapporto all’azione di Planck) di quelle di arrivo.

⁸ Evidentemente, si può qui pensare che la metrica dello spazio dei parametri induca una corrispondente “metrica”, e quindi una topologia, nello “spazio” (per così dirlo) degli enunciati. Qualcosa di simile varrà poco più avanti, quando accenneremo alla “stabilità strutturale” di una TFM; anche in quel caso, lo spazio degli enunciati si dovrà pensare dotato di una topologia indotta da una conveniente metrica.

Il **ragionamento assiomatico-deduttivo** è quello delle teorie matematiche standard (cioè non formalizzate in senso stretto, vedi S.sez. 0.1.3). Dobbiamo peraltro tener conto del fatto che una parte preponderante delle nostre attività mentali è legata al **ragionamento induttivo**. Con un inevitabile grado di approssimazione, l'attributo "induttivo" qualifica un tipo di attività intuitiva e immaginativa, che partendo da asserti isolati e particolari cerca di pervenire (e spesso perviene) a conclusioni *che comunque li superano*. In pratica, un ragionamento induttivo rende plausibile (con vario grado di "credibilità", o "probabilità induttiva" (Carnap (Rudolf, 1891-1970)), una certa conclusione senza implicarla genuinamente. I tentativi di trasportare il ragionamento induttivo in un ambito strettamente "logico" non sembrano aver condotto finora a conclusioni definitive, per cui da parte di molti si contesta l'esistenza di una vera e propria "logica induttiva". Ciò non toglie che il ragionamento induttivo abbia un'importanza capitale nel procedere del nostro pensiero: senza di esso, la stessa matematica non esisterebbe, sia di per sé che come strumento di possibile conoscenza esatta del mondo fenomenico.

D'altra parte la scienza basa la sua fiducia nell'induzione sul successo che essa le ha fin qui assicurato, cioè giustifica l'induzione mediante un'altra induzione. Questa non profondissima scoperta di segno scettico, con le pretese difficoltà che ne deriverebbero, è quello che i filosofi chiamano "circolo vizioso di Hume" (David, Edimburgo, 1711-1776). Un'altra idea di segno scettico è stata accolta e portata alle sue estreme conseguenze da Popper (Karl, 1902-1994) e dalla sua scuola.⁹ Ma fortunatamente per le scienze esatte, disponiamo ormai di sistemi di verità empiriche "previste teoricamente" così ampi e solidi da lasciare allo scetticismo alla Hume (o alla

⁹ Come è presumibilmente noto, Popper rifiuta di ammettere che un sistema di conoscenze S avente per oggetto il mondo fenomenico possa essere comunque convalidato dall'esperienza, e quindi rifiuta l'induzione come base fondativa delle scienze empiriche. Secondo Popper ("Logik der Forschung", Springer (1934); 1^a trad. ingl., "Logic of Scientific Discovery", Hutchinson (1959); trad. it., "Logica della scoperta scientifica", Einaudi (1970), § 4-6 e § 21-22), un asserto T di S ha in generale la natura di un quantificatore universale ($T \equiv \forall(x)P(x)$) su un dominio *non finito*, e quindi non può essere convalidato da un numero necessariamente finito di verifiche. Al più, S potrà essere «corroborato» da conferme sperimentali; e da ciò deriva la provvisorietà e l'incertezza del sapere scientifico, nonché l'impossibilità che esso aspiri a raggiungere verità definitive. Al contrario, una sola istanza negativa x_0 basta ad invalidare T, ammesso di poter refutare x_0 isolatamente, e non in congiunzione con altre x_1, x_2, \dots (Duhem). Nelle notazioni formali standard (qui usate per brevità), ciò avviene perché $\neg P(x_0) \Rightarrow \exists x\{\neg P(x)\} \Leftrightarrow \neg \forall x P(x) \equiv \neg T$; il che induce Popper a proporre come «criterio di demarcazione» tra la "scientificità" e la "non-scientificità" di S la sua «falsificabilità» piuttosto che la sua «verificabilità». Vale a dire, S è falsificabile, e quindi scientifico, se esiste qualche suo enunciato (un suo «falsificatore potenziale») che è in contraddizione con S; ed è «falsificato» se tale falsificatore (che secondo Popper (loc. cit. § 22) sarà «di bassa universalità») è convalidato dall'esperienza. Confessiamo un certo disagio di fronte a queste tesi, che pure hanno avuto un ruolo centrale nella gnoseologia popperiana: la tentazione di associarle al celebre giudizio di Gauss (Carl-Friedrich, Brunswick 1777, Gottinga 1855) sulla distinzione kantiana tra proposizioni analitiche e sintetiche (vedi S.sez. 9.1.1, nota ⁽⁵⁾) è veramente forte. Ciò che più colpisce, agli occhi dell'uomo di scienza normalmente estraneo al mondo della filosofia, è la prolissità (rispetto agli standard logici dell'epoca – siamo nel 1934!) con cui esse tesi sono esposte, *nonché la loro totale ovvietà*; e in senso opposto, le innumerevoli discussioni cui hanno dato vita nella letteratura epistemologica degli ultimi 60/70 anni. Quasi superfluo aggiungere che le idee di Popper hanno avuto un impatto trascurabile sulla grande maggioranza degli scienziati e degli stessi logici, e (ovviamente) sulle loro concrete attività.

Popper) spazi veramente esigui; senza contare che quello scetticismo è *per un verso ineludibile e per l'altro sterile*.¹⁰

La stessa ripetibilità indefinita della capacità predittiva di una TFM affidabile permane inoltre sul piano della *confidenza* (per quanto alta questa possa essere), in quanto un oggetto fisico radicalmente caotico non sarebbe prevedibile da alcuna TFM; detto diversamente, una rappresentazione predittiva di quell'oggetto si deve comunque fondare sulla presunzione di una sua sufficiente *regolarità*. Se questo è il caso, la regolarità dell'oggetto si traduce in una sua rappresentazione di natura *invariabilmente matematica*. Perché? Vi è una varietà di posizioni che interpretano e giustificano questa connessione tra il mondo fisico e quel particolare prodotto della umana immaginazione che è appunto la matematica. Ne menzioniamo le due forse più accreditate, e in certo senso antitetiche. 1): Senza rendersene conto, l'uomo crea e seleziona proprio quella matematica che più si adatta al mondo fisico (risposta "realista"); 2): la fisica è così ben descritta dalla matematica perché «il mondo è del tutto matematico nella sua ultima sostanza, isomorfo ad una struttura matematica che l'uomo viene scoprendo a poco a poco, bit per bit.» (risposta "idealista", nella forma recentemente datale dal cosmologo M. Tegmark (2007)).

È impossibile, a questo punto, non menzionare il notissimo articolo "The Unreasonable Effectiveness of Mathematics in the Natural Sciences"¹¹ di Wigner (Eugen, 1902 - 1995), che nel 1960 ha riproposto in chiave interrogativa lo stesso tema, già vivo da tempo. Possiamo aggiungere che alla risposta a suo modo trascendente offerta da Wigner (vedi la precedente nota) si arriva quasi obbligatoriamente seguendo l'idea che la matematica consista di mere manipolazioni su segni privi di significato; ma che se questa idea si supera nel modo corretto, allora «l'irragionevole efficacia della matematica può diventare ragionevole» (F. Browder e S. Mac Lane, in *Mathematics Today*, 1978).

Che la matematica sia in pratica sempre presente nei modelli predittivi efficaci del mondo fenomenico non è materia di opinione, ma un fatto incontestabile, confermato da innumerevoli successi. La sola barriera che ci separa dalla immensa regione della conoscenza non dominabile da

¹⁰ Una semplice metafora può meglio illustrare la situazione. Sotto certi aspetti, l'obiettivo di una TFM può essere paragonato a quello di decriptare una successione potenzialmente infinita di messaggi cifrati. È allora ovvio che non potremo mai esser certi che il messaggio successivo sarà decriptato come i precedenti (pur nella confidenza che la chiave di codifica non sia cambiata); ma questo fatto non può e non deve farci dubitare della legittimità ed efficacia dell'uso del decodificatore. Un'altra nota metafora, più o meno equivalente sotto il presente punto di vista, accosta l'attività scientifica esatta al tentativo di comprendere le regole di un gioco competitivo leale – ad esempio, non contando le differenze di complessità e di accessibilità, del gioco degli scacchi – assistendo ad un numero potenzialmente infinito di partite. Ci si può allora sempre aspettare che la partita successiva si riveli incomprensibile alla luce delle regole fino a quel momento presunte valide.

¹¹ *Comm. Pure Appl. Maths.*, **13**, 1 (1960). Nelle parole di Wigner: «It is difficult to avoid the impression that a miracle confronts us here, quite comparable in its striking nature to the miracle that the human mind can string a thousand arguments together without getting itself into contradictions, or to the two miracles of laws of nature and of the human mind's capacity to divine them.»

questo strumento è quella della complessità (non necessariamente anche in senso formale ¹²) del modello; la quale spesso implica l'instabilità delle (pseudo)risposte che esso ci propone, o addirittura la loro inesistenza. Ma questo non può essere rifiutato, almeno in generale. Non contando una possibile insufficiente abilità a “modellare” in senso matematico fenomeni complessi, osservabili con difficoltà e presuntivamente non regolari, non c'è infatti niente di assurdo nella incapacità di calcolare efficacemente entro una data teoria matematica.

La conclusione alla quale arriviamo sembra dunque essere la seguente. Attraverso i suoi lenti e misteriosi meccanismi, la selezione naturale avrebbe prodotto una specie, quella di homo sapiens, capace di inventare modelli predittivi *di certe particolari sezioni del mondo*, la natura dei quali si rivela sostanzialmente “matematica”; essendo inteso che quella natura matematica si accorda con la matematica inventata dalla specie stessa. Quindi homo sapiens avrebbe inventato la “matematica giusta”; e ciò non può avere compiuto se non ispirandosi alla sua esperienza del mondo reale. Ma anche così sorge una difficoltà: alcune fondamentali predizioni scientifico-esatte ¹³ suggeriscono infatti che mentre l'attenzione del sapiens si dirige verso regioni empiriche sempre più remote e difficili da esplorare, e delle quali egli certamente non aveva fatto alcuna esperienza nel corso della sua storia evolutiva, può succedere che la sua matematica funzioni!

La tesi di Tegmark si configura insomma come una sorta di nuovo platonismo, in cui al mondo delle idee si sostituisce il mondo reale, e alle idee la regolarità di questo, per quanto sempre più nascosta. In breve, la matematica del sapiens pre-esisterebbe nella regolarità ultima del mondo. Ora se è vero, come è vero, che non esiste linguaggio più efficace di quello matematico per descrivere ciò che è strutturalmente regolare, allora il presunto “miracolo” segnalato da Wigner si spiega, anzi diventa quasi inevitabile. Il problema si sposta peraltro, in questo modo, sull'interrogativo: “perché il mondo è regolare – e regolare così e così – nella sua realtà ultima?” Al quale una risposta accettabile per quanto audace potrebbe essere: “perché un mondo non regolare – e non regolare così e così –, sempre nella sua realtà ultima, non potrebbe esistere” (si cominciano a recepire segnali in questo senso).

Ma possiamo ormai chiudere su questo problema citando le parole che su di esso esprimeva Einstein (Albert, Ulm Germ. 1879, Princeton N.J. 1955; da una lettera all'amico M. Solovine) ben prima della pubblicazione del lavoro di Wigner, e come sempre ispirate al più profondo buon senso: «Ciò che dovremmo aspettarci a priori è un mondo caotico del tutto inaccessibile al pensiero. Ci si potrebbe (o piuttosto ci si dovrebbe) aspettare che il mondo sia governato da leggi soltanto nella

¹² L'esempio che per primo si cita in proposito è quello del modello meteorologico di Lorenz, consistente in un sistema differenziale ordinario non lineare che può essere scritto in poche righe, e che in certe condizioni genera soluzioni caotiche, praticamente imprevedibili.

¹³ La previsione dell'esistenza del positrone da parte di Dirac (1930-31) è un esempio classico, relativamente recente, a questo proposito.

misura in cui interveniamo ad ordinarlo mediante la nostra capacità ordinatrice: qualcosa di simile all'ordine alfabetico del dizionario. Invece il tipo di ordine creato ad esempio dalla teoria della gravitazione di Newton ha tutt'altra natura. Sebbene gli assiomi della teoria siano imposti dall'uomo, il successo di una tale costruzione *presuppone un alto grado di ordine del mondo oggettivo* (corsivo d.r.), cioè qualcosa che non siamo per nulla autorizzati ad aspettarci a priori. È precisamente questo il “miracolo” che sempre più si manifesta con lo sviluppo delle nostre conoscenze.»

Dopotutto, è proprio la ricerca delle regolarità del mondo dei fenomeni il più grande incentivo che ci sollecita alla scienza, e il loro scoprirle/verificarle/prevederle la più grande gratificazione che ne deriviamo; al punto che si potrebbe proporre l'equivalenza informale “ricerca scientifica” ≡ “ricerca di regolarità non banali del mondo fenomenico”. Verificare le regolarità del mondo significa innanzitutto constatare che da esperimenti identici si ricavano (o che situazioni identiche generano) risultati identici. In non rari casi, tuttavia, esperimenti/situazioni *quasi* identici/identiche (in una conveniente accezione) danno luogo a risultati *molto* diversi, e possibilmente diversi in modo disordinato e imprevedibile. Questi casi sono detti “critici”; ma spesso le TFM che se ne occupano prevedono il carattere caotico dei risultati, cogliendone aspetti regolari nascosti, che riflettono la semplicità delle loro autentiche sorgenti.¹⁴ Si aggiunge così una terza ragionevole proprietà da richiedere ad una TFM (oltre alla sua affidabilità/adequatezza): quella della sua **stabilità strutturale** (R. Thom, 1923-2002) rispetto a “piccole perturbazioni” (sempre in una conveniente accezione). La stabilità strutturale di una TFM è di solito assai difficile da dimostrare, e si è così costretti a convivere con la semplice fiducia che essa sia verificata; più o meno come si convive con la fiducia che la teoria assiomatica degli insiemi (mettiamo) sia coerente. D'altra parte la proposta di Thom di accettare la stabilità strutturale di una TFM come un suo *assioma* specifico suscita non poche perplessità.

Continuando ad esprimerci in modo informale, diremo “solida” una TFM della quale un insieme adeguatamente ampio di teoremi sia stato convalidato entro le precisioni osservative. Al presente, le TFM correnti sono in buona parte molto, o accettabilmente, solide. Cionondimeno, vi sono situazioni in cui una TFM non può ancora giudicarsi abbastanza solida, né deve essere rifiutata, alla luce di esperimenti *oggi* realizzabili; pur nella prospettiva che essa possa ben raccordarsi, da un punto di vista concettuale, con altre TFM solide. Sembrerebbe allora conveniente o necessario accettare quella TFM come “valida fino a prova contraria”, eludendo così i dubbi alla Hume o alla Popper in modo anche più radicale di quello correntemente praticato. Ad esempio, i presenti “abbozzi di teorie” della gravità quantistica sono almeno parzialmente in questa

¹⁴ Si pensi ad esempio alla semplicità formale della cosiddetta “mappa logistica”, che in certe condizioni genera una evoluzione imprevedibile, sebbene completamente determinata. È questo un caso canonico di “caos deterministico”.

condizione. Per quanto possiamo aspettarci, una loro possibile selezione su base sperimentale potrà infatti (forse!) ottenersi soltanto mediante procedure molto indirette, essendo le procedure dirette fuori della portata degli strumenti osservativi oggi disponibili, o ragionevolmente immaginabili. D'altra parte, un argomento in favore della massima apertura di principio verso i tentativi di estendere le nostre conoscenze in questa e altre simili direzioni è il fatto che alcune delle attuali TFM fondamentali sono manifestamente *incomplete* rispetto ai loro oggetti empirici già oggi osservabili: ad esempio, il cosiddetto "modello standard" della fisica delle particelle elementari non è in grado di prevedere i valori di un discreto numero di parametri, che devono esservi introdotti ricavandoli dall'osservazione.

0.1.3) MATEMATICA SIGNIFICANTE VS. MATEMATICA NON-SIGNIFICANTE ¹⁵

Evidentemente, le argomentazioni delle sottosezioni precedenti presuppongono che ci si riferisca ad una accezione "quieta" della matematica, nella veste di sistema assiomatico-deduttivo (apparentemente) libero da problemi fondativi. Questo non fu più completamente giustificato a partire dalla seconda metà del secolo XIX, con l'aprirsi della cosiddetta e ben nota "crisi dei fondamenti" che assalì il pensiero matematico con insidiosi (auto)interrogativi. Se la scoperta delle geometrie non euclidee aveva scosso la fede nella corrispondenza tra le leggi del pensiero matematico e quelle dell'universo fisico, anche la convinzione della coerenza dei principi matematici e dello stesso metodo assiomatico-deduttivo era destinata ad essere messa in questione. Alcuni di questi interrogativi hanno trovato nel frattempo risposta, segnando un indiscutibile progresso della disciplina. In altri casi questo non è avvenuto, e a poco a poco ha generato quella che qualcuno ha chiamato la "perdita della certezza matematica". ¹⁶

Tra i tentativi di (auto)riparazione si devono innanzitutto menzionare quello logicista (Russell, Whitehead), quello intuizionista (Brouwer, Weyl) e quello formalista (Hilbert, Bourbaki). Particolarmente efficace e ricco di prospettive doveva rivelarsi quest'ultimo, con l'idea fortemente propugnata da Hilbert (David, Königsberg 1862 - Gottinga 1943, probabilmente il massimo matematico tra i due secoli), e peraltro non inedita, di configurare un sistema assiomatico-deduttivo come un gioco sequenziale astratto su "segni" privi di significato. ¹⁷ Tuttavia, con il formalismo hilbertiano si è aperta una divaricazione senza precedenti tra "significante" e "non-significante" in

¹⁵ In questa sottosezione (e oltre) sono talvolta introdotti oggetti che non vengono definiti in modo preciso ed esaustivo. Essi non sono quindi scritti in grassetto al loro primo apparire, ma soltanto posti tra " (da non confondere con le " , ")".

¹⁶ È questo il titolo di una lucida monografia dello storico della matematica M. Kline ("Mathematics: The Loss of Certainty", Oxford Un. Press 1980, trad. italiana in Mondadori, 1985).

¹⁷ Vedi, per cominciare, D. Hilbert: "Axiomatisches Denken", in *Mathematischen Annalen* **78**, 146-156 (1918).

matematica. Lo studio approfondito di questo iato ha dato origine alla “teoria dei Modelli (maiusc.)”¹⁸, zoccolo duro della moderna logica-matematica e direttamente afferente al nominato problema epistemologico (cfr. S.sez. 0.1.1). La teoria dei Modelli ha sfiorato il mondo della matematica quotidiana, sostanzialmente ispirata alla teoria descrittiva (altrimenti detta “intuitiva” o “ingenua”) degli insiemi; e a maggior ragione, è rimasta praticamente estranea alla ricerca fisico-matematica e fisico-teorica.

Si è così venuta a creare una strana e complessa situazione, nella quale gioca un ruolo di grande rilievo la cosiddetta “Informatica Astratta” (o “Computer Science”). Infatti la fisica teorica si è sempre più matematizzata (lo è ormai del tutto), sebbene la matematica alla quale si richiamano quasi senza eccezione i suoi modelli resti quella quotidiana-intuitiva: un terreno sul quale evidentemente non può muoversi un cosiddetto “automa sequenziale”. D'altra parte, la formalizzazione stretta della matematica intuitiva, cioè la sua traduzione in un linguaggio adeguato a interloquire con l'automa (in quel linguaggio, l'automa ci deve “capire” e noi dobbiamo “capire” lui), solleva micidiali interrogativi intorno al linguaggio stesso, ai quali l'automa non è in grado di rispondere “sì” aut “no” in un numero finito di passi, almeno in generale.

Un nuovo soggetto, la cui precisa natura ontologica abbiamo per un momento affidato all'intuizione, si è così affacciato al nostro discorso. Cosa è un automa? La definizione non può avere altro carattere che quello di una tesi a priori, e può esprimersi come segue: un automa (sequenziale) è il soggetto di una successione discreta di azioni, governata da regole inalienabili e disponibili passo dopo passo, a partire da una prescritta condizione iniziale compatibile. Si è convenuto che la successione di azioni in oggetto dovesse consistere nel calcolo di un certo tipo di funzioni “numeriche”, cioè definite sui, e aventi per valori, numeri naturali (\equiv interi non negativi). Come è stato accertato, le funzioni (numeriche) cosiddette “ λ -definibili” secondo Church-Kleene (Alonzo, 1903-1995; Stephen, 1909-1994), le funzioni cosiddette “ricorsive generali” secondo Gödel (Kurt, Brno 1906 - Princeton N.J. 1978), e le funzioni cosiddette “computabili” secondo Turing (Alan, 1912-1954) sono le stesse funzioni. Secondo la “tesi di Church” (1936) è allora ragionevole definire come un automa il soggetto capace di calcolare (i valori di) queste funzioni numeriche, partendo dai loro argomenti e con affidabilità idealmente totale, e niente altro.¹⁹

È facile, e didatticamente remunerativo, descrivere a grandi linee un linguaggio “strettamente formalizzato” L , o sistema formale (SF). Si parte con un insieme per lo più

¹⁸ Dai suoi esordi all'inizio degli scorsi anni '20, la teoria dei Modelli si è largamente sviluppata grazie all'opera di logici come A. Tarski, A. Robinson, H. Keisler, C. Chang, K. Gödel ... ecc., trovando una accettabile sistemazione almeno \approx quaranta anni dopo.

¹⁹ Tra i grandi logici del secolo XX, quello che più si sforzò di raggiungere una definizione operativa di automa fu probabilmente Turing. La condizione che l'automa operi in un numero *finito* (ma non necessariamente limitato a priori) di passi è intuitivamente ovvia, e verrà sempre presupposta nel seguito.

numerabilmente infinito di segni primitivi distinti che diremo "alfabeto" di L.²⁰ Con questi segni, seguendo regole prescritte si formano successioni *finite*, o "stringhe", di segni che si dicono "parole" (di vari tipi distinti); e similmente seguendo regole prescritte, si formano stringhe di parole che si dicono "frasi" (di vari tipi distinti). Di norma, le parole si scrivono da sinistra a destra e le frasi dall'alto in basso. Come non sarebbe necessario in linea di principio, possiamo interrompere qui questa gerarchia. Le regole per la formazione di parole costituiscono la "morfologia" di L, e quelle per la formazione di frasi la sua "sintassi". Un tipo particolare di parola, specificato dalla morfologia di L, è un suo "enunciato" (detto anche "formula chiusa"); e un tipo particolare di stringa di soli enunciati, specificato dalla sintassi di L, è una sua "dimostrazione". Gli enunciati di una dimostrazione si dicono "Teoremi" (maiusc.), e si dividono nelle classi distinte degli "assiomi" e dei "teoremi" (minusc.). Le definizioni di questi oggetti di L sono tali che il nostro automa è in grado di distinguere se una stringa di segni è una parola (o no), se una parola è un enunciato, se un enunciato è un assioma e se una frase che sia una stringa di enunciati è una dimostrazione. Pur essendo le parole [le frasi] successioni finite per definizione, l'insieme delle parole [delle frasi] di L ha cardinalità numerabilmente infinita (\aleph_0), perché non si pone alcun limite a priori alla loro lunghezza (\equiv numero degli elementi di cui esse si compongono). Questa cardinalità \aleph_0 è all'origine di questioni intorno a L alle quali un automa non è in grado di rispondere con un sì aut con un no (in un numero finito di passi), ed eleva L al disopra del livello di un gioco combinatorio finito.

Se l'alfabeto, le regole morfologiche/sintattiche e gli assiomi di L soddisfano certe condizioni progressivamente più stringenti, L si dice "logico"²¹, "logico-quantificato", "logico-quantificato-egualitario", "logico-quantificato-insiemistico", ... e così via. Correntemente, un linguaggio logico si dice anche "calcolo proposizionale", e un linguaggio logico-quantificato anche "calcolo predicativo del 1° ordine" (o "LPC", Lower Predicate Calculus)²². In genere, ci si occupa di linguaggi che siano almeno logici. Nell'alfabeto di qualunque linguaggio L di qualche interesse figura il segno \neg , che si dice "non", e che premesso ad un generico enunciato A produce un altro enunciato $\neg A$ che si dice "negazione di A". Se L è logico, tra due suoi enunciati può sussistere una

²⁰ A parte l'insieme delle cosiddette "variabili", di norma numerabilmente infinito, l'alfabeto consta di un numero assai ristretto di segni; ma parole e frasi che mimano nozioni apparentemente elementari constano tipicamente di stringhe di un numero enormemente maggiore di quei segni. Per fare un esempio significativo: affinché un automa possa gestire senza errori la (apparentemente) semplice nozione matematica di "1" (cardinale) occorrerebbe trasmettergliela, nell'alfabeto elementare (\equiv l'alfabeto meno l'insieme dei segni delle variabili) usato da Bourbaki (vedi "Théorie des ensembles", E III 24), mediante una stringa di parecchie decine di migliaia di segni, soltanto sette dei quali sono diversi. È chiaro che il prodotto del numero dei segni dell'alfabeto elementare per la lunghezza della generica stringa è approssimativamente costante a parità di contenuto di informazione. Su questa base, sarebbe anche possibile stimare l'ordine di grandezza dell'informazione contenuta nella nozione di "1" cardinale, espressa in bit.

²¹ Questo attributo si riferisce qui tacitamente ad una logica a due valori (vero, falso).

²² L'espressione "del 1° ordine" si riferisce al fatto che la quantificazione è limitata ai "termini" che sono oggetto dei predicati, e non supera questo livello (cioè non vi è quantificazione sui predicati).

specifica relazione riflessiva, simmetrica e transitiva (cioè una specifica relazione di equivalenza) che si dice di "bicondizionalità". Se due enunciati sono in tal senso equivalenti, le regole di L permettono di scambiarli nei loro ruoli deduttivi. Le definizioni sono allora tali che per qualunque enunciato A, A e $\neg\neg A$ sono in relazione di bicondizionalità.

L si dice (sintatticamente) "coerente" ["completo"] se *non* esiste un suo enunciato A tale che *entrambi* gli [almeno uno degli] enunciati A, $\neg A$, siano Teoremi [sia un Teorema], e "incoerente" ["incompleto"] in caso contrario. Evidentemente, L può essere coerente e completo, coerente e incompleto, incoerente e completo, ma non incoerente e incompleto. Se tuttavia L (logico) è incoerente (cioè un enunciato A tale che A e $\neg A$ siano Teoremi *esiste*), allora – si dimostra in base alle sue regole sintattiche (assiomi più regole deduttive) – tutti i suoi enunciati sono Teoremi, e L perde per questo ogni interesse.

Un enunciato A che sia un Teorema si dice "deducibile" (o "provabile"); se poi $\neg A$ è deducibile, si dice anche che A è "refutabile". Aggiungendo (*ceteris paribus*) assiomi ad un L incompleto [ad un L coerente], L può trasformarsi in un L' completo [in un L' incoerente]. Sia L incompleto e sia A un suo enunciato per il quale né A né $\neg A$ sono Teoremi, o suo enunciato "indecidibile". Supponendo L anche coerente, aggiungiamo il suo enunciato indecidibile A (oppure $\neg A$) ai suoi assiomi: L si trasforma allora in un L' ancora coerente, ma possibilmente completo. Se questo non succede, cioè se L' è ancora incompleto, e la situazione si riproduce tale e quale con il nuovo L', e così via indefinitamente, allora l'L originario si dice "essenzialmente incompleto".

Diamo ora alcuni esempi di questioni intorno a L alle quali un automa non può in generale rispondere sì aut no in un numero finito di passi. Il primo esempio è quello di un enunciato generico che *non* sia un assioma (in questo caso l'automa lo riconoscerebbe infatti come tale) del quale l'automa dovrebbe *decidere* se è, aut non è, un teorema.²³ A tale scopo, infatti, l'automa dovrebbe esaminare un numero *non prevedibile* di dimostrazioni prima di possibilmente trovarne una nella quale figura l'enunciato in oggetto. (Se comunque l'automa *trova* una tale dimostrazione durante la sua ricerca, allora l'enunciato è un teorema per definizione.) Sottolineiamo anche che un enunciato che non sia un assioma è necessariamente un teorema aut non lo è; quindi la risposta all'Entscheidungsproblem esiste unica, sebbene l'automa non possa provvederla in generale. Ed ecco altre due questioni irrisolvibili "per automata" (1) "è L coerente (aut incoerente)?" (problema della coerenza); (2) "è L completo (aut incompleto)?" (problema della completezza). Un altro problema ancora è quello della "indipendenza degli assiomi": supposto L coerente, immaginiamo di sottrargli l'assioma A, con il che L diventa L', e domandiamoci: "è A un teorema di L'?"

²³ Per ragioni storiche questo problema si dice appunto, comunemente e in ogni lingua, Entscheidungsproblem.

Ovviamente, questo è di nuovo un Entscheidungsproblem. Talvolta, un problema di indipendenza ha avuto grande importanza storica: basti pensare all'indipendenza dai rimanenti assiomi dell'assioma delle parallele nella geometria euclidea.

Il problema della coerenza di un linguaggio (linguaggio che a questo punto possiamo anche chiamare "teoria matematica formalizzata") era in prima linea tra gli interessi logico-matematici di Hilbert, e tale rimase per tutta la sua vita, fino agli ultimi giorni. Dopo aver sempre meglio formulato il suo programma, per tutti gli anni venti Hilbert nutrì grande fiducia sulla possibilità di dimostrare, via formalizzazione, e con quei mezzi minimi che aveva chiamato "processi finitari" (finite Prozesse – essenzialmente costruttivi²⁴), la coerenza delle teorie matematiche correnti.

Di fatto, la formalizzazione aveva precedenti storici degni del massimo credito: nel lontano passato, passi di progressiva importanza in senso assiomatico-deduttivo erano stati compiuti da Pitagora e da Aristotele (384-322 a.C.), aprendo la strada verso Euclide (330? - 275? a.C.): il primo a cimentarsi con il tentativo di assiomatizzare l'unica "fisica" dell'epoca, la geometria. Tuttavia la formalizzazione *in senso stretto* prese piede soltanto nell'Ottocento avanzato: per cominciare, con l'"Analisi matematica della logica" di Boole (George, 1815-1864),²⁵ poi con la fondamentale "Scrittura di concetti" di Frege (Gottlob, 1848-1925),²⁶ e poi ancora con l'"Algebra della logica" di Peirce (Charles, 1839-1914),²⁷ con il "Formulario" di Peano (Giuseppe, 1858-1932),²⁸ e naturalmente con Hilbert e la sua scuola.

Al Congresso Internazionale di Matematica di Parigi del 1900 (ma avrebbe ripreso la stessa idea in molte occasioni fino alla fine degli anni '20), Hilbert aveva orgogliosamente affermato: «Dentro di noi sentiamo il continuo richiamo: c'è un problema, cerchiamone la soluzione. Possiamo trovarla con la sola ragione, perché in matematica non c'è posto per gli "ignorabimus".» Ed effettivamente i formalisti, dei quali Hilbert era leader indiscusso, avevano compiuto progressi significativi nelle direzioni da lui auspiccate. Nel 1921 erano state provate la coerenza e la completezza del calcolo proposizionale (E. Post, 1921²⁹), il passo più semplice sulla via della

²⁴ Hilbert aveva un'idea più o meno intuitiva dei suoi metodi finitari. Più tardi (1934), nei "Grundlagen der Mathematik" scritti in collaborazione con P. Bernays, precisò cosa dovesse intendersi per "finitismo" nei termini seguenti: «Useremo l'attributo finitario (finitist) per indicare che la discussione, asserzione, o definizione in questione si mantiene entro i limiti della completa producibilità degli oggetti e della completa praticabilità dei processi, e può pertanto essere effettuata entro il dominio della verifica concreta.»

²⁵ G. Boole: "The mathematical Analysis of Logic", Cambridge (1847).

²⁶ G. Frege, "Begriffsschrift, eine arithmetischen nachgebildete Formelsprache des reinen Denkes", Halle (1879). L'importanza dell'opera di Frege rispetto al problema di (possibilmente) formalizzare un linguaggio non potrà mai essere sovrastimata.

²⁷ C.S. Peirce, "On the algebra of logic", Amer. Journ. of Math., t. III, 49 (1880), "On the algebra of logic", Amer. Journ. of Math., t. VII, 190 (1884).

²⁸ G. Peano, "Formulaire de mathématiques" 5 voll., Torino (1895-1905).

²⁹ E. Post: "Introduction to a General Theory of Elementary Propositions", Am. Journ. Math., 43, 163 (1921).

formalizzazione di tutta la matematica. Nel 1930, con la sua tesi di laurea³⁰ K. Gödel aveva poi dimostrato la completezza del calcolo predicativo del 1° ordine, un'impresa ben più ardua della precedente (era nella famosa lista hilbertiana dei ventitrè “problemi aperti” a Parigi); e più tardi, lo stesso Hilbert (con altri) ne avrebbe provato la coerenza, nonché l'indipendenza degli assiomi. Questi ed altri risultati entusiasmarono i formalisti, che erano ormai quasi certi che la sintassi deduttiva, o “teoria della dimostrazione” (Beweistheorie) prima o poi sarebbe riuscita a provare la coerenza e la completezza di tutte le teorie matematiche correnti.

Soltanto alcuni scettici potevano prevedere la catastrofe che di lì a poco si sarebbe abbattuta su questo programma. Proprio l'anno successivo, lo stesso Gödel pubblicò infatti una sua memoria³¹ contenente due sorprendenti risultati ottenuti mediante strumenti di inquestionabile legittimità, e che si possono presentare come segue:

T1. (“teorema di incompletezza” propriamente detto): «Se una teoria formalizzata T abbastanza potente da esprimere l'aritmetica dei naturali è coerente, allora essa è essenzialmente incompleta.» Altrimenti detto, il prezzo della coerenza (per quella T) è l'incompletezza essenziale;

T2. (corollario del precedente, talvolta detto anche “secondo teorema di incompletezza”) «Non è possibile dimostrare la coerenza della sopraddetta teoria formalizzata T servendosi dei soli strumenti ammessi da Hilbert, i nominati metodi finitari.» Infatti, Gödel costruisce l'enunciato aritmetico A che rappresenta l'enunciato metamatematico “l'aritmetica dei naturali è coerente”, e dimostra che A *implica* un enunciato indecidibile G di T. Se A fosse provabile in T, lo sarebbe anche G; ma poiché G non lo è, allora anche A non è provabile in T. In effetti, questo risultato afferma anche di più di quanto appena asserito, perché stabilisce che è impossibile dimostrare la coerenza dell'aritmetica, e a maggior ragione di qualsiasi teoria abbastanza potente da esprimere l'aritmetica, servendosi di metodi “traducibili” nel sistema aritmetico.^{32, 33}

³⁰ K. Gödel: “Die Vollständigkeit der Axiome des logischen Funktionenkalküls”, Mon. Math. Phys., **37**, 349 (1930).

³¹ K. Gödel, “Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I”, Mon. Math. Phys. vol 38, 173 (1931). Di questo epocale lavoro esistono traduzioni (e anche *diverse* traduzioni) praticamente in tutte le lingue importanti (per l'italiano, vedi in E. Agazzi, “Introduzione ai problemi dell'assiomatica”, Vita e Pensiero (1962)). Per le versioni in inglese, vedi: “On Formally Undecidable of «Principia Mathematica» and Related Systems”, Basic Books (1965), ristampato da Dover; oppure in J. van Heijenoort, “A Source Book in Mathematical Logic, 1879-1931”, Harvard Un. Press (1967); o ancora in K. Gödel, “Collected Works” (a cura di S. Feferman), vol I, Publications 1929-1936, Oxford Un. Press (1986).

³² Quanto sopra vale altrettanto sostituendo “refutabile” a “provabile” e “refutare” a “dimostrare”. Si noti che “refutare la coerenza” equivale a “dimostrare l'incoerenza”.

³³ Secondo il secondo teorema di incompletezza, la possibile coerenza del SF considerato non può essere accertata mediante metodi formalizzabili all'interno di quel SF; che è come dire, mediante la più ampia metateoria leale. Una eventualità di questo tipo – che è stata pittorescamente paragonata a quella di «tirarsi fuori dalla palude afferrandosi per il proprio codino», alla maniera del barone di Münchhausen – è infatti la più permissiva entro la quale una prova di coerenza si configurerebbe come ancora del tutto leale. Dobbiamo dunque accettare di convivere al più con la *fiducia*, e non con la *certezza* (ottenuta con mezzi leali) che un SF abbastanza potente da ricevere il nostro pieno interesse sia coerente. Una teoria matematica che sia “espressione semantica” di un SF semplicemente *supposto* coerente può in realtà portare a contraddizioni, e può essere usata soltanto a rischio che questo si verifichi da un momento all'altro. D'altra parte, se la coerenza di quel SF non può dimostrarsi “al suo interno”, essa può invece dimostrarsi “dall'esterno”,

Non meraviglia che Gödel abbia raggiunto queste conclusioni – che segnano una svolta storica nei rapporti tra matematica, logica ed epistemologia – mediante una “realizzazione significativa” di T; affacciandosi così alla sua “semantica”, e inaugurando quel nuovo corso della logica matematica le cui implicazioni, attraverso la nominata teoria dei Modelli, arrivano quasi ai nostri giorni. Ovviamente l’impatto dei teoremi di incompletezza di Gödel (1931) sulla logica matematica è stato enorme: esso ha drammaticamente allontanato la matematica formalizzata dalla matematica intuitiva, rivelando uno spazio tra l’una e l’altra che si confidava non esistere. Per effetto del teorema di incompletezza, delle due combinazioni di proprietà che un SF “interessante” può avere (coerenza-completezza, coerenza-incompletezza), soltanto la seconda sopravvive. Ciò non toglie che esistano SF abbastanza semplici, e tuttavia fondamentali, dei quali si può provare sia la coerenza che la completezza: come abbiamo visto, uno è il calcolo proposizionale (Post 1921), e l’altro quello dei predicati (Gödel 1930), possibilmente con (relazione di) uguaglianza.

Se tuttavia ci chiediamo quanto i teoremi di incompletezza abbiano in qualche modo influenzato lo sviluppo della matematica intuitiva negli ultimi ottanta anni, la sola risposta onesta è “poco o nulla”. Questo non può renderci più perplessi di un tanto. Infatti, i teoremi di incompletezza hanno “soltanto” rivelato due nuove e inattese verità:

- 1) sul fronte della coerenza: la possibile ma probabile assenza di contraddizioni nella maggior parte delle teorie matematiche intuitive non può essere provata *al loro interno*, mediante strumenti completamente leali;
- 2) sul fronte della completezza: l’assenza di risposte del tipo “sì-no” a certi interrogativi “resistenti” (ma legittimi) che nascono in teorie matematiche intuitive forse rispecchia la loro indecidibilità nell’ambito delle versioni formalizzate di quelle teorie.

D’altra parte, e come era prevedibile, queste verità non hanno distolto i matematici che ne erano consapevoli dal loro operare: essi (come pure quelli tra loro che le ignoravano) hanno continuato a fare matematica intuitiva, allargando così sempre più la già enorme sfera delle loro

cioè richiamandosi allo scopo ad un SF più forte di quello di partenza, a sua volta supposto coerente. Ma oltre a costituire un gioco sleale, una tale possibilità dà origine ad un rinvio infinito ad altri SF. Le sole vie di uscita da questo dilemma sembrano dunque essere quelle di (i) rassegnarsi ad usare SF potenzialmente incoerenti, e certamente incompleti se coerenti, oppure (ii) limitarsi a considerare SF così poveri che la prova della loro (possibile) coerenza sia ottenibile mediante strumenti elementari, sulla cui legittimità non vi sia nulla da eccepire (ad esempio i metodi finitari di Hilbert). Sfortunatamente, la seconda opzione è troppo restrittiva. Questo è forse il risultato più suggestivo, ed a suo modo di meno difficile comprensione intuitiva, tra quelli ottenuti da Gödel. Quando nel 1975 (quindi ben 35 anni dopo il suo arrivo all’Institute for advanced Studies di Princeton, transfuga da Vienna) la locale università si decise a conferirgli un dottorato onorario, lo accompagnò con la seguente motivazione ufficiale: «La sua analisi rivoluzionaria dei metodi di dimostrazione usati nella branca più familiare ed elementare della matematica, l’aritmetica dei numeri naturali, ha scosso le fondamenta della nostra comprensione, sia dal punto di vista della mente che nell’ambito di applicazione del suo strumento preferito – il metodo assiomatico. Come tutte le rivoluzioni, essa non ha soltanto evidenziato i limiti dei vecchi metodi, ma si è anche dimostrata una ricca sorgente per quelli nuovi, lasciando fiorenti discipline nella sua scia. La logica, la matematica e la filosofia continuano a trarre immensi benefici da quanto il suo genio ha prodotto.» Per qualche semplice delucidazione sui teoremi di incompletezza, si veda l’App. Gen. A.

conoscenze specifiche. Anche più impercettibile, ovviamente, è stata l'influenza dei teoremi di incompletezza sulle teorie fisiche, che come già detto "nutrono" i loro modelli di matematica intuitiva. Quasi universalmente, si presume ormai infatti che le teorie fisiche (intuitive) siano isomorfe a corrispondenti teorie matematiche (intuitive). Ovvero, una particolare sezione del mondo fenomenico appare agli occhi del fisico teorico come una corrispondente struttura matematica appare agli occhi del matematico: l'"interpretazione" che porta gli enunciati dell'una a corrispondenti enunciati dell'altra deve essere una *biiezione* I rispettosa dei vincoli logici. Per fare un esempio: se \mathcal{F} è un enunciato di una TFM e $A \equiv I(\mathcal{F})$ è l'enunciato matematico che gli corrisponde attraverso I , allora $\text{non-}A \equiv \text{non-}I(\mathcal{F})$ deve avere lo stesso valore di verità di $I(\text{non-}\mathcal{F})$ (o similmente scambiando \mathcal{F} con A e I con I^{-1}). Quindi la TFM in oggetto e la teoria matematica intuitiva che le corrisponde attraverso I possono essere identificate. (Per contro l'analisi della situazione sarebbe ben altra cosa se considerassimo il successivo passaggio dagli enunciati della matematica intuitiva a quelli della matematica strettamente formalizzata.)

Resta da chiedersi – ma la risposta è a questo punto abbastanza ovvia – cosa spinga i logici matematici a ricercare la formalizzazione stretta delle teorie matematiche, cioè la traduzione di queste in un conveniente linguaggio segnico sequenziale non-significante. Il punto è che il livello dei segni di un alfabeto – segni di varia natura a seconda del medium utilizzato³⁴ – di una teoria matematica strettamente formalizzata, o sistema formale, è l'ultimo al di là del quale *non è più possibile arretrare*: esso offre un massimo di concretezza ed affidabilità, quella tipica dell'attività di un automa. Via formalizzazione stretta, un tale automa potrebbe *enumerare* le verità della teoria matematica di cui ci occupiamo, e in particolare quelle relative alla matematica associata alla nostra TFM. Tuttavia, il perseguimento di un tale obiettivo, e con esso quello di possibilmente ridurre la spiegazione ultima dei meccanismi che governano una data sezione del mondo fenomenico alla esecuzione di convenienti algoritmi³⁵, presenta due punti deboli:

³⁴ Una enorme varietà di segni è stata inventata e usata allo scopo nelle varie epoche storiche: si va dai segnali con bandierine, ai segnali acustici, al punto/linea dell'alfabeto Morse, ai segni su carta, agli spots su nastro o disco magnetico, ai micropits di un CD, .., ecc.

³⁵ Questi algoritmi sono di norma "aperti". Per meglio spiegarci: gli algoritmi che affidiamo al calcolatore sono "chiusi" nel senso che esso quasi sempre si arresta ad un certo punto della loro esecuzione. L'arresto è dovuto ad un ordine incorporato nel programma di esecuzione; *ma il criterio che fissa tale ordine è di solito arbitrario*, cioè dettato da un giudizio o da una decisione di solito opinabile di chi ha compilato il programma. Soltanto se quest'ordine di arresto "artificiale" manca, essendo sostituito da un ordine "naturale" che dice "*arrestati se e solo se hai realmente concluso il tuo lavoro*", e il calcolatore si arresta, siamo di fronte ad un autentico algoritmo chiuso.

(i): l'attività dell'automa enumerante è comunque destinata a rimanere sotto la minaccia della comparsa di contraddizioni inattese, perché in generale non possiamo sapere a priori se il sistema formale sul quale esso opera sia coerente, anche se probabilmente lo è;³⁶

(ii): le verità così individuate sono per lo più banali, per cui la loro raccolta risulterebbe in pratica poco interessante.

Senza dubbio, l'introduzione al problema dei fondamenti logico-matematici delle teorie scientifiche-esatte proposta in questa sottosezione è stata troppo concisa per non lasciare dietro di sé un bel numero di lacune e zone d'ombra, quasi tutte collocate nel delicato passaggio tra una generica struttura insiemistica intuitiva e la sua possibile versione strettamente formalizzata. Tentare di eliminare le prime e discutere più a fondo le seconde sarebbe qui del tutto fuori luogo. Abbiamo peraltro detto abbastanza, sulla presente situazione della matematica, da fugare l'illusione che essa possa ancora essere considerata come un sistema compatto di verità indubitabili: la convinzione dell'intima coerenza dei principi e dei teoremi matematici è stata duramente colpita nel corso dell'ultimo secolo. Qual è la matematica giusta? È legittimo definire come tale la matematica che ci assicura i migliori successi nello spiegare il mondo reale, conferendoci il maggior potere predittivo nei suoi confronti? È corretto piegarsi all'«autorità della natura» (Kline) in un campo apparentemente così lontano da essa?

Anche a queste domande si potrebbero dedicare pagine di approfondimenti tecnici e di riflessioni critiche; ma posto che chi scrive ne fosse del tutto all'altezza, questo ci porterebbe ad impegnarci su problemi *troppo lontani da quelli ai quali il nostro libro è dedicato*. Chiudiamo dunque qui l'approssimativa ma necessaria digressione contenuta in questa sottosezione. Varie precisazioni su definizioni per il momento soltanto accennate, nonché altri commenti e/o delucidazioni, si troveranno nella già menzionata App. Gen. A.

0.1.4) SCIENZE MATEMATIZZABILI VS. SCIENZE NON MATEMATIZZABILI, E ALTRI COMMENTI

Se da una parte una TFM possiede una sua capacità predittiva del tipo “se ... allora”, dall'altra non è detto che quest'ultima, o per così dire una sua “approssimazione per difetto”, basti da sola a stabilire la natura teoretica di una ricerca procedente dall'osservazione; e non a caso tale capacità è condivisa, in maggiore o minor misura, da quelle scienze empiriche che non hanno ancora raggiunto il carattere di scienze esatte, o che verosimilmente non potranno mai raggiungerlo

³⁶ La situazione è efficacemente sintetizzata dal seguente arguto ma profondo aforisma di H. Weyl (secondo Kline; secondo altri, di A. Weil): «Dio esiste perché la matematica è coerente; ma anche il diavolo esiste, perché non possiamo dimostrarlo.» (con mezzi leali).

a causa della irrealizzabilità, complessità, irripetibilità, ecc., delle procedure osservative/induttive connesse ai loro oggetti empirici, e della difficoltà, o virtuale impossibilità, di una loro matematizzazione (scienze (empiriche) “non-esatte”). È fuori discussione, infatti, che una certa attitudine predittiva del tipo “se ... allora” si possa attribuire ad una larga varietà di strumenti metodologici, ed al più basso livello conoscitivo a quel **principio di analogia** con cui ha inizio ogni processo di induzione; ma l’uso di questi strumenti appartiene per l’appunto a quel momento induttivo sul quale anche *si conclude* il percorso delle scienze non-esatte, nella gerarchia che ha alla sua base le scienze “descrittive-elencatorie” (≡ che meramente elencano oggetti empirici e fatti intorno ad essi, avendoli univocamente identificati mediante convenienti descrizioni), poi quelle “comparative” (≡ che confrontano, sotto qualche aspetto, oggetti e fatti stabilendo così un ordine più o meno “ragionato” su di essi, generalmente parziale), poi quelle “tassonomiche” o “classificatorie” (≡ che raggruppano oggetti e fatti in classi, non necessariamente disgiunte, secondo qualche criterio), ... e via dicendo.

Ben più significativo appare invece l’insieme delle conseguenze della matematizzazione, mediante la quale una TFM *ordina* i risultati delle osservazioni in una “rete logica” di correlazioni interne, proponendosi di *spiegarli* in una precisa e peculiare accezione (vedi App. Gen. A). Con queste proprietà, la TFM funziona allora come un vero e proprio meccanismo di precisione. Inoltre, la sua elaborazione agisce spesso come formidabile stimolo nei confronti della invenzione matematica, talvolta al punto di spostarne il maggiore interesse su quello specifico terreno (è ad esempio il caso della teoria newtoniana della gravitazione, che non giocò piccola parte nell’invenzione del calcolo infinitesimale, oppure quello della moderna “Teoria delle Stringhe”³⁷ nei confronti dell’Analisi su Varietà).

Riassumendo, la generica TFM possiede la tipica *capacità di ordinare, amplificare e unificare la conoscenza di* (una particolare sezione di) *un mondo fenomenico idealizzato e presuntivamente regolare*. Con ciò, da una parte un insieme più o meno amorfo di verità empiriche, sperimentalmente convalidate entro le approssimazioni osservative, viene logicamente ordinato facendolo risalire, lungo le catene deduttive percorse all’incontrario, ad un piccolo numero di asserti (gli assiomi, convenientemente ma non univocamente individuati³⁸); e dall’altra, emerge un nuovo

³⁷ Una teoria non convalidata che modella le particelle elementari come modi di vibrazione di varietà 1-dimensionali (e poi anche $(m>1)$ -dimensionali), chiuse o aperte, e che nelle speranze dei suoi ideatori era (e forse è ancora) candidata al ruolo di teoria unitaria della fisica. In questo caso, “stringa” è una cattiva traduzione dell’originale inglese, ed è da intendere come “corda”. Con l’occasione, segnaliamo che anche le “stringhe” intese come successioni finite (di cui alla precedente sottosezione) sono il risultato di una cattiva traduzione dall’inglese, ove “string” significa anche, secondariamente, “fila”.

³⁸ Si ritiene spesso – da parte dei non-addetti ai lavori – che gli assiomi siano introdotti in modo diretto e indipendente mediante le opportune procedure osservative/induttive (le ben note “verità evidenti” della tradizione). In realtà l’elaborazione di una TFM, e in generale di ogni sistema formale abbastanza potente, è un processo assai più tortuoso, del quale la fissazione degli assiomi (un’operazione comunque non univoca) è soltanto una parte. Insomma, quella

insieme similmente ordinato di enunciati riconosciuti come “veri”, generati dai meccanismi deduttivi ed eventualmente convalidabili dall’esperienza. L’enorme effetto di *sintesi* conoscitiva che ne segue è ben manifesto: per le date regole deduttive (che in pratica si scelgono una volta per tutte), l’insieme di tutti i teoremi della teoria, numerabilmente infinito, è generato dal piccolo numero dei suoi assiomi e “schemi” di assiomi. Infine il carattere ereditario della deduzione assicura automaticamente la contestualità del sistema di verità in oggetto: esso permane cioè all’interno del contesto fissato dagli assiomi, evitando generalizzazioni arbitrarie e/o prive di senso. Contestualità/specificità e potere generalizzante/predittivo sono proprietà tra loro in competizione: al crescere della prima il secondo tende a ridursi, ponendoci presto sulla strada dell’empirismo radicale, ovvero della negazione di ogni teoria predittiva. Tuttavia il fatto che le TFM esistano, e funzionino egregiamente, ci fa contare su un sufficiente spazio tra quelle proprietà antagoniste.

Alla matematizzabilità (\equiv equivalenza a una teoria assiomatico-deduttiva di tipo intuitivo, supposta coerente) è specificamente legata l’evidente natura autonoma, *indipendente dalla sua interpretazione*, della relativa TFM: natura tale cioè, ad esempio, che «tutti gli enunciati di una teoria matematica dell’elettricità valgano per ogni altro sistema di entità che si sostituiscano al posto delle nozioni di magnetismo, di carica elettrica, .. ecc., purché siano soddisfatti gli assiomi richiesti – supposti coerenti.» (Hilbert).³⁹ E infine, la stessa matematizzabilità assicura la validabilità, delegabile ad un automa, delle dimostrazioni dei teoremi; il che conferisce all’insieme di questi ultimi un totale grado di “certezza” a partire da quella degli assiomi supposti coerenti, e per le date regole di inferenza.

Volendo azzardare una formula conclusiva, potremmo dire che una TFM consiste nell’identificare un sistema organico e non banale di leggi empiriche, relative ad una classe non troppo ristretta di fenomeni idealizzati, la cui essenza è nella capacità di *ordinare e amplificare*

elaborazione ha nel suo insieme un carattere indiretto, non lineare, dinamico ed evolutivo: un «gioco di “avanti e indietro” noto in filosofia come “circolo ermeneutico”, in cui gli assiomi sono giustificati dalla forza dei teoremi che sono in grado di giustificare, ... e gli stessi teoremi, la cui verità è stata resa preventivamente evidente in modo intuitivo, sono resi inevitabili dalle dimostrazioni formali.» (G.C. Rota, in “Discrete Thoughts”, Birkhäuser (1986)). Del resto il formalismo (qui inteso in senso lato) non rispecchia per nulla l’esperienza che la grande maggioranza dei matematici e dei fisici-matematici ha della pratica della propria disciplina, la quale si basa di fatto su ogni sorta di intuizioni e induzioni, spesso procedendo per tentativi. Si legga, per una critica implacabile a qualunque approccio dogmatico alla matematica, la tesi di dottorato di Lakatos (Imre, 1922 - 1974) “Proofs and Refutations: The Logic of Mathematical Discovery” (pubblicata a Cambridge nel 1976, ma risalente a quasi venti anni prima). Al di là di queste argomentazioni, resta il fatto che se il formalismo ben difficilmente può essere considerato come un *metodo di sviluppo* delle teorie matematiche (e tanto meno fisico-matematiche), esso è senza dubbio l’unico modo efficace di *descriverle* inequivocamente, ad impresa compiuta; senza dimenticare che considerazioni puramente formali hanno spesso un importante ruolo euristico nel loro progresso.

³⁹ Con ogni evidenza, Hilbert fu estremamente sensibile a questa peculiare possibilità (già presente in embrione nella antica sillogistica e portata a pieno sviluppo da Frege) di de-interpretare una rappresentazione matematizzata di oggetti e di fatti ad essi afferenti. La frase citata si trova in una sua lettera a Frege, e fa il paio con un’altra assai più nota espressione della stessa idea, con la quale Hilbert proponeva di sostituire i punti, le rette e i piani della sua geometria assiomatica rispettivamente con «tavoli, sedie e boccali di birra».

sostanzialmente, via matematizzazione, il sistema di verità nel dato contesto, empiricamente verificate o verificabili. A questo punto si sarebbe tentati di aggiungere che la natura “scientifica” di una data attività equivalga alla possibilità di principio che essa si evolva in una TFM col progredire del suo sviluppo, anche in una prospettiva di lungo termine;⁴⁰ ma preferiamo non far qui nostra questa proposta definitoria non esente da incertezze pratiche, vedi l’esempio della nota precedente.

Come già detto, gran parte delle nostre attività “ragionative” (e in certo senso anche psichiche) è di natura induttiva. A buona ragione, si potrebbe addirittura affermare che senza questo tipo di procedure mentali la nostra stessa sopravvivenza fisica sarebbe continuamente a rischio: sono esse, infatti, che presiedono ai nostri comportamenti sulla base di stime probabilistiche ispirate dall’esperienza, ormai quasi istintive ma spesso inconsapevolmente raffinate. Nella misura in cui un sistema di conoscenze ottenute per via osservativa/induttiva viene man mano messo insieme, ordinato e organizzato, collegato e confrontato con altri analoghi sistemi, fatto oggetto di discussione e/o di insegnamento/apprendimento, e insomma istituzionalizzato, diventando così patrimonio condivisibile e trasmissibile, esso tende ad essere visto come “catalogo” o “codice” di conoscenze, poi come “sapere (in qualche modo) strutturato”, poi come vero e proprio “sapere esplicativo”, e infine come “scienza”; sebbene quest’ultima estrapolazione possa parzialmente confliggere con la definizione (informale) positivista che abbiamo riportato nella nota (⁵) (la diremo “definizione di scienza del tipo standard”), e possa totalmente confliggere con quella di TFM. Ora è chiaro che mentre una data attività più o meno soddisfacente alla definizione di scienza del tipo standard ha un solo modo di essere anche una TFM (quello di essere matematizzabile, e quindi di fornirci un potere predittivo affidabile), ne ha praticamente infiniti di non esserlo. In linea di principio, la vasta classe delle scienze positive (al presente) non matematizzate potrebbe essere ordinata in base alla misura con cui ciascuna di esse, mediante le procedure osservative/induttive che le sono proprie, è in grado di identificare presunte leggi non banali contestualmente generali, scoprendo così regolarità non banali del mondo fenomenico. Tuttavia è anche troppo evidente quanto incerti sarebbero i giudizi a presidio di tale ordinamento; per non dire di come la precaria gerarchia così possibilmente ottenuta risulterebbe in definitiva irrilevante agli effetti pratici. Quale utilità avrebbe infatti stabilire se al presente la clinica medica (mettiamo) sia “più scientifica” della vulcanologia – posto di poter dare a questa affermazione un significato obiettivo e condivisibile?

Vale la pena di proporre ancora qualche commento sull’uso sempre più libero della parola “scienza” che ricorre nel linguaggio corrente contemporaneo, e che sembra attribuire una natura scientifica sui generis, rispetto ad una definizione del tipo standard, ad attività che a vario titolo con

⁴⁰ Ad esempio, è concepibile che almeno una parte della moderna biologia molecolare sia avviata a diventare una scienza esatta; non solo, ma si cominciano a intravedere potenziali direzioni di sviluppo in tal senso, anche se in tempi per il momento imprevedibili.

la scienza in quella accezione poco, e spesso nulla, hanno a che fare: alludiamo a quelle “scienze” sociali, politiche, storiche, teologiche, occulte, della comunicazione, del comportamento, della formazione, .. e via seguitando, che così insistentemente sono nominate nel comune parlare. Al di là della questione nominalistica, che al più testimonia la drammatica mancanza di comunicazione tra mondi culturali separati anche nel linguaggio, *è bene riflettere sulle distinzioni di sostanza e trarne le dovute conclusioni*. Sistemi di conoscenze come quelli sopra ricordati – non matematizzabili salvo che in alcuni rari casi e sotto loro aspetti particolari⁴¹ – sono tutti assai lontani, nei metodi e nei fini, da quello scientifico-esatto; e ciò, non soltanto perché la loro elaborazione non soddisfa nemmeno le richieste minimali previste da una definizione di scienza del tipo standard, o le soddisfa assai imperfettamente. Infatti, le procedure osservative/induttive delle quali l’elaborazione di un eventuale modello teorico consiste (quando esistono), di norma non sono mirate ad individuare presunte leggi non banali e contestualmente generali; e in conclusione, esse non possono condurre, e di fatto in generale non conducono, alla scoperta di regolarità non banali della sezione di mondo considerata, e quindi ad una capacità predittiva affidabile su di essa. Se si accettano queste ragionevoli argomentazioni, ben difficilmente i sistemi di conoscenze della precedente lista (e simili) possono considerarsi come “scienze” (quantomeno nel senso della definizione standard); sebbene l’uso corrente le denomini con la stessa parola. Siamo cioè dinnanzi ad un tipico caso di “abuso di linguaggio” che si realizza e si accetta correntemente all’interno di un gergo, talvolta radicato nella tradizione e talaltra assai più recentemente introdotto.^{42, 43}

Ci limitiamo qui ad esaminare in qualche dettaglio una situazione esemplare, caratterizzata dall’uso della cosiddetta “logica del detective”: quella della ricerca storica. Ovviamente l’esercizio di questa “logica interpolativa” (un aspetto di quella induttiva) non è estraneo al processo scientifico-esatto – del quale è anzi quasi sempre un componente importante –, ma è ben lontano dal costituire di per sé un atto di quel tipo, o anche una sua premessa. Infatti ogni caso poliziesco

⁴¹ Come ad esempio in quello di certi meccanismi competitivi presenti nei o tra i gruppi umani, modellizzabili e matematizzabili all’interno della Teoria dei Giochi, e atti a rivelare possibili “sezioni esatte” (ideali) delle cosiddette scienze sociali, politiche e perfino storiche.

⁴² È invece assai naturale un’interpretazione maliziosa delle ragioni che spingono le comunità interessate a rivendicare il carattere “scientifico” delle loro attività: per comune (ma spesso recalcitrante) riconoscimento, la scienza – quella vera – ha ormai infatti un impatto formidabile sulla società, sia sul piano culturale che su quello delle sue ricadute socio-economiche, e conferisce in modo quasi automatico autorevolezza a chi la produce, o più semplicemente la sostiene o la utilizza. Di rado questa autorevolezza si traduce in autorità o in vero e proprio potere (piuttosto, di solito è il potere che la usa ai suoi fini); ma comunque, essa richiama emotivamente e soggioga una parte importante della pubblica attenzione.

⁴³ Quanto alla rilevanza della scienza nella società moderna, già nel 1873 (!) così prendeva atto della realtà lo statista inglese B. Disraeli nel suo discorso di insediamento: «Se penso che gli ultimi cinquanta anni sono stati il periodo più importante di qualunque altro nella storia dell’umanità, non mi riferisco all’ascesa o alla caduta di imperi, né al cambio di dinastie o alla costituzione di governi. Mi riferisco invece a quelle rivoluzioni scientifiche che hanno comportato più conseguenze di qualunque azione politica, ed hanno cambiato la condizione e le attese del genere umano più di tutte le conquiste militari, le leggi e i legislatori messi insieme.» (Beninteso, le conseguenze alle quali si riferiva Disraeli non sono necessariamente tutte positive, ma nemmeno tutte negative).

sta a sé, ed il fine del detective è di regola quello di risolvere quel caso specifico, volta per volta, e non certo quello di *indurre* leggi probabilistiche generali da una statistica di casi polizieschi per *dedurre* un sistema sostanzialmente amplificato di altre leggi probabilistiche generali. Se poi per mera ipotesi ciò si verificasse con le necessarie garanzie di universalità contestuale, rigore definitorio/deduttivo e non-banalità, allora a tutti gli effetti quel detective starebbe tentando di elaborare una “teoria probabilistica matematizzabile del caso poliziesco”; e quest’ultima attività, quand’anche seriamente concepibile, sarebbe cosa ben diversa dalla sua finalità originaria (vale a dire, il nostro detective sarebbe sul punto di cambiare mestiere). Con ciò non si nega che il detective possa farsi una sua esperienza generale (non banale ⁴⁴) sulla criminalità, ma soltanto che questa esperienza sia matematizzabile come lo è in una scienza esatta, e che quindi possa offrirgli certezze predittive di qualche valore. La metafora, applicata all’attività *logica* dello storico, è trasparente. ⁴⁵

Ci piace chiudere questa Introduzione citando un aforisma di Rutherford (Ernest, 1871-1937), che recita: «una scienza o è fisica o è filatelia». Al di là del suo gusto per le battute, in questa occasione il Nobel inventore del modello atomico (1911) non scherzava, e anzi si dimostrò di manica larga. Infatti “filatelia” è una ovvia metafora per “inventario”, o al più per “classificazione”; ma la maggior parte delle soprannominate cosiddette scienze non raggiunge nemmeno il livello della tassonomia. ⁴⁶

⁴⁴ Un asserto banale sulla criminalità è ad esempio: “La criminalità aumenta (o piuttosto è probabile che aumenti) in condizioni di disordine sociale”. È facile immaginare un gran numero di amenità di questo tipo.

⁴⁵ Per “Storia” si intende qui la sua accezione usuale, tradizionalmente riferita a gruppi umani abbastanza piccoli, a territori abbastanza poco estesi e a lassi di tempo abbastanza brevi. Se queste scale, e soprattutto quella temporale, si dilatano a sufficienza, la Storia-standard sconfinava nell’antropologia, e come tale si accosta alla scienza; pur restando ben distinta, per metodi e finalità, da una scienza esatta. La conoscenza, in campo antropologico come in moltissimi altri campi scientifici non esatti, è in pratica fondata *soltanto* sull’osservazione, sulla analogia e sui cosiddetti “esperimenti naturali” – cioè su esperimenti non progettati e realizzati dall’uomo. (Queste metodologie sono peraltro tutte accolte senza riserve nelle scienze esatte.) Diversamente da quanto accade nella Storia, nella scala antropologica non è l’uomo, e tanto meno l’uomo *singolo*, ma l’ambiente (e la tecnologia che sull’ambiente si modella), il fattore che veramente determina il corso degli eventi umani. La pretesa “scientificità” della Storia è sostenuta, a vero dire con argomenti ed esiti debolissimi, da certo pensiero filosofico contemporaneo.

⁴⁶ In apparenza in modo indipendente, la sentenza di Rutherford è stata riproposta quasi alla lettera, in tempi recenti e forse ormai un po’ sopra le righe, da G.C. Rota (nei già citati “Discrete Thoughts”), prendendo a termine di confronto la biologia: «Purtroppo il Galileo della biologia non è probabilmente ancora nato; e nell’attesa, le metodiche biologiche poco si discostano da quelle di un collezionista di francobolli.» Anche il fisico L. Alvarez (morto nel 1988) ha ripreso la stessa immagine in una polemica in cui manifestava disprezzo verso i paleontologi e i loro contributi alla scienza: «Davvero (i paleontologi) non sono buoni scienziati. Più che altro sono dei collezionisti di francobolli.», ebbe a scrivere in un duro articolo sul New York Times (citato in B. Bryston, “A Short History of Nearly Everything”, 2003). Insomma l’unica cosa che sembra definitiva, in questa querelle, è che l’interesse per i francobolli non sia tenuto in gran conto da fisici e matematici ...

CAP. 1

LA GEOMETRIA EUCLIDEA

1.1) LA GEOMETRIA COME PROTOTIPO DI TEORIA FISICO-MATEMATICA

1.1.1) GENERALITÀ E INQUADRAMENTO STORICO

La Geometria, per il momento intesa nella corrente accezione di “scienza dello spazio in cui tutto ciò che esiste/accade fisicamente ci appare immerso istante per istante, e delle sue figure”, è l’esempio archetipo di teoria fisico-matematica. Ovviamente essa non si sottrae allo schema generale delineato nella Introduzione 0.1, anche se la sua vicenda storica si presenta con due notevoli e in certo senso antitetiche peculiarità. Da una parte, abbiamo infatti scarsissime informazioni sui primi passi osservativi/induttivi (momento *modellizzante*) del suo sviluppo, come è ben naturale alla luce della estrema ed universale accessibilità della percezione e della esperienza spaziale *comune*, non mediata strumentalmente. Sull’altro versante, quello cioè della crescita del modello verso il possibile traguardo di una teoria formalizzata corretta/completa, quasi senza precedenti *direttamente* documentati la storia ci propone un incipit di impianto già sorprendentemente maturo circa 2300 anni fa: il primo e per lunghissimo tempo unico tentativo – nei fatti sostanzialmente difettoso – di organizzare assiomaticamente (momento *matematizzante*) un sistema di conoscenze empiriche. Con ogni evidenza, ci riferiamo qui ai celebri “Elementi” di Euclide (ca. 300 a.C.).¹

Una vera analisi critica nei confronti di questa opera fondamentale si sviluppò soltanto nel tardo Ottocento, in stretta relazione con quella Logica che nello stesso periodo rifioriva vivacemente dopo una lunga stagnazione, avviandosi a diventare l’universo matematico che è oggi. Una genuina matematizzazione di un sistema di conoscenze empiriche, osservavano Pasch (Moritz,

¹ Fermo restando l’incomparabile valore concettuale e storico dell’opera euclidea, va ricordato che essa ci è pervenuta attraverso traduttori, commentatori e revisori che possono averla modificata migliorandola, e che d’altra parte ci sono circa tre secoli di geometria e di matematica greca (per non nominare quelle babilonese ed egiziana) che la precedono, intorno ai quali abbiamo testimonianze quasi soltanto indirette. Ci riferiamo in particolare a Talete (c. 640 a.C., 546 a.C.), a Pitagora di Samo e alla sua scuola (V secolo), ad Ippocrate di Chio (IV secolo) e ad Eudosso di Cnido (IV secolo). Ben poco ci è pervenuto delle loro opere, ma sappiamo che Ippocrate fu a sua volta autore di “Elementi” che

1843-1930) e Hilbert, deve permettere di *destituire i suoi termini e i suoi enunciati da qualunque riferimento*; ma l'argomentare di Euclide certamente non regge a questa prova di de-interpretazione. La presa d'atto delle deficienze degli Elementi indusse così alcuni matematici di quell'epoca a dedicarsi ad una vera e propria rifondazione della geometria euclidea come teoria fisica matematizzata. Dovremmo anche aggiungere che, negli oltre due millenni trascorsi da Euclide fino ad allora, non si erano registrati avvenimenti decisivi in materia di geometria se si eccettuano due grandi conquiste nate sì in relazione ad essa, ma che poi avrebbero investito l'intero corpo della matematica. Esse furono da una parte l'invenzione delle coordinate, dovuta a Cartesio (o Descartes (René, La Haye Fr. 1596, Stoccolma 1650)) e a Fermat (Pierre de, Beaumont-de-Lomagne Fr. 1601, Castres Fr. 1665), che più tardi permise l'identificazione – via isomorfismo – della geometria euclidea con l'algebra e l'analisi standard su \mathbf{R}^3 (terza potenza cartesiana dell'insieme dei numeri reali \mathbf{R}^2); e dall'altra, la lunghissima riflessione, condivisa da una schiera senza fine di pensatori e matematici di diverse epoche e culture, sul cosiddetto **quinto postulato** (delle parallele)³. Culminata tra il terzo e il quarto decennio del secolo XIX, questa estenuante meditazione collettiva portò infine, essenzialmente per opera di Bolyai (Janos, 1802-1860), di Lobatchewsky (Nikolai, 1792-1856) e di Gauss – il quale ultimo decise tuttavia di non pubblicare nulla sull'argomento –, al riconoscimento della indipendenza del quinto postulato dai rimanenti, e quindi della possibile esistenza “logica” delle cosiddette geometrie non-euclidee.

Sulla base degli Elementi, per oltre due millenni la geometria di Euclide fu unanimemente considerata come insuperabile esempio di rigore per le altre scienze; ma dopo il lavoro critico tardo-ottocentesco, la possibilità/necessità di impugnarne la pretesa eccellenza matematica diventò sempre più evidente.⁴ Tra le opere di rottura, importante fu quella di Pasch,⁵ il quale tornò tra

sono forse serviti ad Euclide come modello o ispirazione. È comunque piuttosto improbabile che ci sia stato qualche serio tentativo di assiomatizzare la geometria anteriore a quello di Euclide.

² Una più comune notazione alternativa per questo insieme è \mathbb{R} .

³ In ordine approssimativamente cronologico, tra i contributi più noti si possono citare quelli di Tolomeo, Proclo, Nadir al Tusi, L. Gerson, P. Cataldi, G. Borelli, G. Vitale, J. Wallis, G. Saccheri, G. Klügel, J.H. Lambert, K.F. Gauss, A.M. Legendre, F. Schweikart, F. Taurinus, J. Bolyai, N. Lobatchewsky, B. Riemann, E. Beltrami, ecc. (altre informazioni su alcuni di questi studiosi sono riportate più avanti).

⁴ Scrive in proposito Russell (Bertrand, Penrhyndeudraeth Ingh. 1872, Trelleck Ingh. 1970) nel 1902: «Le definizioni di Euclide non sempre definiscono, e le sue dimostrazioni richiedono spesso assiomi dei quali egli è del tutto inconsapevole (...) La forza di una dimostrazione valida sta nel non disegnare alcuna figura, ma molte delle dimostrazioni euclidee cadono se sottoposte a questa prova (...) Il valore degli “Elementi” come capolavoro di logica è stato enormemente esagerato.» Il giudizio di Russell (nonché di altri logici e matematici dell'epoca, talvolta espresso in termini anche più duri) è incontestabile, per quanto possa apparire ingeneroso tenendo nel giusto conto l'immensità del tempo trascorso dall'antica Grecia al 1900. Del resto, si può escludere che Euclide avesse un'idea di ciò che, a partire dalla fine del secolo XIX, si intende per “formalismo radicale”. Ma nonostante tutto, e pur cadendo in vistose imperfezioni, del formalismo Euclide colse prodigiosamente l'essenziale, e molto altro.

⁵ “Vorlesungen über neuere Geometrie” (1882); 2^a edizione riveduta da M. Dehn (1926). Pasch richiede che la struttura matematica della geometria, compresi i meccanismi deduttivi che portano ai suoi teoremi, «sia completamente indipendente dal significato dei concetti geometrici e dalle figure; soltanto le relazioni tra quei concetti, stabilite nelle definizioni o nei teoremi, possono essere prese in considerazione.» È forse la prima volta che un'esigenza del genere viene avanzata in modo così esplicito.

l'altro a riconoscere (già Aristotele aveva espresso un'idea analoga) che alcuni concetti della geometria devono restare indefiniti; altrimenti, o diventa necessario basarli sull'esperienza, oppure si cade nel cosiddetto “rinvio senza fine”. Una volta selezionato un tale conveniente insieme di concetti primitivi (cioè, di termini, di formule su termini e di formule chiuse – in senso insiemistico-intuitivo), un più vasto insieme di concetti derivati si può definire a partire dai primi. Come altri matematici prima e dopo di lui, Pasch si ispirò alle idee di punto, retta e piano per introdurre i suoi termini primitivi. Gli assiomi fanno asserzioni sui termini primitivi e derivati, che sebbene suggerite dall'esperienza, devono ritenere carattere matematico-astratto (in quanto relazioni, primitive o derivate, su quei termini); ed i teoremi che ne sono le “conseguenze logiche” devono ricavarsi in base a precise regole deduttive. Una moltitudine di ricerche fondazionali fiorì lungo questa traccia, con maggiore o minor rigore e organicità. Possono in particolare menzionarsi i contributi della scuola italiana (Peano (Giuseppe, 1858 – 1932, nel 1889), Veronese (Giuseppe, 1854 – 1917, nel 1891), Enriques (Federigo, 1871 – 1946, nel 1898), Pieri (Mario, 1860 – 1913, nel 1899 e 1908)), inglese/americana (Whitehead (Alfred, 1861 – 1947, nel 1906), Veblen (Oswald, 1880 – 1960) e Young (John, 1879 – 1932, nel 1910 e 1918)), tedesca, polacca, ... ecc.; ma il sistema geometrico assiomatico (o formalizzato ⁶) che riscosse unanime attenzione e consenso, e che doveva presto affermarsi come il più soddisfacente per profondità e completezza, fu quello proposto all'esatta fine del secolo da D. Hilbert.⁷

Secondo l'interpretazione corrente, l'assiomatizzazione (o meglio il tentativo di assiomatizzazione) di Euclide consta di 23 definizioni, come ad esempio quella di “punto” o quella di “(linea) retta”, di 5 assiomi (comunemente detti “postulati”) e di altrettante “nozioni comuni” (dette “assiomi” da Proclo), le quali ultime formano un piccolo codice di asserti a priori intorno ad oggetti di natura generale e non specificata. Nella sua versione del 1930 (ritenuta dai più essenzialmente stabilizzata ⁸) l'assiomatizzazione hilbertiana elenca 3 tipi di termini primitivi (i \lfloor punti \rfloor ⁹, le \lfloor rette \rfloor e i \lfloor piani \rfloor), 4 tipi di relazioni primitive (il \lfloor giacere su \rfloor , lo \lfloor stare tra \rfloor , la

⁶ Non faremo ormai più distinzione tra “assiomatizzare” e “formalizzare”, e affini. Una formalizzazione “in senso stretto” (v. S.sez. 0.1.3) sarà invece menzionata esplicitamente come tale.

⁷ “Grundlagen der Geometrie” (Leipzig 1899, 7^a edizione, Teubner 1930), trad. ital. in Feltrinelli (1970). In maggiore o minor misura, tutte le successive formalizzazioni della geometria, non soltanto euclidea, sono debitrice a quest'opera fondamentale.

⁸ La possibile non-indipendenza di alcuni assiomi ha dato luogo a ricerche relativamente recenti.

⁹ In questa Sez. 1.1, parole/espressioni “astratte” (o come anche ormai diremo, “formali”) – ossia nomi convenzionali per ben definiti termini/relazioni di un linguaggio insiemistico (qui quello geometrico di Hilbert) – ma dotate di significato *nel linguaggio geometrico intuitivo*, sono racchiuse tra [...] al loro primo apparire. Queste parole/espressioni sono formali in quanto da intendere come prive di “significato” empirico, ma comunque dotate del “senso” che deriva loro dagli assiomi. Ciò vale del resto per tutta la terminologia fisico-matematica, che è sempre “suggestiva” e mai “significante” in senso stretto. (Lo stesso vale ovviamente per la terminologia matematica, che è suggestiva, e significativa soltanto in senso astratto – vedi la teoria degli insiemi che ne è la sorgente semantica.) L'interpretazione naturale può essere un'utile guida alla generale intelligenza del linguaggio astratto cui quella

[congruenza di coppie di punti] e la [congruenza di coppie di angoli], [angolo] essendo un termine derivato) e 20 assiomi divisi in 5 famiglie: 8 assiomi di [incidenza] che descrivono la prima relazione, 4 assiomi di [ordine] che descrivono la seconda, 5 assiomi di [congruenza] che descrivono la terza e la quarta relazione, 1 assioma di [parallelismo], e 2 assiomi di [continuità]¹⁰. Lo [spazio] descritto da questo sistema assiomatico può pensarsi come l'insieme $H_3 \equiv H$ dei suoi punti, e risulta [3-dimensionale] in una precisa accezione formale. Lo stesso Hilbert, inoltre, indicò come generalizzare le nozioni primitive e gli assiomi in modo da ottenere un analogo spazio [n-dimensionale] H_n per qualsiasi $n \geq 1$, partendo, se $n > 3$, dai suoi punti, rette, piani (\equiv 2-piani) ... [(n-1)-piani] come termini primitivi.¹¹ In pratica, e Hilbert lo sostenne in più occasioni, i concetti primitivi sono indirettamente definiti da alcuni degli assiomi.

Come si diceva, fin dal loro primo apparire le Grundlagen der Geometrie suscitarono grande interesse nella comunità matematica, nonché molto lavoro critico e di perfezionamento; il loro stesso autore, in particolare, continuò a dedicarvi attenzione per oltre trent'anni, elaborandone una lunga serie di versioni successive (dieci!), sempre più ampie e sofisticate. Per dare un'idea dello spirito con cui l'assiomatizzazione hilbertiana della geometria euclidea era considerata negli ambienti matematici di pochi anni fa, può valer la pena ricordare che R. Thom, riferendosi ad essa come ad un'opera «di spaventosa complessità», bocciò fermamente la proposta, che tale rimase, di usarne qualche parte come materia di studio nei licei francesi, in sostituzione o a complemento degli Elementi di Euclide. Ed effettivamente, specie se vista nel complesso del suo grande disegno logico, la qualità dell'«impresa geometrica» di Hilbert (tra le tante che il matematico riuscì a portare a termine) è tale da renderla del tutto improponibile alla pur volenterosa attenzione di un neofita di media attitudine; anche se in linea di principio la sua sostanziale comprensione, passo dopo passo, *non presuppone conoscenze di alcun tipo*.

Non facendo appello a procedure extra-geometriche, cioè algebrico-analitiche, le presentazioni assiomatiche della geometria euclidea che stiamo considerando possono definirsi come **sintetiche** (ove l'attributo è usato in contrapposizione ad **analitiche**, cioè legate «all'analisi» su \mathbf{R}^3). Ma come si accennava più sopra, partendo da una conveniente **coordinatazione** dello spazio

terminologia appartiene, ma tale intelligenza sarà tanto più affidabile quanto meno a quella interpretazione si sarà fatto ricorso. Nel seguito, ovviamente rinunceremo ad evidenziare tra [...] le parole astratte, limitandoci ad usare il solito carattere grassetto nel nominare per la prima volta gli oggetti delle definizioni strette, oppure le nozioni di particolare importanza contestuale. Quasi tutte le parole del testo scritte in grassetto sono poi riportate negli appositi glossari alla fine del libro.

¹⁰ Come vedremo i due assiomi di continuità di Hilbert sono spesso riuniti in un unico assioma.

¹¹ In particolare, gli assiomi di incidenza di Hilbert con i quali (assieme agli assiomi delle altre famiglie) si può formalizzare la gerarchia degli spazi H_1, \dots, H_n sono i seguenti: con $1 \leq i \leq n-1$, (1) «ogni H_i è un insieme di punti»; (2) «una retta cui appartengono due punti di H_i giace in H_i »; (3) «esiste un H_i cui appartengono $i+1$ punti dati»; (4) «in ogni H_i esistono $i+1$ punti non appartenenti ad alcun H_{i-1} »; (5) «esistono $n+1$ punti non appartenenti ad H_{n-1} ».

formale è anche possibile *identificare* la geometria euclidea 3-dimensionale [(n≥1)-dimensionale] con l'algebra e/o l'analisi su \mathbf{R}^3 [su \mathbf{R}^n]. Di fatto, un'accettabile assiomatizzazione del campo reale¹², al quale si riconduce la coordinatazione di un insieme continuo di punti, non fu disponibile che nel secolo XIX avanzato; e questo spiega perché una fondazione assiomatica della geometria euclidea per via algebrico-analitica non sia stata possibile prima di allora. Ma a parte questo, e molto più importante, va sottolineato che nessuna coordinatazione *ci è data a priori*: ne occorre una definizione indipendente (sintetica) se si vuole evitare che la geometria euclidea n-dimensionale assiomatizzata scada al ruolo di fotocopia della teoria assiomatizzata della n-ma potenza cartesiana dell'insieme dei numeri reali.

Riferendoci ormai ad accezioni della geometria anche più generali di quella euclidea, potremmo figurarci una geometria come il passaggio da una “dimensione dell'esperienza (sensibile)”, ad una “dimensione dell'intuizione”; da questa si muoverebbe infine ad una associata “dimensione formale”, cioè genuinamente matematica, mediante laboriosi e complicati processi di astrazione, analisi e organizzazione sistematica. Ora, finché la geometria si riferisce allo *spazio fisico* (o empirico), i concetti e le operazioni primitivi devono esservi definiti in termini empirici; le verità fondazionali della disciplina sono quindi enunciati empirici, o kantianamente «sintetici a posteriori», e come tali soggetti a revisione. Ciò è precisamente quanto accaduto alla geometria euclidea a fronte delle scoperte sperimentali che hanno portato alla geometria relativistica. Ma *in quanto matematica*, la geometria – come del resto qualunque teoria fisica-matematica – non può certo esser posta su fondamenti empirici. Come si devono dunque giudicare i suoi assiomi? La risposta è stata ampiamente suggerita in quanto precede. Vale a dire, quegli assiomi interpretano/traducono certe *presunte verità empiriche idealizzate* relative a concetti geometrici primitivi o derivati, che per la loro natura è conveniente, cioè più economico che scegliendone altre nello stesso ruolo, porre al vertice¹³ di un sistema deduttivo capace di produrre, in modo “automaticamente” (≡ delegabile ad un automa) validabile, un sistema amplificato di altre verità empiriche idealizzate dello stesso tipo, sperimentalmente verificate o verificabili entro le solite approssimazioni.

Al giorno d'oggi, sia la geometria fisica che la geometria matematica hanno avuto sviluppi di tale entità e qualità che le loro primitive accezioni sono soltanto dei remoti per quanto importanti punti di partenza. In particolare con riferimento alla seconda, l'incessante opera di

¹² Gli assiomi specifici del “campo reale astratto”, del quale il familiare “campo reale standard \mathbf{R} ” è una particolare accezione, sono 1) gli assiomi di campo; 2) gli assiomi di ordine; 3) l'assioma di Archimede (Siracusa 290 a.C.-212 a.C.); 4) l'assioma di completezza (o di Hilbert). L'assiomatizzazione di Hilbert della geometria presuppone il campo reale \mathbf{R} nella sua completezza strutturale; scelta che è stata giudicata, secondo chi scrive a torto, «non priva di ingordigia» (J. Gray, “Ideas of Space”, Clarendon Press, Oxford 1979).

approfondimento, in direzioni sempre più generali ed astratte, ha da tempo portato ad un corpus matematico di proporzioni gigantesche. In proposito, ci limitiamo a segnalare nella nota ¹⁴ le voci di ordine zero (\equiv grandi blocchi di argomenti interconnessi) legate alla teoria geometrica e alle sue molteplici filiazioni/estensioni che si trovano elencate nella “Mathematics Subject Classification” del 1980.

Utilizzando gli strumenti di cui disponevano, dai tempi più remoti i filosofi si sono “naturalmente” impossessati della riflessione sul problema della conoscenza del mondo fisico, e in particolare della geometria, talvolta inquinandolo con dosi inaccettabili di negligenza concettuale e di sostanziale distacco dalla realtà (questa critica vale pur tenendo conto, volta per volta, dei limiti delle conoscenze specifiche dell’epoca ¹⁵). Un caso esemplare in questo senso ci sembra quello del rapporto di Kant (Immanuel, Königsberg 1724-1804) con la geometria euclidea, che ai suoi tempi, e per qualche decennio ancora, era o sarebbe stata l’unica disponibile. Secondo il padre dell’idealismo tedesco, le proposizioni della geometria (euclidea) sono asserti di tipo intuitivo da lui detti «sintetici a priori»: lo spazio in senso stretto (ma anche il tempo è trattato in modo simile) è una «forma a priori» dell’umana «sensibilità», che condiziona quanto apprendiamo attraverso i sensi, e le sue proprietà sono conoscibili come proiezione e necessità del pensiero. ¹⁶ Quanto poi all’*esistenza* di tali giudizi sintetici a priori, Kant tentò di dimostrarla servendosi di argomentazioni che, prevedibilmente e di fatto, risultano di indiscutibile natura metafisica.

Questo genere di affermazioni, che Gauss fu tra i primi ad avversare fermamente, appare aberrante alla luce dello spirito di ogni seria indagine sul mondo dei fenomeni. L’intuizione fisica

¹³ Quasi universalmente ci si rappresenta la ramificazione deduttiva come procedente dall’alto verso il basso, anche se la scelta opposta sarebbe altrettanto (o forse più) legittima.

¹⁴ Senza pretese di completezza, si possono indicare: il #51 (Geometry, con ca. 80 voci di ordine 2), il #53 (Differential Geometry, ca. 60 voci, sempre di ordine 2), il #54 (General Topology, ca. 90 voci), il #57 (Manifolds, quasi 100 voci), il #58 (Analysis on Manifolds, ca. 100 voci, anche se in parte di non immediata pertinenza geometrica), il #14 (Algebraic Geometry, ca. 50 voci), il #55 (Algebraic Topology, poco meno di 100 voci); e qualcosa ancora, per un totale che non si discosta molto dalle 500 voci di ordine 2. La classificazione degli argomenti matematici nacque in Germania verso la fine dell’Ottocento, e da allora è stata continuamente estesa ed aggiornata. Al presente è gestita dalla American Mathematical Society (AMS), e contiene qualcosa come 4000 voci di ordine 2. Chi ha scarsa familiarità con i volumi delle Mathematical Reviews che mensilmente ne pubblicano i contributi, potrebbe supporre che quelle voci di ordine 2 siano molto specialistiche, e che il loro enorme numero tenda quindi ad amplificare artificialmente l’impressione di vastità della materia. Ma non è così: nel nostro giudizio, forse due o tre di esse, possibilmente scelte tra le più affini tra loro, sarebbero oggi più che sufficienti ad occupare l’intera carriera di un giovane (e poco curioso) matematico di medio talento che ambisse a diventare un esperto ragionevolmente creativo in quel campo.

¹⁵ Un’opinione radicale, a questo o anche più ampio proposito, è quella dell’eminente logico polacco J. Łukasiewicz (1878-1956), che così si esprimeva in “O Determinizmie”: «Se esaminiamo i grandi sistemi filosofici di Platone o Aristotele, Descartes o Spinoza, Kant o Hegel con gli standard di rigore della moderna logica matematica, questi sistemi vanno in pezzi come castelli di carte. I concetti di base sono mal definiti, le tesi più importanti incomprensibili, le dimostrazioni e i ragionamenti inesatti, e le teorie logiche su cui sono fondati tutte sbagliate. La filosofia dovrebbe essere completamente riedificata ispirandosi al metodo scientifico, e sulla base della nuova logica.» Ai tempi in cui Łukasiewicz faceva queste “scandalose” affermazioni, l’emigrazione di certe energie filosofiche verso la Logica Matematica era già cominciata da un pezzo.

¹⁶ «La nozione di spazio non ha origine empirica, ma è una inevitabile necessità del pensiero.» E ancora: «Le proposizioni della geometria sono conosciute sinteticamente a priori e con apodittica certezza.»

stessa (è quasi banale sottolinearlo) si forma infatti *sulla base dell'esperienza*, non soltanto nell'individuo, ma addirittura nella specie, possibilmente accumulata e stratificata come vero e proprio patrimonio inconscio; *inconscio*, ma comunque *non a priori*. Nel corso degli ultimi due secoli circa, inoltre, questa esperienza è andata facendosi mediata e amplificata da strumenti osservativi e logici sempre più sofisticati, al punto che le presenti idee (degli addetti ai lavori) sullo spazio e sul tempo, o piuttosto sullo spazio-tempo nelle sue più avanzate accezioni odierne (come lo spazio-tempo relativistico-generale, oppure, all'estremità opposta, come l'ipotetico spazio-tempo alla scala di Planck, forse per sempre inosservabile), hanno ormai pochissimo a che vedere con le rappresentazioni che mossero i nostri antenati verso le loro immagini intuitive.¹⁷ Per concludere questi rilievi, va anche sottolineato che, sebbene l'idea dei giudizi sintetici a priori fosse già radicalmente contestabile quando fu formulata¹⁸ – tant'è che alcuni spiriti illuminati se ne resero conto appena ne vennero a conoscenza –, essa risultò «completamente annichilita dalle scoperte del nostro (XX, ndr) secolo. La teoria della relatività ha infatti mutato le nostre concezioni sullo spazio e sul tempo, rivelando loro aspetti del tutto nuovi, dei quali non vi è traccia nelle forme a priori kantiane dell'intuizione pura.»^{19, 20}

¹⁷ Può essere istruttivo, a questo proposito, riflettere sul rapporto numerico (congetturalmente almeno 10^{61}) tra le massime e le minime presunte dimensioni dello spazio fisico – dalla decina di miliardi di anni-luce nel cosmo alla lunghezza di Planck $l_{pl} = 4,13 \cdot 10^{-35}$ m o meno (vedi 0.0.1) – con le quali si confronta oggi la fisica nel suo insieme, paragonandolo con la ridicola esiguità della “finestra” su di esso apertaci dalle nostre capacità sensoriali *non amplificate strumentalmente*. Riferendoci alle dimensioni di oggetti fisici direttamente accessibili alla nostra esperienza sensibile, questa finestra può considerarsi larga circa 10^8 , da qualche decina di chilometri a qualche decimo di millimetro. Resta dunque uno spaventoso “buco” largo almeno $10^{61-8} = 53$, sul quale non ci è possibile gettare uno sguardo *diretto, non amplificato*; e al suo interno un buco più piccolo (largo circa 10^{18}), intorno alla cui possibile accessibilità osservativa è difficile fare previsioni. Simili considerazioni potrebbero ripetersi relativamente ai tempi, alle masse-energie, ecc.

¹⁸ I. Kant, “Critica della Ragione pura”, 2ª edizione riveduta (1787).

¹⁹ W. Heisenberg, (1901-1976) “Physics and Philosophy” (1958), trad. ital. in Il Saggiatore (1961). Questo notissimo saggio contiene altri riferimenti alla filosofia naturale di Kant, ad esempio quello alla sua versione del principio di causalità, anche in tal caso di segno negativo.

²⁰ Per quanto possa meravigliare, ancora un secolo fa esistevano “circoli di pensiero” in cui si coltivava Kant come inventore dei giudizi sintetici a priori, e addirittura come assertore di una sorta di presunta “euclideanità soggettiva” dello spazio fisico. Sembra non ci si rendesse conto, in quegli ambienti, che la tesi di fondo della “Critica della Ragione pura” – che nella conoscenza del mondo fisico «non è la mente che si conforma alle cose, ma piuttosto le cose alla mente» (H.J. de Vleeschauwer, 1974) – scade in concreto in un oscuro e gratuito schema psicocentrico, ogni traccia utilizzabile del quale è praticamente scomparsa dal pensiero positivo posteriore. Per ridurre all'osso la questione: nel linguaggio kantiano la geometria fisica è «sintetica», mentre la geometria matematica è «a priori». Il punto è per contro che *nessuna geometria (e a maggior ragione nessuna fisica) è al contempo sintetica e a priori*. Kant non lo capì, e credette di aver dimostrato il contrario. Né ha senso tentare di difendere in questo campo il campione della filosofia tedesca con l'argomento che ai suoi tempi le geometrie non-euclidee (a priori o sintetiche che fossero) non erano state ancora scoperte: ciò che è qui in gioco è un principio generale *intrinsecamente* fallace, che prescindendo dalle conoscenze empiriche e matematiche effettivamente disponibili investe in modo scorretto tutto il nostro rapporto cognitivo con il mondo fisico. Questo tipo di errore non è rimediabile, e non a caso Kant era dogmaticamente convinto che le sue idee in materia sarebbero state «la base di ogni futura metafisica che voglia presentarsi in forma di scienza». Non fa insomma meraviglia che con tali presupposti Kant non abbia dato alcun suggerimento costruttivo finalizzato all'unico obiettivo di una geometria (per tacere della fisica in genere) sperimentalmente e matematicamente sensata, quello di proporre/indicare una rappresentazione delle sue figure rispondente ai criteri di formalizzabilità non-banale e di correttezza/completezza che sono peculiari della conoscenza scientifica in senso stretto. Le stroncature (di molti aspetti) della filosofia naturale di Kant, e più in generale il dis-prezzo del contributo della filosofia alla scienza da quando la seconda si è resa autonoma dalla prima (più o meno nel secolo XVII), non possono certo vedersi, ormai da tempo, come

In conclusione, la nostra rappresentazione dello spazio (o del tempo) e delle loro figure, *nonché ovviamente del mondo fisico nel suo complesso*, non può procedere da altro che non sia l'“osservazione ragionata della natura”; e quindi essa è destinata ad evolversi, anche se nel sostanziale rispetto della capacità cumulativa delle scienze esatte (vedi 0.1), con il raffinarsi dei nostri mezzi osservativi e delle nostre conoscenze matematiche. Non è un caso, a questo proposito, che l'attività fisico-teoretica abbia cominciato a dare risultati veramente significativi, come strumento predittivo del tipo “se ... allora”, da una parte con l'imporsi del metodo sperimentale ed il decollo di genuine tecnologie osservative (ad esempio con lo sviluppo degli strumenti ottici)²¹, e dall'altra con la simultanea crescita e il supporto sempre più efficace dell'arsenale matematico. Ciò cominciò ad affermarsi con chiarezza, nell'ambito di quella che oggi chiamiamo “cultura occidentale”, in un periodo che può collocarsi tra il XVI e il XVII secolo; e segnatamente attraverso l'insegnamento di Galilei (Galileo, Pisa 1564, Arcetri It. 1642), che fu strenuo propugnatore del metodo sperimentale e iniziatore di una fisica radicata nel confronto con il mondo sensibile, nonché convinto assertore del ruolo primario, nel progresso delle scienze esatte, del razionalismo matematico in alternativa alla tradizione verbale di origine aristotelica.²²

1.1.2) GEOMETRIA EUCLIDEA SINTETICA E ANALITICA

Come ben evidenzia ogni illustrazione non elementare dell'argomento, qualunque vera formalizzazione del **modello** (teorico, nel senso dei fisici, vedi S.sez. 0.1.1) **euclideo** dello spazio

posizioni critiche granché originali, ben inquadrandosi invece nella percezione della inadeguatezza della filosofia ad interagire positivamente con il corso del progresso scientifico. Questo diffuso sentire non dovrebbe tuttavia indurre, a nostro avviso, il convincimento che una riflessione filosofica sulle scienze esatte – o sulla “scienza” tout court, per ciò che questa parola significhi – sia necessariamente superflua o addirittura fuorviante; ma piuttosto sottolineare che per affrontarla in modo proficuo occorrono, oggi come già occorre ai tempi di Kant, strumenti specifici acquistabili soltanto attraverso un severo tirocinio diretto, nonché *una sana dose di prudenza*. Chi non riconosce queste regole del gioco, e ugualmente pretende di far sentire la sua voce, quasi certamente commette errori o propone banalità, come prova una lunga galleria di episodi più o meno celebrati, ma comunque effimeri, penosi o esilaranti a seconda dell'attitudine di chi li giudica con più matura cognizione. Vedi anche la nota (5) della Sez. 9.1.

²¹ Per quanto difettoso, quello della geometria di Euclide fu un esito eccezionalmente indipendente da tali tecnologie, giustificato, come si diceva, dalla completa accessibilità della *comune* esperienza spaziale.

²² La fede di Galilei nel disegno matematico della natura è eloquentemente espressa nel celeberrimo giudizio che riportiamo qui per intero: «La natura è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi agli occhi (e dico l'universo), ma non si può intendere se prima non s'impara a intender la lingua, a conoscer i caratteri ne' quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi e altre figure geometriche, senza i quali mezzi è impossibile a intenderne umanamente parola; senza questi, è un aggirarsi vanamente per un oscuro labirinto» (G. Galilei, Opere, IV, 171 (1610)). In forza dei tempi, e forse anche perché non era un matematico, Galileo ravvisava in «triangoli, cerchi ed altre figure geometriche» i «caratteri della lingua matematica»; ma è ugualmente ben chiaro e significativo cosa intendesse dire. In tema di tecnologie osservative, ricordiamo poi che Galileo teorizzò e perfezionò sostanzialmente il cannocchiale inventato dagli ottici-artigiani olandesi, trasformandolo in vero e proprio telescopio astronomico.

fisico, ovvero della **geometria euclidea**²³, è un'impresa di grande spessore culturale, giunta a completo successo soltanto negli ultimi anni del secolo XIX. Sarebbe dunque impensabile di poter illustrare esaurientemente questo grande tema fisico-matematico all'interno di un solo capitolo. D'altra parte, ciò di cui avremo davvero bisogno nel seguito è soltanto un solido inquadramento generale della materia e la chiara definizione dei suoi concetti di base. A questa limitata ma per nulla banale finalità è dedicato il resto del capitolo, oltre alla presente sottosezione che ne è una elementare introduzione informale.

Non vi è dubbio che la geometria euclidea si adatti con alta approssimazione alla comune esperienza spaziale e all'immagine intuitiva che ne ricaviamo. Ma se estendessimo le nostre osservazioni a porzioni di spazio in prossimità di corpi altamente massivi, allora i loro risultati potrebbero differire (e alla luce delle attuali conoscenze con ogni confidenza differirebbero) da quelli che qualunque persona ignara di relatività generale si aspetterebbe (o che chiunque si aspetterebbe da un punto di vista istintivo). Infatti siamo condotti alla geometria euclidea da circostanze tutto sommato accidentali, quali quella di fare esperimenti su regioni spaziali della nostra scala di grandezza e nelle tipiche condizioni offerte dal nostro pianeta, oltre naturalmente che dalla approssimatività dei comuni mezzi osservativi. Se dunque cercassimo di dare ai concetti geometrici primitivi un significato realmente *naturale*, sia nel dominio fisico che in quello ad esso collegato dell'intuizione (momento modellizzante), potrebbe ben non seguire, e di fatto non seguirebbe in generale, la necessità *concettuale* della geometria euclidea. Questa conclusione dà ancora una misura di quanto incaute fossero le concezioni kantiane. D'altra parte problemi di questo tipo non interessano particolarmente i geometri matematici: essi non sono infatti tenuti a scoprire e studiare soltanto la geometria "fisicamente corretta", ma piuttosto ad inventare ed analizzare (ancorché sotto gli ineludibili vincoli imposti dall'esperienza) le possibili geometrie matematicamente sensate, nonché le loro generalizzazioni e mutue connessioni. Ciò spiega bene i grandiosi sviluppi della geometria matematica, le molteplici direzioni di indagine al suo interno, e gli importanti collegamenti con altre matematiche.

Come si diceva, e come non potrebbe essere diversamente, un fisico accetta una teoria fenomenica matematizzata quando i termini e le relazioni su termini ricevono una ragionevole definizione operativa, e gli enunciati introdotti come assiomi, o ricavati come teoremi, corrispondono a verità empiriche verificabili con la conveniente approssimazione. Consideriamo per cominciare la famosa prima (presunta) definizione degli Elementi di Euclide – quella di "punto", secondo la quale esso è «qualcosa che non ha parti». Oltre ad essere alquanto oscura,

²³ Qualcuno chiama "propriamente euclidea" la geometria euclidea, e semplicemente "euclidea" quella che in questo libro è nominata come "pseudoeuclidea". L'idea non sembra troppo felice, perché va contro l'uso comune e può creare qualche equivoco.

questa espressione sembra rinviarci a nozioni più primitive, che tuttavia non sono a loro volta definite. Prendendola alla lettera in chiave insiemistica moderna, un insieme privo di parti proprie e diverso dall'insieme vuoto deve essere un singoletto, deve cioè constare di un solo elemento; il che ci riporta esattamente alla situazione di partenza. Comunque la si consideri, insomma, la definizione euclidea è vuota; e se (come non è da escludere) Euclide stesso intendeva che così rimanesse, tanto valeva allora, seguendo l'ammonimento aristotelico, introdurre la nozione di punto come dichiaratamente primitiva.

D'altra parte, almeno implicitamente Euclide sembra anche ammettere tra le nozioni primitive quella "estensiva" di *grandezza*, e quindi in particolare quella ad essa strettamente collegata di *distanza*; ad esempio quando sostiene la possibilità di replicare indefinitamente una lunghezza lungo una retta. Alla luce di ciò, si sarebbe tentati di definire la nozione intuitiva di punto dello spazio fisico come quella di una sua regione connessa di "estensione abbastanza piccola" (in senso vago). E in effetti, un approccio intuitivo al problema probabilmente privilegia questa idea di "piccola regione spaziale", in quanto più vicina alla nostra esperienza immediata, rispetto a quella, decisamente astratta, di punto.²⁴ Tuttavia questa prospettiva appare anche più insoddisfacente, in quanto l'idea di regione spaziale, e a maggior ragione quella di *piccola* regione spaziale, è troppo complessa per essere accettata come primitiva ed essere posta alla base di un discorso formalistico. Ciò è vero anche se il comune osservatore effettivamente pensa ad un punto dello spazio fisico come ad una sua regione piccola (in senso vago) rispetto alle distanze tra le piccole regioni spaziali (sempre in senso vago) di interesse; il che gli riesce soltanto perché fa un uso essenziale, sostanzialmente corretto anche se su base intuitiva e inconsapevole, di una varietà di nozioni che esulano da quelle strettamente insiemistiche, e che gli consentono di ragionare senza errori grossolani. L'aver dunque intuito, da parte di Euclide, che l'iter definitorio "piccola regione spaziale (termine primitivo, cioè *non* definito) → punto (termine derivato)" non era utilmente percorribile, ma che lo era il suo opposto "punto (termine primitivo) → piccola regione spaziale (termine derivato)", sebbene lungo uno sviluppo estremamente laborioso e sofisticato, prova la sua altissima capacità di immaginazione astratta.

A nostro avviso, questo del punto geometrico è un esempio estremamente significativo di come, anche partendo con la più forte inclinazione mentale verso la cosiddetta "realtà concreta dei fatti" e il cosiddetto "sano buon senso", dobbiamo convivere con nozioni empiriche *idealizzate*, cioè definite al prezzo di convenienti e problematici *atti di astrazione*, se vogliamo pervenire a modelli (parziali) del mondo di reale utilità; sia che di quegli atti abbiamo coscienza/cognizione o

²⁴ Lamb (Horace, 1849-1934), un fisico-matematico inglese noto in particolare per il suo trattato di Idrodinamica (1879), si proponeva di «erigere un monumento allo sconosciuto inventore del "punto matematico" (leggi geometrico, ndr), tipo supremo di quell'astrazione che è stata la condizione necessaria del lavoro scientifico fin dal suo inizio».

no. È appunto una conveniente scelta di tali nozioni empiriche idealizzate che forma l'ossatura logica di una teoria fisica matematizzata, cioè l'insieme dei suoi termini primitivi se quelle nozioni hanno natura oggettuale, quello delle sue relazioni (o formule) primitive se esse hanno natura predicativa, e infine quello dei suoi assiomi se esse si identificano con relazioni chiuse *introdotte* nel gioco inferenziale come enunciati veri per definizione. Tornando alla geometria (fisica), chiameremo qui **punto empirico idealizzato**, o semplicemente **punto empirico** l'oggetto che abbiamo cercato di descrivere più sopra; ed è chiaro che *nel presente ordine di idee* l'effettivo "diametro" di un tale punto empirico non può essere più piccolo del più corto regolo materiale disponibile, cioè del passo di un tipico reticolo cristallino, che è dell'ordine di qualche decina di angstrom ($\sim 10^{-9}$ m, essendo $1 \text{ \AA} \equiv 10^{-10}$ m). Questo è consistente con la dimensione di un corpo macroscopico, che abbiamo valutato (cfr. Presentazione 0.0) essere "abbastanza più grande" di $\approx 3 \cdot 10^{-6}$ m (prodotto della velocità della luce per 10^{-14} s).

La geometria analitica ²⁵ euclidea è ovviamente contenuta nel sistema assiomatico di Hilbert. Precisamente, in esso si dimostra che nello spazio (sintetico) che abbiamo denotato con $H \equiv H_3$ (vedi S.sez 1.1.1, H come Hilbert) quando pensato come insieme dei suoi punti, è possibile, e in infiniti modi, stabilire una [coordinatazione cartesiana (generalmente) obliqua]. Il caso della [retta orientata] $H_1 \subset H$ è elementare: prefissato su tale H_1 un punto O, e avendo scelto una [unità di misura] $|U|$ per le lunghezze, la coordinata cartesiana (rispetto a O e a $|U|$) del generico punto X ($\in H_1$) è la [distanza con segno], nell'unità di misura $|U|$, tra O e X, e quindi un numero reale ²⁶ in corrispondenza biunivoca ²⁷ con X. Esiste dunque una ben definita corrispondenza biunivoca (\leftrightarrow)

²⁵ In italiano (ma non soltanto), nel passato la geometria analitica fu per lo più detta "geometria algebrica". Più accuratamente, in inglese essa era ed è denominata "coordinate geometry", cioè "geometria con coordinate", o "geometria coordinata" (absit iniuria verbis). Tuttavia da parecchio tempo l'espressione "geometria algebrica" ha un significato assai diverso da quello corrente di "geometria analitica". A partire da circa la metà del secolo XIX, con le ricerche sugli invarianti algebrici e le trasformazioni birazionali (Cayley (Arthur, 1821-1895), Sylvester (James, 1814-1897)) quella che continuò a chiamarsi geometria algebrica venne infatti progressivamente ad assumere un carattere sempre più astratto, culminando nell'opera di matematici del XX secolo come Weil (André, 1906-1998), Dieudonné (Jean, 1906-1992) e Grothendieck (Alexandr, 1928-). La geometria algebrica in senso moderno studia certi oggetti geometrici che sono connessi con gli anelli commutativi, cioè le "varietà algebriche" e le loro successive generalizzazioni (gli "schemi" e gli "spazi algebrici"). Informalmente, potremmo anche dire che la geometria algebrica studia le equazioni algebriche da un punto di vista geometrico- astratto. In questo libro non occorrerà ricorrere a definizioni/concetti propri della geometria algebrica nella sua accezione moderna.

²⁶ In realtà ogni misura empirica dovrebbe essere un numero razionale. Il passaggio ai reali è necessario agli effetti teoretici (scoperta degli irrazionali), ma non ha rilevanza pratica nei confronti delle correnti applicazioni computazionali. Vale la pena di ricordare a questo proposito che, salvo una loro ristretta classe (quella dei cosiddetti reali "finitamente descrivibili", o "computabili"), i reali generici sono ineffabili in un linguaggio con parole di lunghezza finita, e che quindi possiamo calcolare (in senso numerico) soltanto con i razionali.

²⁷ La biunivocità di questa corrispondenza segue dagli assiomi hilbertiani di continuità. Ciò vale anche con riferimento agli analoghi e seguenti casi di un piano e dell'intero spazio.

tra H_1 e \mathbf{R} , e tra le strutture su di essi definite (la struttura d'ordine e la struttura metrica)²⁸, che rende H_1 e \mathbf{R} isomorfi²⁹ rispetto a quelle strutture.

Passando al caso del piano $H_2 \subset H$, mediante certi assiomi del sistema di Hilbert³⁰ si dimostra che, date in H_2 due rette orientate distinte e incidenti in (\equiv aventi in comune) O , diciamo l_1 e l_2 , ed un generico punto X , esiste esattamente una (retta parallela a) l_1 ed esattamente una retta parallela a l_2 (entrambe in H_2), passanti per X , le quali intercettano su (\equiv hanno in comune con) l_2 e rispettivamente su l_1 punti X'_2, X'_1 unicamente determinati. Le distanze con segno ξ_1 da O a X'_1 e ξ_2 da O a X'_2 , in certe unità di misura $|U_1|, |U_2|$ (in generale diverse), sono allora le coordinate cartesiane, generalmente oblique, di X ; e la coppia ordinata di reali $\langle \xi_1, \xi_2 \rangle$ è in corrispondenza biunivoca con X . In questo modo si stabilisce ancora una corrispondenza biunivoca (\leftrightarrow) tra H_2 e \mathbf{R}^2 , e tra le strutture (di ordine e metrica) su di essi definite, che similmente rende H_2 e \mathbf{R}^2 isomorfi rispetto a quelle strutture.³¹ In modo ancora analogo si procede nel caso dello spazio H , partendo da tre sue rette orientate incidenti nel punto O e (non-coplanari) l_1, l_2, l_3 , ogni coppia delle quali determina univocamente un corrispondente (piano coordinato) $\pi_{12} = (l_1, l_2), \pi_{23} = (l_2, l_3), \pi_{31} = (l_3, l_1)$. Per il generico punto X passa allora esattamente un (piano parallelo a) ciascuno dei tre piani $\pi_{12}, \pi_{23}, \pi_{31}$, che intercetta sulla corrispondente retta trasversa (l_3 per il piano parallelo a π_{12} , ecc.) un corrispondente punto unicamente determinato. (Naturalmente tutti i punti nominati appartengono a H e tutte le rette e i piani vi sono inclusi.) Le distanze con segno di questi punti da O , in certe unità di misura generalmente diverse tra loro, sono le coordinate cartesiane, generalmente oblique, di X , diciamo ξ_1, ξ_2, ξ_3 ; e la terna ordinata di reali $\langle \xi_1, \xi_2, \xi_3 \rangle$ è in corrispondenza biunivoca con X . Ancora, si stabilisce così una corrispondenza biunivoca (\leftrightarrow) tra H e \mathbf{R}^3 e tra le strutture su di essi

²⁸ Affidiamo per il momento all'intuizione le corrispondenti nozioni di "orientamento" (di una retta) e di "distanza" (tra due punti di una retta).

²⁹ Cioè, sotto la biiezione \leftrightarrow gli enunciati veri della teoria formalizzata di H_1 sono enunciati veri di quella di \mathbf{R} , e viceversa.

³⁰ Precisamente, se si utilizzano due teoremi della geometria proiettiva noti come teorema di Pappo (fine III sec. d. C.) e di Pascal (Blaise, Clermont-Ferrant Fr. 1623, Parigi 1662) e rispettivamente di Desargues (Girard, 1591-1661) (a loro volta deducibili nel sistema hilbertiano), oltre ai convenienti termini e relazioni occorrono soltanto alcuni assiomi di incidenza e di ordine, nonché l'assioma delle parallele.

³¹ Anche il piano H_2 può essere orientato in uno di due modi possibili; ma preferiamo rimandare la discussione generale del concetto di orientamento (sia per H_2 che per H , o addirittura per H_n) alla S.sez 1.4.2, ad esso dedicata. Inoltre anche su H_2 (o su H_n) può essere definita una distanza assoluta (o semplicemente "distanza"), in unità $|U_1| = |U_2| = |U|$ (vedi appresso) tra due punti distinti. Come nel caso della retta H , anche in quello del piano H_2 (o in generale di uno spazio H_n) le nozioni standard di orientamento e di distanza (assoluta) sono indipendenti tra loro: cioè se si cambia l'orientamento del piano la distanza tra i due punti non cambia. Ad esempio, e come si vedrà, la distanza tra $X \neq O$ e O è data da $(\xi_1^2 + \xi_2^2 + 2\xi_1\xi_2\cos\omega)^{1/2}$, dove ω è l'angolo (con segno) dalla retta orientata l_1 alla retta orientata l_2 . (Si verifica che il radicando è intrinsecamente > 0 .) Cambiando l'orientamento del piano, ω cambia segno (vedi S.sez. 1.4.2), mentre ξ_1 e ξ_2 non cambiano; quindi la distanza non cambia. Inoltre essa non cambia nemmeno cambiando l'orientamento di una, o di entrambe, le rette l_1, l_2 . Nel primo caso, ad es. cambiando l'orientamento di l_1 , ξ_1 cambia segno, mentre ω diventa $\omega \pm \pi$, e quindi $\cos\omega$ cambia segno; nel secondo caso ω non cambia mentre ξ_1 e ξ_2 cambiano entrambe segno. La conclusione è ancora che la distanza tra X e O non cambia.

definite, che rende H e \mathbf{R}^3 isomorfi rispetto a quelle strutture. Si potrebbe anche proseguire con riferimento a $H_{n>3}$ e $\mathbf{R}^{n>3}$.

Nella loro interpretazione naturale, le parole astratte del sistema di Hilbert (o di qualunque altro sistema che formalizzi la geometria euclidea) corrispondono a termini ed enunciati idealizzati relativi allo spazio fisico della esperienza comune; e in particolare, gli enunciati veri corrispondono, nelle solite approssimazioni osservative, a verità empiriche idealizzate. In forza dell'isomorfismo tra H e \mathbf{R}^3 (rispetto alle strutture su di essi definite), potremo ripetere questa affermazione sostituendo \mathbf{R}^3 ad H (oppure, riferendoci al piano o alla retta fisici della esperienza comune, sostituendo \mathbf{R}^2 a H_2 , e rispettivamente \mathbf{R} a H_1). In particolare, alla distanza tra due punti X e Y di H , diciamo $d(X,Y)$, definita come la [lunghezza] (assoluta, e misurata in una unità di misura $|U|$) del [segmento di estremi] X e Y , corrisponde la “distanza pitagorica” d_{pit} ³² tra le terne di coordinate cartesiane [ortogonali] (con comune unità $|U|$) $x \leftrightarrow X$ e $y \leftrightarrow Y$ ³³, e la uguaglia numericamente. Ma X e Y corrispondono biunivocamente a certi punti empirici X^* e Y^* , e $d(X,Y)$ è uguale, entro le approssimazioni osservative, alla distanza empirica tra X^* e Y^* , diciamo $\text{dist}(X^*,Y^*)$, se l'unità del regolo corrisponde a $|U|$. In conclusione, in questo caso abbiamo

$$(1) \quad d(X,Y) = d_{\text{pit}}(x,y) \simeq \text{dist}(X^*,Y^*),$$

dove \simeq significa “uguale a meno delle approssimazioni osservative”. Questo è un esempio canonico di isomorfismo (un isomorfismo metrico, o “isometria”) tra H , \mathbf{R}^3 e (con le solite approssimazioni) lo spazio fisico.

Passando da rette coordinate cartesiane ortogonali con comune unità $|U|$ a rette coordinate generiche, con origine e unità anch'esse generiche, le vecchie coordinate cartesiane ortogonali con comune unità $|U|$ subiscono una trasformazione lineare affine, e quindi d_{pit}^2 diventa una forma quadratica definita positiva nelle differenze delle nuove coordinate omologhe, i cui sei coefficienti sono unicamente determinati dagli angoli (nell'intervallo $[0,\pi]$) tra le rette coordinate, e dalle tre unità di misura su di esse adottate; quindi quei coefficienti non variano per [rotazioni proprie] della nuova terna coordinata. Se le tre unità di misura restano uguali ad $|U|$, il *valore* della forma quadratica si mantiene quello di d_{pit} ; e se la nuova terna coordinata è anch'essa ortogonale, i coefficienti si riducono ai sei ($6 = n(n+1)/2$ per $n = 3$) **simboli di Kronecker** (Leopold, (1823-1891)) δ_{ik} ($\delta_{ik} = 1$ se $i = k$ e $\delta_{ik} = 0$ se $i \neq k$, con $i, k = 1,2,3$). Quindi la forma resta del tipo

³² Ricordiamo che la distanza pitagorica tra due n -ple ordinate di reali $x \equiv \langle x_1, \dots, x_n \rangle$ e $y \equiv \langle y_1, \dots, y_n \rangle$ è per definizione la radice quadrata della somma dei quadrati delle differenze tra i reali i -mi x_i e y_i ($i = 1, \dots, n$). Ad esempio per $n = 2$ si ha $d_{\text{pit}}(x,y) = :[(x_1-y_1)^2 + (x_2-y_2)^2]^{1/2}$. È ben noto, e comunque si prova facilmente, che tale d_{pit} soddisfa ai quattro assiomi metrici canonici – non-negatività, simmetria, identità tra le n -ple con distanza nulla e disuguaglianza triangolare (quest'ultima si prova a sua volta mediante la disuguaglianza di Minkowski, conseguenza di quella di Cauchy-Schwarz).

pitagorico. In conclusione, passando da una terna cartesiana ortogonale con comune unità $|U|$ ad un'altra (comunque dislocata e orientata) con la stessa unità, la distanza pitagorica tra le due terne ordinate di reali relative ai due riferimenti cartesiani ortogonali con la stessa comune unità, diciamo x e x' e rispettivamente y e y' , mantiene lo stesso valore numerico. Questo valore $d_{\text{pit}}(x,y) = d_{\text{pit}}(x',y')$ è a sua volta uguale a $d(X,Y)$, ove x, x' corrispondono a X e y, y' corrispondono a Y ; e se il regolo unitario ha la lunghezza $|U|$, sempre sotto le solite approssimazioni, a $\text{dist}(X^*,Y^*)$. Risultati del tutto analoghi si hanno riferendosi a piani o rette, la sola differenza essendo che si avranno allora coppie ordinate di reali, o rispettivamente singoli reali, piuttosto che terne ordinate di reali.³⁴

La coordinatazione cartesiana dello spazio (o del piano, o della retta) empirici – i primi due supposti euclidei –, e quindi il passaggio dalla loro geometria euclidea *sintetica*, intuitiva o matematizzata, a quella con coordinate (o *analitica*) ha reso immensi servizi; ma le difficoltà legate al dimostrare la possibilità teoretica di questo passaggio non sono solitamente apprezzate nella giusta misura. Ciò in quanto la scuola secondaria ci ha formati con l'idea intuitiva di quella possibilità, al punto che siamo abituati a considerarla come qualcosa di completamente ovvio; e addirittura al punto che la maggior parte delle persone con una formazione scientifica-esatta di livello universitario pensa alla versione euclidea della geometria *né più né meno come a \mathbf{R}^3* con la metrica della distanza pitagorica (o con una metrica equivalente), cioè identifica (e confonde) la teoria *sintetica* formalizzata con il suo modello numerico canonico, unico a meno di isomorfismi.³⁵ In conclusione, la coordinatazione dello spazio e la definitiva divulgazione della teoria dei numeri reali sembrano aver sottratto alle grandi conquiste formalistiche dei geometri “euclidei” (a cavallo tra il XIX e il XX secolo) quella popolarità che esse certamente meritano.

³³ Ovviamente la nozione di “terna cartesiana ortogonale con unità $|U|$ ” riceve una precisa definizione nel sistema assiomatico di Hilbert.

³⁴ La giustificazione degli asseriti degli ultimi quattro paragrafi riposa su teoremi della geometria euclidea, e quindi su una assiomatizzazione di questa (ad es. quella di Hilbert, o quella equivalente presentata nella Sez 1.2); su di essi ritorneremo ampiamente.

³⁵ Gran parte delle moderne teorie fisico-matematiche sono presentate in modo analogo. Qualcuno ha osservato che dopotutto anche questa è una didattica *metaforica* (vedi la Presentazione 0.0), in quanto all'oggetto empirico si sostituisce il suo modello matematico; ma l'obiezione contiene a nostro avviso un equivoco fondamentale, perché ben difficilmente le proposizioni matematiche potrebbero considerarsi come oggetti empirici. Anzi, si potrebbe facilmente controbattere che una possibile descrizione del mondo fenomenico via siffatte “metafore matematiche” (per così dirle) è quanto di più ambizioso ci si possa proporre nei riguardi di una sua rappresentazione predittiva.

1.2) UNA FORMALIZZAZIONE METRICA DELLA GEOMETRIA EUCLIDEA

1.2.1) INTRODUZIONE

Mirando ad un opportuno approfondimento di quanto abbiamo esposto a grandi linee nella precedente sezione circa l'isomorfismo tra $H \equiv H_3$ e \mathbf{R}^3 (o tra H_2 e \mathbf{R}^2 , ecc.) come modelli (teorici) euclidei dello spazio empirico (o del piano empirico, ecc.), l'idea più naturale sembrerebbe quella di offrire al lettore una sintesi del sistema assiomatico di Hilbert. Ci siamo tuttavia rapidamente convinti della modesta qualità pedagogica di una tale scelta, anche prescindendo dalle difficoltà che essa comporterebbe per riuscire veramente efficace. Da una parte, infatti, le Grundlagen der Geometrie sono facilmente accessibili al lettore desideroso di cimentarsi con il loro studio; e dall'altra, ci è parso che il carattere sostanzialmente astratto di quelle pagine potesse riuscire ostica agli occhi di un tipico fisico-matematico. Per meglio spiegarci, ci è parso che la formalizzazione della geometria euclidea più congeniale alla mentalità fisico-matematica dovesse essere fondata su un insieme per quanto possibile ristretto di nozioni primitive, dotate della massima suggestione intuitiva e aventi controparti empiriche verificabili nei modi più naturali; e che la assiomatizzazione di Hilbert non fosse necessariamente il miglior punto di partenza da questo punto di vista. Ad esempio: la cosiddetta relazione dell'“essere tra” (di un punto rispetto ad altri due, tutti e tre essendo distinti e allineati tra loro), che pure è la prima delle relazioni primitive di Hilbert, non può utilmente tradursi in termini di operazioni empiriche idealizzate senza inoltrarsi alquanto nello sviluppo della teoria; e precisamente fino a quando, avendone introdotto un congruo numero di assiomi, l'interpretazione di quella relazione in termini di “misura di un segmento” non sia diventata intuitivamente trasparente. Sembra ovvio che un fisico-matematico preferirebbe procedere subito alla verifica di una semplice relazione numerica tra le distanze tra le tre coppie di punti distinti in oggetto, naturalmente avendo introdotto la nozione di distanza come primitiva, e avendone definita la “misura empirica” mediante operazioni idealizzate di esecuzione inequivoca – almeno in linea di principio.

L'idea di porre la nozione di distanza alla base della geometria risale a Riemann (Bernhard, Reselenz Germ. 1826, Selasca It. 1866), e avrà un ruolo centrale nello sviluppo delle geometrie non euclidee (e in particolare, della geometria relativistica). Più precisamente, tra le tante straordinarie intuizioni che offrì alla sua udienza di Gottinga nella esposizione della sua tesi di abilitazione (1854), Riemann suggerì che la geometria si potesse completamente sviluppare partendo dalla nozione di “distanza” (di un punto da un punto) e da quella di “verso” (un ordine tra due punti

distinti) – oltre che ovviamente da quella stessa di “punto”. Se si considera che la scelta di un ordine tra due cose distinte non è dopo tutto un concetto specificamente geometrico, l’idea di Riemann si riduce insomma ad identificare nella coppia primitiva {punto (un termine), distanza (una relazione, coinvolgente una coppia, a priori ordinata, di punti e un numero reale)} una possibile base concettuale della geometria. Partendo da questa coppia di nozioni primitive, secondo Riemann doveva essere possibile da una parte ricostruire tutta la geometria di Euclide ammettendone il quinto postulato, e dall’altra inventare nuove geometrie di pari dignità logica rifiutandolo, ed eventualmente sostituendolo con un diverso e conveniente assioma. Sempre secondo Riemann, la cui impressionante capacità di previsione non potrà mai essere sopravvalutata, queste nuove geometrie avrebbero potuto avere applicazioni indipendenti, *e segnatamente alla fisica*. Una proposta così radicale ed apertamente ispirata ad una visione “fisica” del problema non poteva avere grande seguito tra i matematici dell’epoca; e difatti essa fu raccolta soltanto oltre mezzo secolo più tardi, e precisamente da M. Pieri nel 1908¹. Più o meno da allora in avanti, cioè dopo che le Grundlagen hilbertiane erano divenute ormai di comune dominio,² lo sviluppo di assiomatizzazioni della geometria (euclidea e non) *alternative* a quella di Hilbert si era infatti ridotto al livello di un esercizio non proibitivamente difficile.

Proprio da qui prende le mosse la nostra presente opzione didattica, quella cioè di illustrare una assiomatizzazione della geometria euclidea fondata sulla coppia primitiva {punto, distanza}, o **assiomatizzazione metrica**, intenzionalmente indipendente da precedenti operazioni dello stesso tipo. Il conseguimento di questo obiettivo sarà anche una preziosa occasione per offrire al lettore un’idea *di come si sviluppa una teoria fisico-matematica non strettamente formalizzata* (cioè non identificata con mere sequenze di stringhe di segni privi di significato, contenuti in un dato alfabeto.)³

Venendo dunque alla nostra assiomatizzazione metrica, innanzitutto presupporremo in essa l’intera teoria degli insiemi *sin dal principio*, come non è sempre fatto in altre assiomatizzazioni. Formuleremo quindi convenienti assiomi sull’unico enunciato specifico disponibile, quello che afferma che “la distanza di un dato punto da un dato punto è uguale ad un ben definito numero reale non negativo” (naturalmente l’aritmetica dei reali è contenuta nella teoria degli insiemi). Su questa base, introdurremo poi termini ed enunciati derivati, e formuleremo altri assiomi coinvolgenti in generale tutti gli oggetti fino a quel momento introdotti, nell’intento di pervenire a termini ed

¹ M. Pieri, “La geometria elementare istituita sulle nozioni di punto e di sfera”, Mem. Mat. Fis. Società Italiana delle Scienze, ser. 3. 15, (1908).

² Su questo cospicuo ritardo ebbe probabilmente un peso l’ancora immaturo stadio di sviluppo della logica formale a metà Ottocento. La situazione cambierà radicalmente soltanto verso la fine del secolo, in particolare con le ricerche di Frege, e più tardi, di Hilbert stesso, in parte proprio con le Grundlagen der Geometrie.

enunciati consistentemente sovrapponibili a quelli accolti come primitivi nella assiomatizzazione hilbertiana. Quest'ultima ci guiderà infine al completamento della teoria in condizioni di relativa sicurezza. Ovviamente non sarebbe ragionevole proporsi di sviluppare *qui* un programma del genere nella sua interezza: vale a dire, se oltre alla prova dell'equivalenza delle due teorie – quella di Hilbert e la presente –, esso prevedesse l'enunciazione, o addirittura la dimostrazione esplicita, di tutti i teoremi di qualche importanza derivabili nella seconda mediante le usuali regole della deduzione (o inferenza) naturale. Tuttavia molti passi, e tutti quelli essenziali, possono essere fatti in questa direzione senza grandi difficoltà.⁴

1.2.2) I PRIMI NOVE ASSIOMI E LA GEOMETRIA “SPECIALE” DELLA RETTA

Come dicevamo, la assiomatizzazione della geometria euclidea che ci accingiamo ad illustrare presuppone la teoria degli insiemi, della quale useremo liberamente le nozioni intuitive e il linguaggio “naturale”, nonché le notazioni standard (vedi App. Gen. A). Partiremo dunque da un **insieme-universo**, che per il momento denoteremo con S , di oggetti primitivi che diremo suoi **punti** e denoteremo $X, Y, Z, \dots, X_1, X_2, \dots$ ecc.⁵, e da un'applicazione $d: S^2 \rightarrow \mathbf{R}_{\geq 0}$ ($\mathbf{R}_{\geq 0} \equiv$ insieme dei reali non-negativi) che diremo **distanza** (del primo punto dal secondo punto della coppia ordinata degli argomenti di d).⁶ Questi oggetti primitivi (i punti e il loro insieme S , l'insieme dei reali non-negativi $\mathbf{R}_{\geq 0}$ e l'applicazione distanza di S^2 in $\mathbf{R}_{\geq 0}$), nonché i termini e gli enunciati derivati che verremo via via introducendo, saranno per definizione soggetti a certi assiomi specifici progressivamente elencati in quanto segue in questa S.sez. 1.2.2. Denoteremo gli assiomi specifici con la lettera A munita di pedice numerico (e in certi casi, di apici), o anche con acronimi ad hoc (usando il grassetto all'atto della prima enunciazione degli assiomi), e man mano segnaleremo alcune delle loro “conseguenze naturali”, o teoremi, di maggior interesse ai nostri fini. I teoremi menzionati nel testo e non dimostrabili in modo banale saranno riportati tra « ... ». ⁷ In pratica, quasi di nessun teorema sarà data la dimostrazione; ma è inteso che la sua convalida *non* richieda l'uso di assiomi che lo *seguono* nel testo. Inoltre, in linea di massima ogni teorema riportato

³ Questo livello di formalizzazione è quello praticato nella matematica corrente, cioè nella matematica tout court salvo quella orientata alle applicazioni logiche del calcolatore.

⁴ Seppur in senso assai diverso, il ruolo protagonista della nozione di distanza permarrà negli sviluppi relativistici dei prossimi capitoli.

⁵ I punti che compaiono (in numero finito) nei quantificatori \exists e \forall si sottintenderanno quindi correnti su S , oppure (non tacitamente, vedi App. Gen. A) su suoi sottoinsiemi propri.

⁶ Sottolineiamo che l'applicazione d di S^2 è per il momento *in*, e non *su* $\mathbf{R}_{\geq 0}$.

presupporrà, tra (alcuni de)gli assiomi che lo precedono, l'uso dell'*ultimo* di essi.⁸ Queste circostanze faciliteranno molto il lavoro del lettore che volesse esercitarsi a ricostruire le varie dimostrazioni per suo conto (a nostro giudizio, *quasi mai difficili*).⁹ A seconda della convenienza, useremo indifferentemente il linguaggio naturale o quello simbolico (in questo caso, con le notazioni dell'App. Gen. A).

Il primo assioma corrisponde ad una richiesta ovvia:

(A₀) “In S esistono (almeno) due punti distinti”.

Seguono gli assiomi:

(A₁) $\forall(X,Y)\{d(X,Y) = d(Y,X)\}$

(quindi $d(X,Y)$ potrà dirsi “distanza *tra* X e Y”, ovvero d è simmetrica), e

(A₂) $\forall(X,Y)\{d(X,Y) = 0 \Leftrightarrow X = Y\}$.

Per arbitrari X,Y,Z , la notazione $[X,Z,Y]$ sarà usata come abbreviazione della relazione ternaria “ $\neq(X,Y,Z) \wedge (d(X,Y) = d(X,Z) + d(Z,Y))$ ”; quindi $[X,Z,Y] \Leftrightarrow [Y,Z,X]$. $[X,Z,Y]$ si leggerà “Z è tra X e Y” (ciò definisce la relazione **essere tra** di un punto rispetto ad altri due). Seguono altri due assiomi alquanto simili tra loro, e alcune nuove definizioni e conseguenze ad essi legate. Il primo è:

(A'₃) $\forall(X,Y)\{X \neq Y \Rightarrow \exists(Z)\{[X,Z,Y]\}$ ¹⁰.

Per dati X,Y distinti, l'insieme $\{Z|[X,Z,Y]\}$ (che ora sappiamo avere “qualche” elemento distinto da X e da Y) si dirà **segmento aperto di estremi** X e Y , e si denoterà (X,Y) ($\equiv (Y,X)$); l'insieme $\{Z|[X,Z,Y]\} \cup \{X\}$ si dirà **segmento di estremi** X,Y **chiuso in X e aperto in Y**, e si denoterà $[X,Y)$ ($\equiv (Y,X]$); e infine l'insieme $\{Z|[X,Z,Y]\} \cup \{X,Y\}$ si dirà **segmento chiuso di estremi** X,Y , e si denoterà $[X,Y]$ ($\equiv [Y,X]$).¹¹ Da (A'₃) segue facilmente che «la cardinalità di un

⁷ L'uso standard è quello di far precedere al generico teorema (e talvolta anche agli assiomi) il simbolo \vdash , introdotto da Frege. Pur essendo molto comodo nelle formalizzazioni totalmente simbolizzate, esso è parso qui sconsigliabile, perché non segnalerebbe chiaramente dove l'enunciato del teorema finisce.

⁸ Quest'ultima condizione è talvolta difficile da soddisfare volendo dare all'esposizione un ragionevole carattere di organicità.

⁹ Esistono ormai esempi di interi trattati matematici presentati in questa chiave. Ne citiamo un paio, anche per la loro pertinenza ad alcuni contenuti di questo libro: 1) I. Glazman, Y. Liubitch, “Analyse linéaire dans les espaces de dimensions finies”, 400 pp, trad. francese in Éditions MIR, Moscou (1974); 2) M. Berger, “A Panoramic View of Riemannian Geometry”, 826 pp, Springer 2^a ristampa (2007).

¹⁰ Ricordiamo che $\exists(Z)\{\dots\}$ si traduce nel linguaggio naturale con “per qualche z , $\{\dots\}$ ”, o anche meglio, con “esiste *almeno* un Z per cui $\{\dots\}$ ”. Del resto, anche nel linguaggio matematico corrente la locuzione “esiste un oggetto per cui \dots ” significa sempre “esiste *almeno* un oggetto per cui \dots ”.

¹¹ Nota bene: le scritture (X,Y) , $[X,Y)$, $[X,Y]$ *non* sono definite, e quindi sono prive di senso, se $X = Y$. La scelta dei simboli (X,Y) , $[X,Y)$, e di altri analoghi che compariranno più tardi, non è soddisfacente, perché suggerisce l'esistenza di un ordine, che non c'è, tra X e Y ; ma non vi sono alternative tipograficamente accettabili. Ricordiamo che in teoria degli insiemi è importante distinguere con notazioni diverse gli insiemi finiti *non ordinati* da quelli *ordinati*. Una convenzione abbastanza diffusa, alla quale cercheremo di uniformarci nel seguito, è quella di porre i simboli con cui si denotano gli oggetti dell'insieme considerato tra parentesi $\{ \}$ (o talvolta $()$) nel primo caso e $\langle \rangle$ nel secondo, separandoli con virgole. Se poi l'insieme è costituito da oggetti distinti da un indice numerico naturale, tale indice può

segmento aperto, e quindi quella di S che lo include, è (almeno) quella di N (insieme dei naturali), o \mathbb{N}_0 ».

Il secondo assioma della coppia annunciata è:

$$(A''_3) \quad \forall(X,Y)\{X \neq Y \Rightarrow \exists(Z)\{[Z,X,Y]\}\},$$

quindi anche lo stesso con $\exists(Z)\{[X,Y,Z]\}$ come conseguente. Per arbitrari X,Y,Z, la relazione ternaria (simmetrica in X,Y,Z) $[X,Y,Z] \vee [Z,X,Y] \vee [Y,Z,X]$ si abbrevierà in $\text{Col}(X,Y,Z)$ e si leggerà “la terna $\{X,Y,Z\}$ è **collineare**”. In forza di (A'_3, A''_3) , esistono terne (di punti) collineari, e insiemi numerabili di punti collineari a tre a tre.¹² Conseguenze immediatamente dalle definizioni l’asserto:

$$(1) \quad \forall(X,Y,Z)\{[X,Z,Y] \Rightarrow \neg[Z,X,Y]\},$$

quindi anche lo stesso con $\neg[X,Y,Z]$ come conseguente. In forza di (1), nella definizione di $\text{Col}(X,Y,Z)$ le disgiunzioni *inclusive* \vee si potranno sostituire con disgiunzioni *esclusive* $\underline{\vee}$.¹³ Per dati X,Y (distinti), l’insieme simmetrico in X,Y $\{Z | \text{Col}(X,Y,Z)\} \underline{\cup} \{X,Y\}$ (ove $\underline{\cup}$ è l’unione esclusiva) si dirà **retta per X,Y**, e si denoterà $\mathcal{R}(X,Y)$ ($\equiv \mathcal{R}(Y,X)$). Un sottoinsieme p (di S) per il quale esistono due punti distinti X,Y tali che $p = \mathcal{R}(X,Y)$ si dirà una **retta**.^{14, 15} Ovviamente esistono rette. Nel seguito, denoteremo rette con p, q, r, ..., $p_1, p_2 \dots$ ecc. Almeno una retta contiene due punti distinti dati X,Y, la $\mathcal{R}(X,Y)$; tale retta si dirà anche “passante per X,Y”. Se tre punti distinti appartengono ad una retta, sono collineari; e viceversa, se tre punti sono collineari, esiste una retta cui essi appartengono. Quindi tre punti distinti sono collineari sse (se e solo se) appartengono ad una retta. Vale peraltro anche l’assioma:

(A₄) “Non esistono due rette distinte passanti entrambe per due punti distinti dati” (o se si preferisce, “per due punti distinti dati passa *al più* una retta”).

Quindi per due punti distinti X,Y passa esattamente una retta, che è la $\mathcal{R}(X,Y)$; ovvero, se X,Y sono punti distinti della retta p, allora $p = \mathcal{R}(X,Y)$. Evidentemente, due rette distinte p, q sono disgiunte aut hanno esattamente un punto in comune; in quest’ultimo caso, cioè se

pensarsi come *parte del simbolo* con cui si denota l’oggetto, oppure come denotante un ordine, di regola quello crescente degli indici. Si scriverà dunque ad es. $\{a_1, a_2 \dots a_n\}$, o $\{a_i\}_{i=1, \dots, n}$ nel primo caso, e $\langle a_1, a_2 \dots a_n \rangle$ o $\langle a_i \rangle_{i=1, \dots, n}$ nel secondo. La notazione (X,Y) per il segmento aperto di estremi X e Y è coerente con la regola enunciata perché vi è corrispondenza biunivoca tra i segmenti e le coppie di punti che ne sono gli estremi.

¹² Si ricordi che se A e B sono formule di una teoria logica quantificata, $\exists(x)\{A \vee B\} \Leftrightarrow \exists(x)A \vee \exists(x)B$, e che $\forall(x)\{A \wedge B\} \Leftrightarrow \forall(x)A \wedge \forall(x)B$.

¹³ Nella lingua italiana l’“o” inclusivo si scrive spesso “e/o”, ma non esiste una precisa congiunzione per designare l’“o” esclusivo. Nel seguito, sarà comodo usare convenzionalmente il “vel” per il primo e l’“aut” per il secondo. Quindi “vel” significherà “almeno una” di due cose, e “aut” “esattamente una” di esse. In inglese, si usa spesso “or” per “vel” e “either ... or” per aut.

¹⁴ La “definizione” di Euclide di retta (anch’essa vuota) recita: “linea retta è quella che giace ugualmente rispetto ai suoi punti”. Va anche ricordato che, per Euclide, “linea retta” è il nostro segmento rettilineo.

¹⁵ Il lettore è fortemente dissuaso dall’acceptare passivamente l’immagine intuitiva di retta (e in generale, ancora una volta, *qualunque* immagine intuitiva). Per quanto euristicamente utile o indispensabile, ripetiamo, l’intuizione è inessenziale dal presente punto di vista. Ad esempio, come è stata fin qui definita la retta è un insieme soltanto *numerabile* di punti collineari.

$\exists!(Z)\{p \cap q = \{Z\}\}$, esse si diranno **incidenti** (in quel punto). Non possiamo affermare, per il momento, che esistano rette distinte, e a maggior ragione disgiunte aut incidenti.

Per dati X, Y (distinti), l'insieme $\{Z|[X, Z, Y] \vee [X, Y, Z] \vee Z=Y\}$ si dirà **semiretta**, o anche **raggio, di origine X e contenente Y** , e si denoterà $\mathcal{SR}(X|Y)$; evidentemente, $\mathcal{SR}(X|Y) \subset \mathcal{R}(X, Y)$, e $\mathcal{SR}(X|Y) \cap \mathcal{SR}(Y|X) = (X, Y)$. Sempre per dati X, Y (distinti), l'insieme $\mathcal{SR}^*(X|Y) =: (\mathcal{R}(X, Y) \setminus \mathcal{SR}(X|Y)) \setminus \{X\}$ (\setminus è il segno di sottrazione insiemistico) si dirà **semiretta complementare di $\mathcal{SR}(X|Y)$** . Si verifica subito che la semiretta complementare di $\mathcal{SR}^*(X|Y)$ è $\mathcal{SR}(X|Y)$. Risulta inoltre $\mathcal{SR}^*(X|Y) = \{Z|[Z, X, Y]\}$. Dalle definizioni segue facilmente che $\mathcal{SR}(X|Y) \cap \mathcal{SR}^*(X|Y) = \emptyset$ e $\mathcal{SR}(X|Y) \cup \mathcal{SR}^*(X|Y) = \mathcal{R}(X, Y) \setminus \{X\}$. Ovviamente, se due punti distinti X, Y appartengono ad una retta p , sia i segmenti (di qualunque tipo) di estremi X e Y che le semirette $\mathcal{SR}(X|Y)$ e $\mathcal{SR}^*(X|Y)$ sono tutti inclusi in p . Una **semiretta di origine X** è un insieme del tipo $\mathcal{SR}(X|Y)$ (o equivalentemente, del tipo $\mathcal{SR}^*(X|Y)$) per qualche coppia di punti distinti X, Y ; se tali X, Y appartengono alla retta p , parleremo di **semiretta di origine X della retta p** . Denoteremo semirette con $p', q' \dots, p'_1, p'_2 \dots$ ecc. Le notazioni $\mathcal{SR}(X|Y)$ e $\mathcal{SR}^*(X|Y)$ presuppongono tacitamente che $X \neq Y$.

Si consideri ora la relazione binaria sull'insieme delle semirette di p : “ $p' \subset q' \vee q' \subset p'$ ”.¹⁶ Sussiste il teorema: «“ $p' \subset q' \vee q' \subset p'$ ” è una (relazione di) equivalenza», rispetto alla quale l'insieme delle semirette di p si decompone *in esattamente due* classi (disgiunte)¹⁷ che si diranno **orientamenti** (o **versi**) di p . Due semirette di p dello stesso orientamento si diranno **equiverse**, e **antiverse** nel caso contrario. Una semiretta di p è unicamente determinata dalla sua origine e dall'orientamento cui appartiene. La coppia costituita da una retta e da uno dei due orientamenti delle sue semirette si dirà **retta orientata**; una retta orientata è quindi univocamente definita dando quella retta e una delle semirette dell'orientamento prescelto. Si comprende allora che «la relazione di inclusione sull'insieme delle semirette *equiverse* di una retta orientata è una relazione di ordine totale (o t-ordine)».¹⁸ Su una retta orientata, una semiretta dell'orientamento prescelto è completamente determinata dalla sua origine; quindi, se su quella retta orientata denotiamo con p'_X e p'_Y le due semirette (equiverse) del suo orientamento aventi origine X e rispettivamente Y , è chiaro che la relazione “ $p'_Y \subset p'_X$ ” può equivalentemente scriversi come relazione di t-ordine tra punti, diciamo $X \preceq Y$. In conclusione abbiamo con ciò ricondotto la nozione di retta orientata, e di

¹⁶ Seguendo l'uso oggi prevalente, si è scelto di usare il simbolo \subset per esprimere l'inclusione generalmente impropria, rinunciando al più appropriato \subseteq . Ciò ha il blando inconveniente di dover denotare altrimenti l'inclusione propria (\subsetneq), vedi anche l'App. Gen. A.

¹⁷ Ricordiamo che un insieme qualsiasi sul quale esiste una equivalenza viene da essa decomposto in classi che sono comunque *disgiunte*.

t-ordine (o di s-ordine) su di essa, a quella primitiva di distanza, oltre a quella (non specificamente geometrica) di “ordine tra cose (distinte)”. Un orientamento su una data retta può anche prescriversi dando una coppia di suoi punti distinti e un *ordine* su di essi (cioè, dando una coppia ordinata di suoi punti distinti). In questo caso, se la coppia ordinata (di punti distinti della retta) di riferimento è $\langle X_0, Y_0 \rangle$, potremo sempre decidere se una generica coppia ordinata $\langle X, Y \rangle$ (di punti distinti della stessa retta) ha orientamento uguale (“coppia equiversa”) aut opposto (“coppia antiversa”) a quello della coppia di riferimento. Se $\langle X, Y \rangle$ e $\langle X_0, Y_0 \rangle$ sono coppie di punti distinti (cioè $X \neq Y$ e $X_0 \neq Y_0$) della stessa retta, scriveremo $\langle X, Y \rangle \uparrow \uparrow \langle X_0, Y_0 \rangle$ [$\langle X, Y \rangle \uparrow \downarrow \langle X_0, Y_0 \rangle$] per significare che $\langle X, Y \rangle$ è equiversa [antiversa] a $\langle X_0, Y_0 \rangle$; ed è ovvio che

$$\langle X, Y \rangle \uparrow \uparrow \langle X_0, Y_0 \rangle \Leftrightarrow ((X < Y) \wedge (X_0 < Y_0)) \vee ((Y < X) \wedge (Y_0 < X_0))$$

$$[\text{che } \langle X, Y \rangle \uparrow \downarrow \langle X_0, Y_0 \rangle \Leftrightarrow ((X < Y) \wedge (Y_0 < X_0)) \vee ((Y < X) \wedge (X_0 < Y_0))].^{19}$$

Aggiungiamo ancora la seguente coppia di assiomi sull’“essere tra”, ciascuno dei quali coinvolge una quaterna di punti distinti, e che scriviamo trascurando di rendere espliciti gli universali:

$$(A'_5) \quad [X, Y, Z] \wedge [Y, Z, U] \Rightarrow [X, Y, U]$$

(quindi anche lo stesso con $[X, Z, U]$ come conseguente); e

$$(A''_5) \quad [X, Y, U] \wedge [Y, Z, U] \Rightarrow [X, Y, Z]$$

(quindi anche lo stesso con $[X, Z, U]$ come conseguente).

Essi permettono di dimostrare teoremi come i seguenti.

$$(2) \quad \ll \text{se } \neq(X, Y, Z, U), \text{Col}(X, Y, Z) \wedge \text{Col}(U, Y, Z) \Rightarrow \text{Col}(Y, X, U) \wedge \text{Col}(Z, X, U) \gg;$$

$$(3) \quad \ll \text{se } \neq(X, Y, Z), ((X, Y) \cap (Y, Z) = \emptyset) \Leftrightarrow [X, Y, Z] \gg;$$

$$(4) \quad \ll \text{Col}(X, Y, Z) \Rightarrow ((X, Y) \cap (Y, Z) \cap (X, Z) = \emptyset) \gg;$$

$$(5) \quad \ll \text{Col}(X, Y, Z) \Rightarrow (((X, Z) \setminus \{Y\}) \subset ((X, Y) \cup (Y, Z))) \gg;$$

$$(6) \quad \ll [X, Y, Z] \Rightarrow (((X, Z) \setminus \{Y\}) = ((X, Y) \cup (Y, Z))) \gg.$$

Non sarebbe inoltre difficile dimostrare, utilizzando (A'_5, A''_5) , che «la relazione binaria $\neg[U, X, V]$ sui generici punti U, V di una retta p privata del suo punto X è una equivalenza, e che tale $p \setminus \{X\}$ si decompone in esattamente due corrispondenti classi (disgiunte), che sono proprio le due semirette complementari di origine X ».

Possiamo ora definire la **congruenza di coppie** (arbitrarie) **di punti distinti**: diremo che “la coppia $\{X, Y\}$ è congruente alla coppia $\{Z, V\}$ se $d(X, Y) = d(Z, V)$ ”. Nel seguito converrà scrivere

¹⁸ Una breve rassegna di definizioni/teoremi sulle relazioni d’ordine è presentata come App. Spec. 1.A alla fine del presente capitolo.

¹⁹ Nella Sez. 1.4 sarà illustrato un criterio alternativo (ma equivalente) per giudicare della equiversità o antiversità delle coppie ordinate $\langle X, Y \rangle$ e $\langle X_0, Y_0 \rangle$ di una retta, passibile di una immediata e importante generalizzazione.

semplicemente $X|Y$ per la coppia non ordinata di punti *distinti* $\{X,Y\}$ (la notazione $X|Y$ implicherà cioè tacitamente $X \neq Y$ ²⁰). Denoteremo con $X|Y \approx Z|V$ la congruenza tra $X|Y$ e $Z|V$, quindi $X|Y \approx Y|X$. La congruenza tra coppie (non ordinate) di punti distinti è manifestamente una equivalenza, e quindi decompone l'insieme di tali coppie in corrispondenti classi (disgiunte) di equivalenza, che si diranno **coppie libere** (di punti distinti) e si denoteranno $\alpha, \beta, \dots, \alpha_1, \alpha_2, \dots$. In pratica, invece che a coppie di punti distinti ci si potrà equivalentemente riferire ai segmenti (di qualunque tipo) aventi per estremi quei punti. In questo caso le corrispondenti classi di equivalenza di segmenti congruenti si diranno **segmenti liberi**. Sebbene “coppie di punti distinti” e “segmenti” siano oggetti manifestamente diversi, essi potranno liberamente interscambiarsi nel presente specifico contesto. Una coppia libera²¹ è unicamente determinata da un suo rappresentante; quindi $X|Y$ è un rappresentante di α sse $X|Y \in \alpha$. La coppia libera identificata dal rappresentante $X|Y$ si denoterà $[X|Y]$; per cui per definizione $X|Y \in [X|Y]$ e $[X|Y] = [Z|V]$ equivale a $X|Y \approx Z|V$. Sull'insieme delle coppie libere si può stabilire una struttura di semigrupp abeliano, senza elemento neutro, definendovi una operazione (commutativa e associativa) di **somma** (+); e inoltre una struttura di insieme s-ordinato definendovi la relazione di **più piccolo di** (<). La definizione di + è la seguente: “ $c = [X|Y]$ è la somma di α e di β , $c = \alpha + \beta$, se esiste Z per cui $[X,Z,Y]$ (cioè per cui $d(X,Y) = d(X,Z) + d(Z,Y)$) e per cui $\alpha = [X|Z]$ e $\beta = [Z|Y]$ ”. (Questa definizione, si vede facilmente, non dipende dalla scelta del rappresentante di c). La definizione di < è la seguente: “ $\alpha = [X|Y]$ è più piccolo di $\beta = [Z|V]$, $\alpha < \beta$, se $d(X,Y) < d(Z,V)$ ²² (anche in questo caso, la definizione non dipende dalla scelta dei rappresentanti di α e di β). Quanto abbiamo sopra asserito circa la natura commutativa e associativa di +, e di quella di s-ordine di <, segue dalle definizioni senza difficoltà. Prendendo la disgiunzione esclusiva ($\underline{\vee}$) della relazione di s-ordine < con quella di uguaglianza (=), si ottiene l'associata relazione di t-ordine, che denoteremo \leq , sull'insieme delle coppie libere.

Arricchiremo ora la struttura di questo insieme introducendo l'operazione di “**sottrazione** (–) (di una coppia libera da una coppia libera)”: diremo che “ $c = \alpha - \beta$ ” (**differenza di α da β**) se “ $\alpha = c + \beta$ ” (si verifica facilmente che tale c esiste unico sse $\beta < \alpha$, per cui la definizione è consistente sotto questa condizione). Valgono allora le seguenti leggi, che coinvolgono una delle tre coppie estraibili da $\{+, -, <\}$, o anche l'intera terna $\{+, -, <\}$. «Per ogni α, β, c, d :

leggi con $\{+, -\}$: $(\alpha + \beta) - \beta = \alpha$;

leggi con $\{+, <\}$: 1) $(\alpha < \beta) \Rightarrow (\alpha + c < \beta + c)$; 2) $(\alpha < \beta \wedge c < d) \Rightarrow (\alpha + c < \beta + d)$;

²⁰ Invece la notazione $\{X,Y\}$ (per la coppia *non* ordinata di punti X,Y) è compatibile con $X = Y$. Secondo alcune formalizzazioni della teoria degli insiemi, il singoletto $\{x\}$ è *definito* come $\{x,x\}$ (vedi ad es. N. Bourbaki, “Théorie des Ensembles” 1, E II.4).

²¹ Per brevità, nel seguito, diremo ormai “coppia libera” per “coppia libera di punti distinti”.

²² “ $d(X,Y) < d(Z,V)$ ” equivale a “esiste U per cui $[Z,U,V]$ e $X|Y \approx Z|U$ (o $X|Y \approx U|V$)”.

leggi con $\{-, <\}$: 1) $(a < b \wedge c < a) \Rightarrow (a - c < b - c)$; 2) $(b < c \wedge a < b) \Rightarrow (c - b < c - a)$;

leggi con $\{+, -, <\}$: 1) $(b - a) \Rightarrow ((a - b) + b = a)$; 2) $(c < a \wedge c < b) \Rightarrow (a + (b - c)) = ((a - c) + b)$.»

La struttura sull'insieme delle coppie libere sarà ulteriormente arricchita introducendo l'operazione di **moltiplicazione** (\cdot) di una coppia libera per un naturale $n \geq 1$, mediante la regola ricorsiva: " $1 \cdot a = a$ e $(n+1) \cdot a = n \cdot a + a$ ". $n \cdot a$ si dirà **prodotto di n per a**.

Possiamo anche enunciare qui il seguente **teorema di congruenza tra rette**. «Siano p_1 e p_2 due rette, $\{X_1, Y_1\}$ e $\{X_2, Y_2\}$ coppie di punti distinti di p_1 e rispettivamente di p_2 per cui $X_1|Y_1 \approx X_2|Y_2$; e sia $Z_1 \in p_1$ distinto da X_1 e da Y_1 . Allora esiste al più un punto $Z_2 \in p_2$ per cui $Z_1|X_1 \approx Z_2|Y_2$ e $Z_1|Y_1 \approx Z_2|X_2$ ».

Concludiamo questa sottosezione con l'introduzione dell'assioma (il nono della nostra lista):

(A₆) "Per ogni semiretta p' con origine X , e per ogni coppia di punti distinti U, V esiste esattamente un punto $Y \in p'$ per cui $X|Y \approx U|V$ ".

La parte relativa alla unicità, in (A₆), è in realtà superflua. Infatti, se esistessero due siffatti punti Y e Y' , avremmo $X|Y \approx X|Y'$, da cui seguirebbe subito, disgiungendo i casi $[X, Y, Y']$ e $[X, Y, Y']$, $d(Y, Y') = 0$, ossia $Y = Y'$. L'avverbio "esattamente" è stato aggiunto soltanto per mettere in risalto l'analogia tra (A₆) e (A₁₁) della S.sez. 1.2.3. Inoltre i due punti distinti U, V dell'assioma (A₇) possono pensarsi come entrambi appartenenti alla retta includente p' (e non in generale a S), in modo da restare nell'ambito di una geometria "speciale" della retta, fino all'assioma (A₇) escluso. L'assioma (A₆) è il primo degli assiomi di congruenza di Hilbert, e può essere ignorato in tutto quanto segue fino alla presentazione di (A₁₁). (Dopo l'assioma (A₆), nel teorema di congruenza tra rette potremo tuttavia sostituire "esattamente un" ad "al più un".) La sopraddetta geometria speciale della retta si completerà in una sua "geometria diadica" utilizzando il "teorema della bisezione di un segmento" (v. S.sez. 1.2.3), e in una genuina "geometria della retta" aggiungendovi l'"assioma di continuità" (v. S.sez. 1.2.4).

1.2.3) I SUCCESSIVI OTTO ASSIOMI E LE GEOMETRIE DIADICHE DEL PIANO E DELLO SPAZIO

Ci siamo fin qui occupati di una retta e dei suoi sottoinsiemi, ma il prossimo assioma allarga questo orizzonte. Precisamente, esso afferma che

(A₇) "(In S) esistono (almeno) tre punti distinti non collineari".

Quindi esistono rette (distinte e) incidenti. Siano ora X_i ($i = 1, 2, 3$) tre tali punti distinti e non collineari, e sia $\mathcal{B}(X_1|X_2, X_3) \equiv \mathcal{B}(X_1|X_3, X_2)$ l'insieme (di punti di S) definito come l'unione, al

variare di V in $[X_2, X_3]$, delle rette $\mathcal{R}(X_1, V)$ (qui $[X_2, X_3]$ gioca dunque il ruolo di “insieme d’indici”). Sia poi $\mathcal{B}_i =: \mathcal{B}(X_i | X_{i+1}, X_{i+2})$ (con $i = 1, 2, 3$, leggi gli indici mod 3). L’unione di questi tre insiemi, diciamo $\mathcal{P}(X_1, X_2, X_3)$, è simmetrica nei suoi argomenti, e si dirà **piano per** X_1, X_2, X_3 . Un **piano** è un insieme di punti del tipo $\mathcal{P}(X, Y, Z)$ per qualche terna di punti distinti non collineari X, Y, Z .²³ È chiaro che esistono piani; li denoteremo con $\pi, \rho, \sigma, \dots \pi_1, \pi_2, \dots$ ecc. Anche l’intersezione dei tre insiemi \mathcal{B}_i , diciamo $\mathcal{T}(X_1, X_2, X_3)$, è simmetrica nei suoi tre argomenti, e si dirà **triangolo** (o **3-edro**) **chiuso di vertici** X_1, X_2, X_3 . L’insieme $\cup_i [X_i, X_{i+1}]$ ²⁴, anch’esso simmetrico nei suoi tre argomenti, si dirà **frontiera di** $\mathcal{T}(X_1, X_2, X_3)$; esso è manifestamente incluso in $\mathcal{T}(X_1, X_2, X_3)$. Sottraendolo (insiemisticamente) da quest’ultimo, si ottiene il **triangolo aperto di vertici** X_1, X_2, X_3 , che si denoterà $\mathcal{T}^\circ(X_1, X_2, X_3)$. Converremo che tutte le notazioni introdotte implicino tacitamente che X_1, X_2, X_3 siano distinti e non collineari.

Per ogni terna di punti distinti non collineari esiste un piano che li contiene, ma vale anche l’assioma (simile ad (A₄)):

(A₈) “Non esistono due piani distinti passanti entrambi per tre punti distinti non collineari dati” (ovvero, se si preferisce, “per tre punti distinti non collineari dati passa *al più* un piano”).

Quindi per tre punti distinti non collineari arbitrari X, Y, Z passa esattamente un piano, che è il piano $\mathcal{P}(X, Y, Z)$; e se X, Y, Z sono tre punti distinti non collineari del piano π , risulta $\pi = \mathcal{P}(X, Y, Z)$. Due rette incidenti p, q determinano esattamente un piano, che potrà denotarsi con $\mathcal{P}(p, q)$ ($\equiv \mathcal{P}(q, p)$). Quattro punti distinti X, Y, Z, U si diranno **coplanari** se esiste un piano cui essi appartengono; in questo caso si scriverà $\text{Cop}(X, Y, Z, U)$, convenendo al solito che questa relazione implichi automaticamente che $\neq(X, Y, Z, U)$. (La relazione $\text{Cop}(X, Y, Z, U)$ è manifestamente simmetrica nei suoi argomenti.) Esistono quaterne di punti coplanari distinti.²⁵ Due rette distinte p, q si diranno a loro volta **coplanari** se esiste un piano che le contiene; in questo caso si scriverà $\text{Cop}(p, q)$, convenendo che questa relazione implichi automaticamente $p \neq q$. (Anche la relazione $\text{Cop}(p, q)$ è manifestamente simmetrica.) Esistono coppie di rette distinte coplanari. Come due rette qualsiasi, due rette distinte coplanari sono disgiunte aut hanno esattamente un punto in comune (\equiv sono incidenti in quel punto). Sappiamo dunque che esistono rette incidenti, ma non ancora che esistono coppie di rette coplanari disgiunte. Il prossimo assioma è:

(A₉) “Se una retta p e un piano π hanno due punti distinti in comune, allora $p \subset \pi$ ”.

²³ Nelle parole di Euclide: “piano è quello che giace ugualmente rispetto alle sue rette” (altra definizione vuota).

²⁴ Si verifica subito che i tre segmenti semichiudi che si uniscono sono disgiunti, e quindi \cup potrebbe qui sostituirsi con \cup .

²⁵ Su quattro punti distinti si può fare la più forte richiesta che essi siano coplanari senza che nessuna delle quattro terne che da essi si estraggono sia una terna collineare. In questo caso i quattro punti si diranno distinti e “coplanari*”. Esistono quaterne di punti distinti e coplanari*.

Se la retta p è inclusa nel piano π , $\pi - p \neq \emptyset$. Data una retta p ed un punto X fuori di essa, è ovvio che esiste esattamente un piano π per cui $p \subset \pi$ e $X \in \pi$; denoteremo con $\mathcal{P}(p, X)$ questo piano.²⁶ « $(Y \in \mathcal{P}(p, X) \wedge Y \notin p) \Rightarrow (\mathcal{P}(p, Y) = \mathcal{P}(p, X))$ ». Due punti distinti X, Y e una retta p cui essi non appartengono si diranno ancora **coplanari** se esistono due punti distinti U, V di p per cui la quaterna $\{X, Y, U, V\}$ è coplanare. Scriveremo in questo caso $\text{Cop}(p|X, Y)$ ($\equiv \text{Cop}(p|Y, X)$), con la solita convenzione sulle nuove notazioni (in questo caso, che $\text{Cop}(p|X, Y)$ implichi che $X \neq Y$, $X \notin p$, $Y \notin p$). La relazione $\text{Cop}(p|X, Y) \wedge \exists (Z \in p) \{[X, Z, Y]\}$ si abbrevierà in $[X, p, Y]$ ($\equiv [Y, p, X]$) e si leggerà “ p è tra X e Y ”. Si può ora introdurre l’assioma:

(A₁₀) “Dati tre punti distinti e non collineari X, Y, Z e una retta p del loro piano non passante per Z , $[X, p, Y] \Rightarrow [Y, p, Z] \vee [X, p, Z]$ ”;

inoltre, poiché le due relazioni nella disgiunzione, si dimostra, sono incompatibili, la \vee del conseguente si può sostituire con $\underline{\vee}$.²⁷

Gli assiomi fin qui elencati sono sufficienti (non contando (A₆), che non serve allo scopo) per introdurre consistentemente due nuove importanti nozioni. La prima ricalca da vicino quella delle semirette complementari di una data origine (e su una data retta), ed è la nozione di **semipiano di bordo p di un piano π** (dove si presuppone $p \subset \pi$). Precisamente, fissata una retta p di un piano π , si dimostra che «la relazione $\neg[X, p, Y]$ tra generici punti X e Y di $\pi \setminus p$ (“ X e Y sono dalla stessa parte rispetto a p ”) è una equivalenza, rispetto alla quale $\pi \setminus p$ si decompone in esattamente *due* corrispondenti classi (disgiunte)». Ciascuna di queste due classi è completamente determinata, oltre che dal bordo p , da un suo rappresentante X , e potrà perciò essere denotata con $\mathcal{SP}(p, X)$; quindi, per ogni coppia di punti distinti X, Y appartenenti a $\pi \setminus p$, $\mathcal{SP}(p, X) = \mathcal{SP}(p, Y)$ sse $\neg[X, p, Y]$. $\mathcal{SP}(p, X)$ si dirà **semipiano di bordo p e contenente X** . Denoteremo semipiani con $\pi', \rho', \dots \pi'_1, \pi'_2, \dots$ ecc. È ovvio che, sotto $X \notin p$, X e p determinano unicamente anche il piano $\mathcal{P}(p, X)$ che include $\mathcal{SP}(p, X)$. L’insieme $\mathcal{SP}^*(p, X) =: (\mathcal{P}(p, X) \setminus \mathcal{SP}(p, X)) \setminus p$ si dirà **semipiano complementare di $\mathcal{SP}(p, X)$** . Si verifica subito che il semipiano complementare di $\mathcal{SP}^*(p, X)$ è lo stesso $\mathcal{SP}(p, X)$. Segue dalle definizioni che $\mathcal{SP}(p, X) \cap \mathcal{SP}^*(p, X) = \emptyset$, che $\mathcal{SP}(p, X) \underline{\cup} \mathcal{SP}^*(p, X) = \mathcal{P}(p, X) \setminus p$, e che « $Y \in \mathcal{SP}^*(p, X)$ equivale a $[X, p, Y]$ ». Per tre punti distinti non collineari X_1, X_2, X_3 di π , si considerino ora i tre semipiani $\mathcal{SP}(\mathcal{R}(X_{i+1}, X_{i+2}), X_i)$ ($i = 1, 2, 3$, leggi gli indici mod3). Si dimostra

²⁶ Qui e nel seguito, trascureremo ormai di avvertire che notazioni (per insiemi aut relazioni) introdotte sotto certe condizioni implicano tacitamente quelle condizioni. Così $\mathcal{P}(p, X)$ implica $X \notin p$.

²⁷ Una conseguenza di (A₁₀) è la seguente. Si consideri un triangolo e una retta del suo piano non passante per alcuno dei suoi vertici ma intersecante un suo lato: allora quella retta interseca esattamente uno dei due altri lati. Questo asserto è noto come “Assioma di Pasch”. In unione con gli altri assiomi, (A₁₀) e l’assioma di Pasch si equivalgono. Si noti che se nella formulazione di (A₁₀) mancasse l’ipotesi di non collinearità di X, Y e Z , e questi punti si supponessero collineari, allora l’implicazione $[X, p, Y] \Rightarrow [Y, p, Z] \underline{\vee} [X, p, Z]$ sarebbe conseguenza di assiomi precedenti. L’ipotesi di non collinearità di X, Y e Z serve ad evitare questo caso banale.

allora che «l'insieme intersezione di questi tre semipiani, manifestamente simmetrico nei suoi tre argomenti, è uguale al triangolo aperto $\mathcal{T}^\circ(X_1, X_2, X_3)$ ». ²⁸

La seconda nozione è quella di **fascio** (o **stella**) delle semirette (che in questo contesto diremo più convenientemente **raggi**), di un piano π e origine $X \in \pi$, diciamo $\mathcal{F}(\pi, X)$. L'unione di due raggi complementari di origine X di questo fascio e di $\{X\}$ è evidentemente una retta p di π ; e come sappiamo questa retta decompone $\pi \setminus p$ nell'unione di due semipiani complementari. La parte del fascio $\mathcal{F}(\pi, X)$ inclusa in uno di questi semipiani si dirà **semifascio** (di π) **di vertice** X del semipiano in oggetto. Un semifascio è quindi unicamente determinato dal bordo p del semipiano π' che lo include, dal suo vertice $X \in p$ e da un rappresentante $Y \notin p$ di π' , per cui si potrà denotare con $\mathcal{SF}(p|X|Y)$ (ove si sottintenderanno le due condizioni $X \in p$ e $Y \notin p$ della definizione). Su tre raggi *distinti* di un semifascio, diciamo s', q', r' , si può consistentemente definire la nozione di “essere tra”, come su tre generici punti distinti: precisamente, diremo che “ q' è **tra** s' e r' ”, scrivendo per questo $[s', q', r']$, ($\equiv [r', q', s']$) se per le loro rispettive intersezioni X', Y', Z' con una qualunque retta t che li interseca tutti e tre risulta $[X', Y', Z']$. Gli assiomi fin qui introdotti permettono di verificare che la definizione è consistente, cioè che «tale t esiste e che la relazione $[s', q', r']$ non dipende dalla sua scelta». Si può quindi esportare sui raggi di un semifascio la nozione del s -ordine (e associato t -ordine) indotto.

L'ordine su un semifascio è anche unicamente determinato assegnando una coppia di riferimento di suoi raggi distinti, diciamo p'_0, q'_0 , e un ordine su di essi; una coppia ordinata di raggi distinti del semifascio, diciamo $\langle p', q' \rangle$, sarà allora univocamente giudicata come equiversa o antiversa (alla coppia di riferimento) in modo ovvio, utilizzando una retta che interseca i quattro raggi, e le loro intersezioni su di essa. Per significare l'equiversità o l'antiversità di coppie ordinate di raggi distinti di un semifascio useremo la stessa notazione già vista per le coppie ordinate di punti distinti di una retta. Un semifascio può infine equivalentemente ordinarsi suddividendolo in due semi-semifasci complementari mediante un suo raggio, esattamente come si suddivide una retta in due semirette complementari mediante un suo punto. La relazione di inclusione tra semi-semifasci equiversi di un semifascio è una relazione di ordine totale, e corrisponde ad un ordine totale sui raggi-origine, come si voleva ottenere. ²⁹

²⁸ L'esistenza di questi triangoli aperti permetterebbe di definire una topologia del piano già a questo livello. Quando disporremo della “disuguaglianza triangolare” (vedi oltre in questa sezione), troveremo più naturale topologizzare il piano mediante d ; ma si potrebbe dimostrare che la topologia che usa i triangoli aperti come base di intorni, e quella che usa i dischi aperti (cioè la topologia fondata su d) sono equivalenti. (Il “disco aperto di centro O e raggio r ” è l'insieme dei punti X del piano per i quali $d(O, X) < r$).

²⁹ Una procedura analoga può applicarsi per dare un orientamento ad un intero fascio. Basterà prefissare in esso una coppia di riferimento p'_0, q'_0 di raggi distinti e non complementari, e un ordine su di essi. Per una coppia ordinata $\langle p', q' \rangle$ di raggi distinti e non complementari del fascio, se i quattro raggi p'_0, q'_0, p', q' appartengono ad un unico semipiano, si

L'unione dei raggi di un semifascio che stanno tra due suoi raggi distinti si dirà **angolo** tra quei raggi, che ne sono i **lati**.³⁰ Un angolo è dunque unicamente determinato dai suoi lati p'_1 e p'_2 , presupposti essere distinti (ciò che denoteremo con $p'_1|p'_2$) e appartenere ad un semifascio avente per vertice la loro comune origine (quindi non essere complementari), e potrà denotarsi con $\mathcal{A}(p'_1|p'_2)$ ($\equiv \mathcal{A}(p'_2|p'_1)$). Per introdurre consistentemente una nozione di congruenza tra angoli ci occorrono due nuovi assiomi, che sono:

(A₁₁) “Dato un semipiano π' di bordo p , una coppia $X|Y \subset p$ e una terna $\{V_1, V_2, V_3\}$ di punti distinti tra loro tali che $X|Y \approx V_1|V_2$, esiste esattamente un punto Z di π' per cui $(X|Z \approx V_1|V_3) \wedge (Y|Z \approx V_2|V_3)$.”³¹

(A₁₂) “Date due rette p_i ($i = 1, 2$), non necessariamente distinte, e punti distinti X_i, Y_i, Z_i di p_i , nonché V_i fuori di p_i , se $[X_i, Y_i, Z_i]$, e se $(X_1|Y_1 \approx X_2|Y_2) \wedge (Y_1|Z_1 \approx Y_2|Z_2) \wedge (V_1|X_1 \approx V_2|X_2) \wedge (V_1|Y_1 \approx V_2|Y_2)$, allora $V_1|Z_1 \approx V_2|Z_2$ ”;

Si considerino ora due angoli $\mathcal{A}(p'_1|p'_2)$ e $\mathcal{A}(q'_1|q'_2)$: detta X l'origine di p'_1 e p'_2 , e Y quella di q'_1 e q'_2 , diremo che il primo è **congruente** al secondo se esistono punti $Z_i \in p'_i$, $V_i \in q'_i$ ($i=1, 2$) tali che $(X|Z_1 \approx Y|V_1) \wedge (X|Z_2 \approx Y|V_2) \wedge (Z_1|Z_2 \approx V_1|V_2)$. In forza dei due ultimi assiomi si può dimostrare che la definizione è consistente, cioè «non dipende dalla scelta dei punti Z_i e V_i sotto le condizioni stipulate, e nemmeno dalle origini delle due coppie di lati, ma soltanto da queste coppie di lati di comune origine». Quindi se gli angoli $\mathcal{A}(p'_1|p'_2)$ e $\mathcal{A}(q'_1|q'_2)$ sono congruenti, potremo scrivere semplicemente $p'_1|p'_2 \approx q'_1|q'_2$ (dando per inteso che la notazione $p'|q'$ sia simmetrica rispetto ai suoi due argomenti e che p'_1 e p'_2 , nonché q'_1 e q'_2 , oltre che distinti non siano complementari). Si dimostra poi che «la relazione di congruenza (di un angolo ad un angolo) è una equivalenza». Corrispondentemente, l'insieme di tutti gli angoli può decomporsi in classi (disgiunte) di angoli congruenti. Queste classi si diranno **angoli liberi**, e si denoteranno $\mathfrak{A}, \mathfrak{B}, \dots, \mathfrak{A}_1, \mathfrak{A}_2, \dots$ ecc. Un angolo libero è unicamente determinato da un suo rappresentante, del quale interessano solo i lati, diciamo p' e q' , e che potremo ancora denotare, in questa occorrenza, con $p'|q'$. Quindi tale $p'|q'$ è un rappresentante di \mathfrak{A} sse $p'|q' \in \mathfrak{A}$. L'angolo libero identificato dal rappresentante $p'|q'$ si denoterà

procederà come nel caso precedente per decidere se le due coppie sono equiverse o antiverse; in caso contrario, i quattro raggi si ridurranno in un semipiano prendendo il complementare di uno di essi e *rovesciando* il giudizio di equiversità o antiversità. Cioè, se così facendo si trovasse ad es. $\langle p', q' \rangle \uparrow \downarrow \langle p'_0, q'_0 \rangle$, si concluderebbe che $\langle p', q' \rangle \uparrow \uparrow \langle p'_0, q'_0 \rangle$. Intuitivamente, con la scelta della coppia di riferimento si è fissato un “verso di circolazione” sull'intero fascio, che diventa così un **fascio orientato**.

³⁰ Come nel caso dei segmenti (aperti, chiusi, semichiusi), un angolo può non comprendere i suoi lati (angolo “aperto”) o comprenderne uno, ecc. A prima vista sembrerebbe esserci completa equivalenza formale tra angoli di dati lati e segmenti di dati estremi; ma come vedremo, vi è una limitazione a carico dell'operazione di “somma” di due angoli “liberi” (vedi oltre) che non ha corrispondenza nell'analogia operazione di “somma” di due segmenti liberi.

³¹ Questo assioma è l'analogo di (A₆) per i piani.

con $[p'|q']$; cioè per definizione $p'|q' \in [p'|q']$, e $[p'|q'] = [r'|s']$ equivale a $p'|q' \approx r'|s'$. Si noti la fin qui totale analogia con il caso dei segmenti liberi.³²

Sia ora $\text{Col}(X,Z,Y)$; se $X|Z \approx Z|Y$, Z si dice *il punto di mezzo tra* X e Y . « $\text{Col}(X,Z,Y) \wedge (X|Z \approx Z|Y) \Rightarrow [X,Z,Y]$ ». La definizione di Z è consistente, perché con gli assiomi disponibili si dimostra che «esiste esattamente un punto di mezzo di ogni coppia $X|Y$ ». ³³ Ciò equivale ad affermare che per ogni coppia libera α esiste unica la coppia libera α per la quale $\alpha = 2 \cdot \alpha$. Per definizione, porremo $\alpha =: 2^{-1} \cdot \alpha$. Quindi per ogni naturale $k \geq 0$ esiste unico $2^{-k} \cdot \alpha$; e infine esiste unico, per ogni $k \geq 0$ e ogni $n \geq 1$, $n \cdot (2^{-k} \cdot \alpha) \equiv (2^{-k}n) \cdot \alpha$. In conclusione, oltre alle precedenti strutture di semigruppato commutativo senza elemento neutro e con sottrazione (sotto la condizione “sottraendo < diminuendo”), e di insieme s -ordinato, l'insieme delle coppie libere (di punti distinti) è un'algebra sui razionali del tipo $2^{-k}n$, per ogni $n \geq 1$ e ogni $k \geq 0$, cioè sui cosiddetti **numeri diadici positivi**.

Diamo adesso alcuni teoremi derivabili dagli assiomi di cui fin qui disponiamo e che permettono di introdurre consistentemente nuove fondamentali nozioni. Sia π' un semipiano di bordo p , X un punto di p , e p' , p'^* i raggi complementari di origine X per cui $p = p' \cup p'^* \cup \{X\}$. Allora «esiste un unico raggio r' del semipiano π' tale che $p'|r' \approx p'^*|r'$ ». Due angoli di comune vertice X , che come $p'|r'$ e $p'^*|r'$ hanno un lato comune e per gli altri lati hanno raggi complementari (di origine X), si diranno **angoli supplementari**. Manifestamente l'essere supplementari, per due angoli, è una relazione simmetrica, e ogni angolo ha esattamente due supplementari, uno per lato. I due angoli supplementari di uno stesso angolo si diranno **angoli opposti**, e «risultano congruenti». «Gli angoli supplementari, oppure opposti, di angoli congruenti sono congruenti». Due angoli liberi si diranno **supplementari** se un rappresentante dell'uno è supplementare di un rappresentante dell'altro (la definizione è consistente). Un angolo congruente con il *suo* supplementare si dirà **angolo retto**; alla luce del primo teorema di questo paragrafo esistono angoli retti, e sono congruenti fra loro. La classe di equivalenza degli angoli retti si dirà **angolo libero retto** e si

³² Si ha qui una sovrapposizione concettuale tra “angolo” e “coppia di raggi distinti” (di comune vertice e non complementari) del tutto analoga a quella che abbiamo incontrato tra “segmento” e “coppia di punti distinti”. Come già in quel caso, i due concetti, sebbene diversi, sono liberamente interscambiabili nell'ambito dell'algebra degli angoli liberi (vedi appresso).

³³ Per dimostrare l'esistenza (non l'unicità) del punto di mezzo di una data coppia $X|Y$ occorre immergere $X|Y$ in un piano (un'idea che risale ad Euclide), e ciò può a prima vista disturbare. D'altra parte la sufficienza allo scopo di tale immersione può anche considerarsi in certo senso fortuita. Volendo evitare questi aspetti blandamente imbarazzanti, sul piano intuitivo, della teoria, l'unica via di uscita sembra quella di ricorrere ad un assioma di esistenza ad hoc. Questo assioma, che si potrebbe dire a ragione **assioma di divisibilità** (beninteso, non necessariamente per 2) è stato più in generale ravvisato come necessario all'istituzione di una teoria formalizzata della misura di grandezze generiche. Esso si trova enunciato nella forma «ogni quantità può essere suddivisa in n parti uguali, dove n è un naturale qualunque ≥ 1 » nella “Allgemeine Funktionentheorie” (1882) di Du Bois-Reymond (Paul, 1831-1889), e ivi denominato “assioma di linearità”. Nella presente formalizzazione metrica della geometria continueremo comunque a far uso del sopravvisto

denoterà con \mathfrak{R} . Con qualche importante diversità, risultati, definizioni e notazioni simili a quelli stabiliti per l'insieme dei segmenti liberi si ottengono/introducono con riferimento all'insieme degli angoli liberi. Precisamente, diremo che " $\mathfrak{C} = [p'|q']$ è la **somma degli angoli liberi \mathfrak{A} e \mathfrak{B}** , $\mathfrak{C} = \mathfrak{A} + \mathfrak{B}$, se esiste un raggio r' (dello stesso semifascio) per cui $[p',r',q']$ e $\mathfrak{A} = [p'|r']$ e $\mathfrak{B} = [q'|r']$ "; e diremo che " $\mathfrak{A} = [p'|q']$ è **più piccolo di $\mathfrak{B} = [r'|s']$** , $\mathfrak{A} < \mathfrak{B}$, se esiste un raggio t' per cui $[r',t',s']$ e $p'|q' \approx r'|s'$ ". Queste due definizioni sono identiche alle loro corrispondenti nel caso delle coppie libere. Compare tuttavia una condizione per l'esistenza della somma, che è « $\mathfrak{A} + \mathfrak{B}$ esiste sse $\mathfrak{B} < \mathfrak{A}^*$ (\equiv angolo libero supplementare di \mathfrak{A})». La definizione di sottrazione ($-$) condizionata è identica a quella già vista per le coppie libere: diremo che " $\mathfrak{C} = \mathfrak{A} - \mathfrak{B}$ " (**differenza di \mathfrak{A} da \mathfrak{B}**) se esiste \mathfrak{C} per cui $\mathfrak{A} + \mathfrak{C} = \mathfrak{B}$ " (tale \mathfrak{C} esiste unico sse $\mathfrak{B} < \mathfrak{A}$). Il carattere di semigruppato abeliano dell'insieme degli angoli liberi segue analogamente al caso dei segmenti liberi (o coppie libere di punti distinti), ma con certe precauzioni di tipo "esistenziale". Vale a dire, per quanto riguarda la commutatività della somma, abbiamo il teorema «se $\mathfrak{A} + \mathfrak{B}$ esiste, allora $\mathfrak{B} + \mathfrak{A}$ esiste, e $\mathfrak{A} + \mathfrak{B} = \mathfrak{B} + \mathfrak{A}$ »; e per quanto riguarda la sua associatività, «se $\mathfrak{A} + \mathfrak{B}$ e $(\mathfrak{A} + \mathfrak{B}) + \mathfrak{C}$ esistono, allora $\mathfrak{B} + \mathfrak{C}$ e $\mathfrak{A} + (\mathfrak{B} + \mathfrak{C})$ esistono, e $(\mathfrak{A} + \mathfrak{B}) + \mathfrak{C} = \mathfrak{A} + (\mathfrak{B} + \mathfrak{C})$ ». Si hanno anche leggi che coinvolgono la coppia $(+, <)$, cioè

- 1) «se $\mathfrak{A} < \mathfrak{B}$, e $\mathfrak{B} + \mathfrak{C}$ esiste, allora $\mathfrak{A} + \mathfrak{C}$ esiste, e $\mathfrak{A} + \mathfrak{C} < \mathfrak{B} + \mathfrak{C}$ »; e
- 2) «se $\mathfrak{A} < \mathfrak{B}$ e $\mathfrak{C} < \mathfrak{D}$, e $\mathfrak{B} + \mathfrak{D}$ esiste, allora $\mathfrak{A} + \mathfrak{C}$ esiste, e $\mathfrak{A} + \mathfrak{C} < \mathfrak{B} + \mathfrak{D}$ »; e una legge che coinvolge la terna $\{+, -, <\}$, cioè
- 3) «se $\mathfrak{B} < \mathfrak{A}$, la somma $(\mathfrak{A} - \mathfrak{B}) + \mathfrak{B}$ esiste, e $(\mathfrak{A} - \mathfrak{B}) + \mathfrak{B} = \mathfrak{A}$ ».

Vale infine il teorema: «se $[p',q',r']$, allora $[p'|q'] + [r'|p'^*] = [r'|q'^*]$ »³⁴

Una retta p si dirà **perpendicolare** ad una retta q se p e q sono incidenti, diciamo in un punto X , ed esistono raggi p' e q' di origine X , inclusi in p e rispettivamente in q , tali che l'angolo $p'|q'$ sia retto. Ovviamente l'essere perpendicolari, per due rette incidenti p e q , è una relazione simmetrica, che si denoterà $p \perp q$ ($\equiv q \perp p$). «Data una retta p e un punto X fuori di essa, nel piano $\mathcal{P}(p,X)$ esiste unica la perpendicolare q alla p passante per X .» L'intersezione (unica) X_p di questa q con la p si dirà la **proiezione ortogonale di X sulla p** . «Per una data retta p ed un dato $X \notin p$, risulta $[X|U] > [U|X_p]$ (e quindi anche $[X|U] > [X|X_p]$) per ogni $U \in p \setminus X_p$ ». Ciò si esprime comunemente dicendo che in un triangolo rettangolo "ciascuno dei due cateti è più piccolo dell'ipotenusa", e porta direttamente alla ben nota "disuguaglianza triangolare" (vedi più sotto). «Data una retta $p \subset \pi$ e un

teorema della bisezione di un segmento immerso in un piano. La necessità di immergere $X|Y$ in un piano per definire il loro punto di mezzo lede quella che Hilbert chiamava "purezza di metodi".

³⁴ Non avrebbe senso invece dire che "la somma di un angolo libero \mathfrak{A} e del suo supplementare \mathfrak{A}^* è indipendente da \mathfrak{A} ", come l'intuizione fortemente suggerirebbe, semplicemente perché quella somma non esiste.

punto $X \in p$, esiste unica la perpendicolare $q \subset \pi$ a p e passante per X ». Questa perpendicolare si potrà denotare con $\perp(X \in p \subset \pi)$. «Date due rette incidenti p e q *non perpendicolari*, e dati tre punti distinti X, Y, Z su q , le proiezioni ortogonali su p di questi punti sono tali che $[X, Z, Y] \Leftrightarrow [X_p, Z_p, Y_p]$ ».³⁵

Altri importanti teoremi, derivabili dagli stessi assiomi, sono i seguenti. Se per tre raggi distinti p', q', r' di un semifascio avviene che $p'|q' \approx q'|r'$, r' si dice *il bisettore* dell'angolo $p'|q'$. Infatti, «esiste esattamente un bisettore di un dato angolo, e se r' è bisettore di $p'|q'$, allora $[p', r', q']$ ». (Questo teorema è analogo a quello del punto di mezzo di una data coppia di punti, e rende conto della consistenza della definizione.). Quindi «per ogni angolo libero \mathfrak{A} esiste esattamente un angolo libero \mathfrak{X} per cui $\mathfrak{A} = 2 \cdot \mathfrak{X}$ ». *Con le dovute precauzioni esistenziali*, questo fatto consente ancora di ottenere un'algebra sull'insieme degli angoli liberi e su quello dei numeri diadici positivi. Non insistiamo su questi sviluppi, che il lettore potrà ottenere da sé senza difficoltà significative.

Risulta ancora che «per arbitrari tre punti distinti X, Y, Z , $d(X, Y) \leq d(X, Z) + d(Z, Y)$ ». Questa è appunto la preannunciata **disuguaglianza triangolare**, ed insieme agli assiomi (A_1, A_2) , permette di pensare alla distanza (o a qualunque applicazione dello stesso tipo con valori proporzionali a quelli di d secondo un fattore strettamente positivo) come “metrica” del piano dei tre punti X, Y, Z , e quindi di topologizzare il piano con la topologia indotta da questa metrica (vedi App. Gen. C)³⁶. Se poi $p = \mathcal{R}(X, Y)$, e $Z \notin p$ è tale che $[X, Z_p, Y]$, il punto Z_p può pensarsi come quello che rende minima la somma $d(X, U) + d(U, Y)$ al variare di U sulla perpendicolare q a p per Z , e per i dati X, Y fissi.^{37, 38}

«Due rette coplanari, distinte ed entrambe perpendicolari a una terza retta, sono disgiunte (ovviamente esistono coppie di tali rette)» Quindi possiamo ormai affermare che in un dato piano π esistono coppie di rette coplanari distinte e disgiunte; e quindi, che data una retta p ed un punto $X \notin p$, esiste *almeno* una retta coplanare a , e disgiunta da p , passante per X . Due rette p, q coplanari e disgiunte, oppure identiche, si dicono **rette parallele**; questo si denota con $p \parallel q$. La

³⁵ Intuitivamente si sarebbe tentati di supporre valida una relazione del tipo $[X|Z]/[X_p|Z_p] = [Y|Z]/[Y_p|Z_p]$. Questa è una tipica “insidia euclidea”: infatti tale ipotesi non sarebbe qui giustificabile (ma lo sarà in un piano euclideo).

³⁶ La disuguaglianza triangolare è qui da pensare con riferimento ad un piano, perché ancora non sappiamo se esistono piani distinti; ma è ovvio che se esistono piani distinti essa vale in ognuno di essi. Inoltre la riserva che i tre punti X, Y, Z siano distinti è da intendere nel senso che in caso contrario la disuguaglianza è soddisfatta banalmente. Si noti che non è necessario, per la sua validità, che il piano in questione sia *euclideo* (vedi S.sez. 1.2.4).

³⁷ A proposito della disuguaglianza triangolare, un'altra “insidia euclidea” è la seguente: si sarebbe tentati di supporre che, se V è un punto di q distinto da Z e dalla sua stessa parte rispetto a Z_p , si abbia $d(V, X) <$ (rispettivamente $>$) $d(Z, X)$ per ogni $X \in p$ distinto da Z_p , se $d(V, Z_p) <$ (risp. $>$) $d(Z, Z_p)$. Questo sarebbe vero in un piano euclideo, ma non può essere ancora giustificato qui.

³⁸ È chiaro che l'aver definito l'allineamento di tre punti in termini di distanza ha l'inconveniente pratico della cattiva sensibilità della sua verifica *empirica*, nello spazio fisico reale. Infatti l'allineamento è di fatto verificabile in termini di un problema di minimo, la cui soluzione ha sensibilità nulla rispetto a variazioni del primo ordine. Questo spiega

relazione \parallel tra rette è manifestamente simmetrica e riflessiva; ma non possiamo ancora affermare che \parallel sia anche transitiva, e quindi che sia un'equivalenza. Riformuleremo il precedente teorema affermando che «esiste almeno una parallela ad una retta data passante per un punto dato» (se il punto è sulla retta, allora la parallela è la retta stessa).

Diamo ancora il teorema della “congruenza tra piani” che è l'ovvia generalizzazione di quello della congruenza tra rette. «Siano π_1 e π_2 due piani, $\{X_1, Y_1, Z_1\}$ e $\{X_2, Y_2, Z_2\}$ terne di punti distinti non collineari di π_1 , e rispettivamente di π_2 , per cui $X_1|Y_1 \approx X_2|Y_2$, $Y_1|Z_1 \approx Y_2|Z_2$ e $Z_1|X_1 \approx Z_2|X_2$; e sia $V_1 \in \pi_1$ e distinto da X_1, Y_1, Z_1 . Allora esiste esattamente un $V_2 \in \pi_2$ per cui $V_1|X_1 \approx V_2|X_2$, $V_1|Y_1 \approx V_2|Y_2$ e $V_1|Z_1 \approx V_2|Z_2$ ».

Gli assiomi fin qui introdotti sono necessari e sufficienti ad istituire una **geometria diadica del piano**, che naturalmente include quella della retta fornendole il teorema di bisezione. Per passare alla cosiddetta “geometria assoluta del piano” occorrerebbe introdurre un assioma di continuità (vedi S.sez. 1.2.4). Prima di procedere in tal senso, ci interessa tuttavia estendere la geometria diadica del piano ad una analoga **geometria diadica dello spazio** (che la include). L'assioma che apre questa prospettiva è, come ben naturale,

(A₁₃) “Esistono (almeno) quattro punti distinti non coplanari”.

In particolare, questo assioma ci assicura l'esistenza di rette non incluse in un piano dato, e quindi quella di piani distinti. A questo punto si ripropongono sviluppi completamente analoghi a quelli che, partendo da una terna di punti distinti non collineari, hanno condotto alla definizione del piano per quei tre punti e del triangolo chiuso (o aperto) avente quei punti per vertici; ciò che così si ottiene è uno **spazio** (euclideo) **3-dimensionale**”, diciamo S_3 , che denoteremo $\mathcal{P}(X_1, X_2, X_3, X_4)$ (simmetrico nei suoi quattro argomenti) e rispettivamente il **tetraedro** (o **4-edro**) **chiuso** [**aperto**] di tali vertici, che denoteremo $\mathcal{T}(X_1, X_2, X_3, X_4)$ [$\mathcal{T}^\circ(X_1, X_2, X_3, X_4)$], anch'esso simmetrico nei suoi argomenti. Si potrebbe continuare nello stesso modo mediante un assioma che assicura l'esistenza di una quintupla di punti distinti non appartenenti a uno spazio 3-dimensionale, generando uno spazio 4-dimensionale e un 5-edro chiuso (o aperto), ... e così via fino ad un n per il quale sussiste un assioma che *nega* l'esistenza di un punto *fuori* dello spazio n -dimensionale. Per i fini presenti, noi potremo tuttavia fermarci a $n = 3$, cioè con un assioma che nega l'esistenza di un punto fuori di $S \equiv S_3$. A questo assioma può darsi la forma seguente:

(A₁₄) “Se due piani distinti hanno un punto comune, allora ne hanno anche un altro distinto dal primo”.

perché, in pratica, nella geometria fisica si verifica la collinearità di tre o più punti servendosi di un “campione di collinearità” (o “riga” ideale).

(A₁₄) è uno degli assiomi di incidenza (il settimo) di Hilbert, e oltre ad affermare che due piani distinti sono disgiunti (se ne esistono) aut hanno in comune un'intera retta (cfr. l'assioma (A₉)), implica infatti (si dimostra) la tridimensionalità dello spazio $S \equiv S_3$.³⁹

In questo caso, così come una retta p di un piano π ed un punto $X \in \pi \setminus p$ identificano i due semipiani complementari $\mathcal{SP}(p,X)$ e $\mathcal{SP}^*(p,X)$ di π , analogamente un piano π ed un punto $X \in S \setminus \pi$ identificano due **semispazi complementari** di S , che denoteremo $\mathcal{SP}(\pi,X)$ e $\mathcal{SP}^*(\pi,X)$. Inoltre, se X e Y sono punti distinti di $S \setminus \pi$, la relazione $\exists(Z \in \pi)\{[X,Z,Y]\}$ si leggerà “ π sta tra X e Y ” e si denoterà $[X,\pi,Y]$ ($\equiv [Y,\pi,X]$)⁴⁰; mentre la relazione $\neg[X,\pi,Y]$ (“ X e Y sono dalla stessa parte rispetto a π ”) risulta essere una equivalenza tra i punti X e Y di $S - \pi$, rispetto alla quale $S - \pi$ si decompone in esattamente *due* classi (disgiunte) identificate ciascuna dal comune bordo π e da un loro rappresentante. Queste classi sono appunto i due semispazi complementari $\mathcal{SP}(\pi,X)$ e $\mathcal{SP}^*(\pi,X)$, se X è un rappresentante del primo di essi. Un semispazio di S è un insieme del tipo $\mathcal{SP}(\pi,X)$ (o equivalentemente, del tipo $\mathcal{SP}^*(\pi,X)$) per qualche piano π e qualche punto $X \in S \setminus \pi$. Denoteremo generici semispazi con S', R', \dots ecc.

Una retta p e un piano π si diranno **incidenti** (tra loro) se hanno in comune esattamente un punto (esistono coppie di rette e piani incidenti), cioè se $\exists!(Z)\{p \cap \pi = \{Z\}\}$. Essi si diranno poi **perpendicolari** (tra loro) se, essendo incidenti, p è perpendicolare a due rette distinte di π passanti entrambe per $p \cap \pi$. La definizione è consistente perché, si dimostra, «se p è perpendicolare a π , allora è perpendicolare a *tutte* le rette di π passanti per $p \cap \pi$.» Significheremo che la retta p è perpendicolare al piano π con la notazione $p \perp \pi$, per definizione equivalente a $\pi \perp p$. «Esiste esattamente un piano [una retta] passante per un dato punto X e perpendicolare a una data retta [ad un dato piano].» Un piano π si dirà **perpendicolare** ad un piano ρ se esiste una retta $p \subset \pi$ perpendicolare a ρ . La relazione “ π è perpendicolare a ρ ” è simmetrica, e sarà denotata $\pi \perp \rho$, o equivalentemente $\rho \perp \pi$. «Due rette distinte ed entrambe perpendicolari ad uno stesso piano sono coplanari e disgiunte (cioè parallele)»⁴¹. Sia X_π l'intersezione con il piano π dell'*unica* retta passante per X e perpendicolare a π ; essa si dirà **proiezione ortogonale di X su π** . Similmente all'analogo teorema della geometria del piano, «se p e π non sono perpendicolari, e X, Y, Z sono punti di p , allora $[X,Z,Y] \Leftrightarrow [X_\pi, Z_\pi, Y_\pi]$ ».

³⁹ Si dimostra che in uno spazio quadridimensionale (nel senso sopra accennato) due piani non disgiunti possono avere *un solo* punto comune.

⁴⁰ $[X,\pi,Y]$ non è tuttavia un tipo essenzialmente nuovo di relazione; nel senso che se ρ è un *qualunque* piano contenente i punti distinti X e Y rispetto ai quali π è “tra” (quindi necessariamente incidente a π), allora $[X,\pi,Y] \Leftrightarrow [X,\pi \cap \rho, Y]$; vale a dire, la relazione $[X,\pi,Y]$ si riconduce a quella dell’“essere tra” di una retta p rispetto a due punti X, Y per i quali $\text{Cop}(p|X,Y)$.

⁴¹ E' questa una dimostrazione alternativa, e indipendente, dell'esistenza di parallele.

Due piani π e ρ entrambi perpendicolari ad una data retta p sono disgiunti aut identici; essi si diranno **paralleli** tra loro (la relazione di parallelismo è riflessiva e simmetrica, ma per il momento non necessariamente transitiva). Significheremo che π e ρ sono paralleli con la notazione $\pi \parallel \rho$ ($\equiv \rho \parallel \pi$). Esistono piani paralleli, ed «esiste *almeno un* piano parallelo ad un piano dato passante per un punto dato (se il punto è sul piano, allora il piano parallelo è il piano originale stesso)». Si noti la completa analogia dei contenuti del presente paragrafo con il precedente caso di due rette parallele in quanto perpendicolari ad una stessa retta data e coplanari, oppure in quanto identiche.

1.2.4) GLI ULTIMI DUE ASSIOMI: LA GEOMETRIA ASSOLUTA E LA GEOMETRIA EUCLIDEA

Per completare la formalizzazione metrica della geometria euclidea occorrono ormai soltanto l'**assioma di continuità**, diciamo (C), e l'**assioma delle parallele** (o di Euclide), diciamo (E). L'uno e l'altro si possono dimostrare essere indipendenti dagli assiomi precedenti e tra loro ⁴². Li illustriamo qui appresso, per il momento senza priorità.

L'assioma (C) traduce in modo preciso l'idea intuitiva di "continuità geometrica" di un prefissato insieme I di punti. Ciò può ricevere varie formulazioni equivalenti, sia nella forma di un enunciato unico che in quella di congiunzione di più enunciati parziali. Diamo ora alcune di queste definizioni, le prime due del primo tipo (enunciato unico) e le altre del secondo tipo (enunciati parziali).

§ Dedekind (Richard, Braunschweig Ger. 1831-1916, "Ded"):

"L'insieme di punti I , supposto s -ordinato ($<$), ha la D -proprietà". Cioè (v. Sez. 1.3), "per ogni separazione $\{I_1, I_2\}$ di $(I, <)$, I_1 ha un ultimo, aut I_2 ha un primo elemento"; ovvero anche "per ogni tale separazione, esiste un $Z \in I$ tale che $\forall (X_1 \in I_1, X_2 \in I_2) \{X_1 \leq Z < X_2 \vee X_1 < Z \leq X_2\}$ ".

Weierstrass (Karl, Ostenfelde Ger. 1815, Berlino 1897, "W"):

"Ogni sequenza (infinita) di punti di $(I, <)$ che sia non decrescente e limitata verso l'alto (rispetto alla relazione $<$) converge". Vale a dire, se $X_1 \geq X_2 \geq \dots$ è la sequenza in oggetto, ed esiste $Z \in I$ tale che $X_i \leq Z$ per ogni $i \geq 1$, allora esiste un $L \in I$ per cui, fissato $\varepsilon > 0$ ad arbitrio, esiste $N_\varepsilon \geq 1$ tale che $d(X_i, L) < \varepsilon$ per ogni $i \geq N_\varepsilon$. §

§ Cantor (Georg, S. Pietroburgo 1845, Halle 1918, "Cant"):

Sia $\langle [X_i, Y_i] \rangle_{i=1, \dots}$ una sequenza infinita di segmenti chiusi tale che

1) $[X_{i+1}, Y_{i+1}] \subset [X_i, Y_i]$ per ogni $i \geq 1$;

2) data arbitrariamente una coppia $U|V$, esiste un $n \geq 1$ per cui $[X_n|Y_n] < [U|V]$.

Allora (Cant) è l'asserto "per ogni sequenza del tipo sopraddetto e di estremi in I , $\bigcap_{i=1}^{\infty} [X_i, Y_i] = \{Z\}$ per qualche punto $Z \in I$." ⁴³ §

§ Cauchy (Augustin-Louis, Parigi 1789, Sceaux Fr. 1857, "Cau"):

"Ogni sequenza "fondamentale" $X_1, X_2 \dots$ di punti di I – cioè per la quale, per $\varepsilon > 0$ fissato ad arbitrio, $d(X_{n+p}, X_n) < \varepsilon$ per ogni $n \geq N_\varepsilon$ e ogni $p \geq 0$ – converge ad un punto di I ". §

§ Archimede (287 – 212 a. C., "Arch"):

"Comunque siano scelte le coppie $X|Y$ e $U|V$ di I esiste un naturale n tale che $n \cdot [X|Y] > [U|V]$ ".

Poiché 2^k può sempre sostituirsi ad n nella precedente disuguaglianza per un $k \geq 0$ abbastanza grande, essa può anche enunciarsi dicendo che esiste un $k \geq 0$ per cui $[U|V] < 2^{-k} \cdot [X|Y]$. (Se si ammettesse l'assioma di divisibilità (Du Bois-Reymond, vedi S.sez. 1.2.3) la prima disuguaglianza si potrebbe trascrivere semplicemente come $[U|V] < n^{-1} \cdot [X|Y]$.) §

Come si diceva, (Ded, W) sono definizioni del primo tipo, cioè esauriscono ciascuna per suo conto la definizione di continuità di I , mentre (Cant, Cau, Arch) non bastano da sole allo scopo. Risulta infatti: (Ded) \Leftrightarrow (W); (Cant) \Leftrightarrow (Cau); (Ded) \Leftrightarrow (Cant) \wedge (Arch). In conclusione, se si presuppone (Arch), le rimanenti (Ded), (W), (Cant) e (Cau) si equivalgono. Le definizioni di continuità di I diventano l'assioma (C) quando le si riferisca ad uno specifico tipo di insieme di punti I . In pratica è sufficiente che I sia un segmento aperto, o una retta, che hanno dunque la "potenza del continuo" (cioè, dell'insieme delle parti di \mathbf{N} , o $\mathcal{P}(\mathbf{N}) \equiv \aleph_1$). Quindi possiamo enunciare l'assioma di continuità nella forma (ad esempio):

(C) "Un segmento aperto gode della D-proprietà". ⁴⁴

Intuitivamente, (Arch) significa che riportando per un numero sufficiente di volte un segmento di lunghezza "comunque piccola" su una semiretta a partire dalla sua origine (sempre dalla stessa parte), si finisce col superare un suo punto assegnato ad arbitrio. Ancora intuitivamente, (Cant) significa che i punti di una retta sono "densi" in essa, nel senso che ce n'è uno (lo abbiamo chiamato Z) che sta tra gli elementi di ogni coppia della sequenza di coppie di punti "incastonate" una dentro la precedente $X_1|Y_1, \dots, X_i|Y_i, \dots$, comunque si scelga la sequenza. Si potrebbe anzi

⁴² Ciò vale se l'aritmetica dei numeri reali è coerente. Sotto questa condizione, è infatti possibile costruire modelli semantici che soddisfano agli assiomi precedenti ma non a (C); oppure anche agli assiomi precedenti più (E) (o più la negazione di (E), vedi oltre) ma non a (C).

⁴³ Generalizzando l'asserto di Cantor ad un generico spazio metrico M con distanza ρ , esso diventa: "per ogni sequenza di sottoinsiemi non vuoti e chiusi di M , diciamo $\langle A_i \rangle_{i=1, \dots}$, per cui $A_{i+1} \subset A_i \forall i \geq 1$ (sequenza "decreciente") e per cui $\lim_{n \rightarrow \infty} \text{diam}(A_n) = 0$ – ove $\text{diam}(A) =: \sup\{\rho(x, y) | x, y \in A\}$ (secondo la definizione standard) –, allora $\bigcap_{i=1}^{\infty} A_i = \{x\}$ per qualche $x \in M$ ". Se questo è il caso, M è **completo** secondo la definizione standard, cioè ogni sua sequenza di Cauchy converge in M . Quindi (Cant) asserisce la completezza del segmento chiuso $[X_1, Y_1]$ per ogni coppia $X_1|Y_1$ di S .

facilmente vedere che quel punto Z , di cui (Cant) assicura l'esistenza, è unico. Esistono ulteriori alternative equivalenti (in generale, nel senso che le teorie con il sistema di assiomi fin qui introdotti, più gli "assiomi di continuità" prescelti, sono equivalenti); ad esempio, alla coppia (Arch) e (Cant) potrebbe sostituirsi (Arch) unito a "per ogni $X|Y$ esiste una biiezione tra il segmento aperto (X,Y) e l'intervallo reale aperto $(0,1)$ ". In una delle parecchie versioni della sua formalizzazione, Hilbert ha usato la congiunzione $(Arch) \wedge (Cant)$ come assioma di continuità.⁴⁵

Anche l'assioma delle parallele può esprimersi in varie forme, di cui diamo qui la più semplice ed espressiva.

(E) "Per qualunque punto dato, passa *al più* una parallela ad una qualunque retta data"⁴⁶

Combinando (E) con il teorema che afferma l'esistenza di *almeno* una parallela nella stessa situazione (v. S.sez. 1.2.3), l'"al più" di (E) può sostituirsi con "esattamente".

Se si rinuncia a (C) e si accoglie (E), *non* si ottiene una geometria dello spazio isomorfa a \mathbf{R}^3 (o del piano isomorfa a \mathbf{R}^2 , ecc.). Infatti questa geometria coinvolge, in luogo di \mathbf{R} , un suo sottoinsieme, diciamo \mathbf{R}_* , così definito: " \mathbf{R}_* contiene 1, la somma, la differenza e il prodotto di due suoi elementi (quindi in particolare $0 = 1-1$), il quoziente di due suoi elementi se il divisore è $\neq 0$, e infine la radice quadrata di un suo generico elemento > 0 ". Benché sia più ampio dell'insieme dei razionali, si intuisce e si può dimostrare che \mathbf{R}_* è ancora *numerabile*; quindi la geometria ad esso connessa è a sua volta una geometria "numerabile". La comparsa dell'operazione di estrazione di radice quadrata, nella definizione di \mathbf{R}_* , si spiega considerando che in uno spazio per cui si accolga (E) vale il teorema di Pitagora (vedi oltre) anche senza (C); e quindi tale estrazione deve essere prevista per poter calcolare la lunghezza dell'ipotenusa di un triangolo rettangolo con cateti unitari.⁴⁷

Se al contrario si accoglie (C) (in una delle sue possibili versioni equivalenti) e si rinuncia a (E) si ottiene uno scenario molto diverso e decisamente più interessante. Innanzitutto, si può dare una definizione di "parallelismo", diciamo "parallelismo*" tra rette p, q (relazione ancora riflessiva e simmetrica), che pur *implicando* la definizione standard data in precedenza (cioè la $(Cop(p,q) \wedge p \cap q = \emptyset) \vee (p=q)$) *non le equivale*, e che stabilisce un limite superiore per il numero di parallele* ad una retta data passanti per un punto dato. Precisamente, si ha il teorema: «data una

⁴⁴ Non a caso, le varie definizioni di continuità di I sono comunemente menzionate come **principi** (di continuità), e sono riferiti a $I = \mathbf{R}$ nella didattica standard della teoria dei numeri reali. Con le opportune modifiche degli "assiomi dimensionali", (C) si applica anche al generico spazio sintetico n -dimensionale.

⁴⁵ In tempi più recenti sono state proposte definizioni alternative della continuità di I , valide per I qualsiasi (anche $I = S$).

⁴⁶ Questo assioma si dimostra equivalente al quinto postulato euclideo. L'acutezza di Euclide nel richiederne la validità *come assioma* non può non colpire profondamente.

⁴⁷ Tuttavia questo è soltanto un esempio appariscente; l'esigenza di estrazione di radice quadrata *permane* infatti anche prescindendo da (E), in forza degli altri assiomi (ad es. dell'assioma (A_{11}) , che conduce a problemi di secondo grado).

retta p ed un punto $X \notin p$, esiste almeno una parallela*, ma anche al più due parallele*, a p passante(i) per X ; se poi $X \in p$, allora esiste esattamente una parallela* a p per X , che è la p stessa». La geometria in oggetto, quella cioè con $(A_0 \div A_{14} + (C))$ come assiomi, si dice **geometria assoluta**. Essa si specializza nella geometria euclidea quando a quegli assiomi si aggiunga (E) ; oppure nella geometria iperbolica (di Bolyai-Lobatchewsky, (BL)) quando vi si aggiunga la negazione $(\neg E)$ di (E) , che ovviamente è “per almeno una retta p e almeno un punto X fuori di essa esistono almeno due rette distinte passanti per X e disgiunte da p .» Nel primo caso, il parallelismo* tra rette *equivale* al loro parallelismo standard (cioè quest’ultimo risulta anche *implicare* il parallelismo*); e inoltre, la relazione di parallelismo così unificata risulta essere anche transitiva, e quindi una equivalenza. Quale che sia la scelta di un successivo assioma, è ben naturale che lo studio di tale geometria assoluta (del piano, dello spazio tridimensionale, del generico spazio sintetico n -dimensionale) meriti il massimo interesse, perché *tutti gli sviluppi* che conseguono dai suoi assiomi restano validi nella geometria euclidea, o in quella iperbolica, o comunque altrimenti si proceda. Illustreremo qui appresso alcuni dei più importanti teoremi della geometria assoluta, ciò che permetterà al lettore di individuare automaticamente le conseguenze *specificamente* “euclidee” della geometria che nasce con l’accoglimento di (E) , ossia della geometria euclidea propriamente detta.

. Per cominciare, abbiamo l’ormai ovvio asserto che «una semiretta, un segmento aperto orientato, una retta orientata, un angolo orientato e un semifascio orientato godono tutti della D -proprietà». Due altri teoremi “gemelli” (che conseguono dal solo $(Arch)$, non richiedendo $(Cant)$), sono i seguenti. «se il segmento libero a [l’angolo libero \mathfrak{A}] è più piccolo ($<$) del segmento libero b [dell’angolo libero \mathfrak{B}], allora per ogni segmento libero c [per ogni angolo libero \mathfrak{C}] esiste un numero diadico positivo δ (cioè del tipo $n \cdot 2^{-k}$) per cui $a < \delta \cdot c < b$ [per cui $\mathfrak{A} < \delta \cdot \mathfrak{C} < \mathfrak{B}$]]. Questi due teoremi possono utilizzarsi per istituire una “teoria della misura” di segmenti o di angoli quando si rinunci, come nel nostro caso, all’assioma di divisibilità. Ma forse il più importante teorema della geometria assoluta, ancora conseguente dal solo $(Arch)$, è quello cosiddetto “di Saccheri (Gerolamo, 1667-1733) e Legendre (Adrien, 1752-1833)”, secondo il quale «la somma di due angoli liberi di un triangolo è minore/uguale all’angolo supplementare del terzo angolo”. Naturalmente ciò che in esso ci appare esotico è l’attributo “minore/uguale” in luogo del familiare “uguale” della geometria euclidea.

Introduciamo ora la nozione di **misura di un segmento** (di dati estremi; o se si preferisce, di misura di una data coppia di punti distinti) come una generica applicazione $\mu: (S^2 \setminus \Delta_S) \rightarrow \mathbf{R}_{>0}$ (dove S è il solito insieme-universo di punti, Δ_S è l’insieme diagonale di S e $\mathbf{R}_{>0}$ è la semiretta reale strettamente positiva) soddisfacente ai requisiti (vedi anche l’App. Gen. D):

1) \mathcal{M} è invariante per congruenza (cioè $(X|Y \approx U|V) \Rightarrow (\mu(X|Y) = \mu(U|V))$); quindi in particolare μ è una funzione simmetrica dei suoi due argomenti;

2) \mathcal{M} è additiva (cioè è tale che $[X,Z,Y] \Rightarrow (\mu(X|Y) = \mu(X|Z) + \mu(Z|X))$).

Manifestamente, la restrizione di d a $S^2 \setminus \Delta_S$ è una misura. Se $\mathbf{a} = [X|Y]$, e μ è una misura di segmenti, diremo **misura del segmento libero** \mathbf{a} , e la denoteremo con $\mu(\mathbf{a})$, il reale positivo $\mu(X|Y)$ (quindi per definizione $\mu([X|Y]) = \mu(X|Y)$). Si verifica facilmente che, con questa definizione, « $\mu(\mathbf{a} + \mathbf{b}) = \mu(\mathbf{a}) + \mu(\mathbf{b})$; $\mathbf{a} < \mathbf{b} \Rightarrow \mu(\mathbf{a}) < \mu(\mathbf{b})$; $\mathbf{a} < \mathbf{b} \Rightarrow \mu(\mathbf{b} - \mathbf{a}) = \mu(\mathbf{b}) - \mu(\mathbf{a})$; e $\mu(\delta \cdot \mathbf{a}) = \delta\mu(\mathbf{a})$ per ogni numero diadico positivo δ ». Sfruttando il primo dei precedenti teoremi gemelli, si dimostra inoltre la seguente proposizione: «per ogni segmento libero \mathbf{a} , una generica misura μ di \mathbf{a} è uguale al prodotto di una “misura di riferimento”, diciamo μ_0 , dello stesso \mathbf{a} , per un reale $\lambda > 0$ indipendente da \mathbf{a} (cioè, per una *costante* $\lambda > 0$)». Esprimeremo questo fatto scrivendo $\mu = \lambda\mu_0$. Il valore di λ , e quindi la stessa μ , è unicamente determinato prescrivendo il valore $\alpha_0 > 0$ di μ per un segmento libero “campione” \mathbf{a}_0 . La condizione che ne risulta, cioè la $\alpha_0 = \lambda\mu_0(\mathbf{a}_0)$, si dice **di normalizzazione di μ** . Se si pone $\alpha_0 = 1$, il campione \mathbf{a}_0 si dirà **unitario** rispetto alla misura μ ; o anche, μ si dirà la misura di **unità** \mathbf{a}_0 . Evidentemente, una misura è del tutto determinata con la scelta della sua unità.

La nozione di misura di coppie di punti distinti è fondamentale in una formalizzazione che, come quella hilbertiana, *non* accolga la nozione di distanza come primitiva; essa non allarga invece l’orizzonte di una formalizzazione metrica, producendo soltanto una generalizzazione banale (moltiplicazione per un fattore > 0) della già disponibile distanza originale. La nozione di misura resta tuttavia importante anche in questo secondo caso di maggior presente interesse, nel senso che essa si applica ad oggetti geometrici *più generali* delle semplici coppie di punti distinti, continuando ad essere definita sotto requisiti (opportunamente modificati e/o generalizzati) di positività, invarianza rispetto alla congruenza, additività e normalizzazione, in sostanziale analogia con quelli introdotti più sopra (v. S.sez. 1.4.1, nonché App. Gen. C). Tornando per un momento ad una formalizzazione in cui la nozione di distanza come primitiva *non è disponibile*, resterebbe ancora da provare l’esistenza (l’unicità è ovvia), in essa, di una misura che valga 1 (diciamo) per il campione \mathbf{a}_0 . Fissato a questo scopo un generico segmento libero \mathbf{a} , suddividiamo l’insieme Δ dei numeri diadici positivi in due classi (disgiunte) $\Delta_1(\mathbf{a})$ e $\Delta_2(\mathbf{a})$ secondo la prescrizione $\delta \in \Delta_1(\mathbf{a}) \Leftrightarrow \delta \cdot \mathbf{a}_0 < \mathbf{a}$, e $\delta \in \Delta_2(\mathbf{a}) \Leftrightarrow \delta \cdot \mathbf{a}_0 \geq \mathbf{a}$; allora $\{\Delta_1(\mathbf{a}), \Delta_2(\mathbf{a})\}$ è una D-sezione di Δ , e quindi individua un unico reale > 0 , diciamo $r(\mathbf{a})$, come suo elemento di separazione. Si prova facilmente che $r(\mathbf{a})$ è una misura di \mathbf{a} , e che $r(\mathbf{a}_0) = 1$; cioè, quanto volevamo accertare. Banalmente, la misura μ che vale α_0 in \mathbf{a}_0 si ottiene infine ponendo α_0 in luogo di λ e r in luogo di μ_0 nella $\mu = \lambda\mu_0$.

Si consideri ora una semiretta p' di origine X_0 , e sia μ una misura (di coppie di punti distinti). In forza della D-proprietà della semiretta, possiamo dimostrare che «per un arbitrario reale $\alpha > 0$ esiste uno e un solo punto $Y \in p'$ per cui $\mu(X_0|Y) = \alpha$ ». A tal scopo, definiamo una separazione $\{p'_1, p'_2\}$ di p' secondo le $X_1 \in p'_1 \Leftrightarrow \mu(X_0|X_1) < \alpha$, e rispettivamente $X_2 \in p'_2 \Leftrightarrow \mu(X_0|X_2) \geq \alpha$. Per la D-proprietà della semiretta, esiste un (unico) elemento di separazione Y di p'_1 e p'_2 . Proviamo ora che $\mu(X_0|Y) = \alpha$. Supponiamo (per assurdo) che sia (ad es.) $\mu(X_0|Y) < \alpha$, e sia $\alpha_0 =: \mu(X_0|Y_0)$ per un certo $Y_0 \in p'$. Allora esiste un δ (diadico positivo) per cui $\mu(X_0|Y) < \delta\alpha_0 < \alpha$. Il punto Z_0 per cui $\mu(X_0|Z_0) = \delta\alpha_0$, esiste e appartiene a p'_1 in quanto $\mu(X_0|Z_0) < \alpha$; ma esso appartiene anche a p'_2 in quanto $\mu(X_0|Y) < \mu(X_0|Z_0)$ equivale a $Y < Z_0$: contraddizione. In modo analogo si proverebbe che non può essere $\mu(X_0|Y) > \alpha$, qed. Questo importante risultato ci assicura che *una misura di una coppia $\{X_0, Y\}$, ove X_0 è l'origine di una semiretta e Y un suo punto arbitrario, e in particolare la distanza $d(X_0, Y)$, è una suriezione di quella semiretta su $\mathbf{R}_{>0}$* . Essendo come sappiamo tale misura anche una iniezione, si conclude che esiste una corrispondenza biunivoca tra la semiretta in oggetto e $\mathbf{R}_{>0}$; e quindi tra una semiretta arbitraria e $\mathbf{R}_{>0}$ stesso.

Con le consuete cautele intorno all'esistenza, quanto sopra si può ripetere con riferimento alle **misure di angoli** (o se si preferisce, di coppie di raggi di comune origine, distinti e non complementari) e alle corrispondenti **misure di angoli liberi**. In questo caso, solitamente si conviene di normalizzare la misura di un angolo libero in modo che essa valga $\pi/2$ per l'angolo libero retto \mathbf{R} (misura **standard** (o **naturale**) degli angoli liberi). Il teorema di Saccheri-Legendre si riformula allora asserendo che «la somma delle misure standard dei tre angoli di un triangolo è $\leq \pi$ ».^{48, 49}

Avendo ormai garantito la biiettività dell'applicazione $d: S^2 \rightarrow \mathbf{R}_{\geq 0}$, possiamo introdurre nella geometria assoluta una ulteriore importante nozione, quella di “similitudine di S su se stesso”.⁵⁰ Una suriezione $\mathcal{S}: S \rightarrow S$ si dirà una **similitudine di S su S** se esiste un reale $\sigma > 0$ per il quale, per ogni coppia $\{X, Y\}$ di punti di S , risulta $d(\mathcal{S}(X), \mathcal{S}(Y)) = \sigma d(X, Y)$. In questo caso \mathcal{S} si dirà **fattore**

⁴⁸ È tuttavia legittimo chiedersi se in una geometria assoluta possano esistere triangoli con somma degli angoli *uguale* a π e triangoli per cui tale somma sia *minore* di π . La risposta è negativa: se esiste un triangolo con somma degli angoli uguale a π , allora ciò vale per ogni triangolo.

⁴⁹ Anticipiamo qui una nozione della quale sarà meglio reso conto nella Sez. 1.4, quella di lunghezza di un arco piano rettificabile, affidandoci per il momento alla sua immagine intuitiva. Facendone uso, si intuisce che a meno di un fattore di normalizzazione la misura dell'angolo $\mathcal{A}(p'|q')$ (ove, ricordiamo, p' e q' hanno comune origine, diciamo O , e sono distinti e non complementari) uguaglia la lunghezza dell'arco del cerchio di centro O e raggio unitario compreso tra p' e q' (questo arco è rettificabile e giace per intero in un semipiano). Questa è del resto la definizione di (misura di un) angolo che si dà comunemente nella trigonometria elementare. Se la misura in oggetto è quella naturale, cioè vale $\pi/2$ per un angolo retto, allora il fattore di normalizzazione più sopra nominato vale 1.

⁵⁰ Usualmente la nozione di similitudine viene introdotta in uno spazio euclideo (se a più di una dimensione), ma quanto segue nel presente paragrafo mostra che ciò può essere fatto anche in un semplice spazio assoluto: in questo caso sono soltanto gli sviluppi della definizione che risultano meno significativi.

(o **coefficiente**) della similitudine \mathcal{S} . Una similitudine è automaticamente iniettiva ($X \neq Y \Rightarrow \mathcal{S}(X) \neq \mathcal{S}(Y)$), e quindi biiettiva. In particolare una similitudine con fattore 1 è una isometria di S su S . Il prodotto (funzionale) di due similitudini di fattori σ e rispettivamente σ' è una similitudine di fattore $\sigma\sigma'$; l'inversa di una similitudine di fattore σ è una similitudine di fattore $1/\sigma$. Segue facilmente che l'insieme delle similitudini (di S su S) è un gruppo di trasformazioni su S . Un sottoinsieme B di S (in questo contesto, una “figura” di S) si dirà **simile** ad una figura A (di S) se è l'immagine di A secondo una similitudine; il fattore σ di questa similitudine si dirà a sua volta **fattore di B rispetto ad A** . La relazione “la figura B è simile alla figura A ” è una equivalenza. Infatti una figura è simile a se stessa (riflessività: la similitudine è allora l'identità, di fattore 1); se B è simile a A con fattore σ , A è simile a B con fattore $1/\sigma$ (simmetria); se B è simile ad A con fattore σ , e C è simile a B con fattore σ' , allora C è simile ad A con fattore $\sigma\sigma'$ (transitività). Poiché tutte le nozioni geometriche fin qui introdotte (salvo quella di ordine, che *non* è strettamente geometrica) sono state definite in termini di uguaglianza o disuguaglianza di distanze, si intuisce e si dimostra che esse sono invarianti rispetto a una generica similitudine. Così abbiamo ad esempio:

- 1) $[X,Z,Y] \Rightarrow [\mathcal{S}(X), \mathcal{S}(Z), \mathcal{S}(Y)]$;
 - 2) L'immagine secondo \mathcal{S} di una retta [di un piano] è una retta [un piano]; quella di una semiretta p' inclusa nella retta p e con origine $X \in p$ è una semiretta inclusa nella retta $\mathcal{S}(p)$ e con origine $\mathcal{S}(X)$. Mutatis mutandis, analoghe conclusioni valgono con “semifascio” in luogo di “retta” e semi-semifascio” in luogo di “semiretta”;
 - 3) Le immagini per similitudine di segmenti [di angoli] congruenti sono congruenti;
 - 4) Se p'^* è la semiretta complementare della p' , $\mathcal{S}(p'^*) = (\mathcal{S}(p'))^*$;
 - 5) L'immagine per similitudine di un angolo è un angolo congruente all'angolo originale; in particolare l'immagine per similitudine di un angolo retto è un angolo retto;
- ...e via dicendo.

È ormai agevole introdurre “coordinate” del generico punto X dello spazio sintetico assoluto $S_{n \geq 1}$; anzi, addirittura due tipi sostanzialmente diversi di coordinate se $n > 1$, quelle cosiddette “assolute” e quelle cosiddette “rettangolari”. Cominciamo dal caso banale della coordinatazione di una retta ($n = 1$), in cui la precedente distinzione (coordinata assoluta vs. coordinata rettangolare) non sussiste. Fissiamo una misura μ (di coppie di punti distinti) su p (ad esempio, ma non necessariamente, la stessa distanza d), e una semiretta p' di origine O su di essa (o se si preferisce un punto $X_0 \neq O$ su p , mediante il quale si ponga $p' =: \mathcal{SR}(O, X_0)$). Definiremo allora la **coordinata** $C(X)$ del generico punto X di p come

$$C(X) =: \mu(O, X) \text{ se } X \in p' \text{ (quindi } X \neq O);$$

$C(X) =: 0$ se $X = O$;

$C(X) =: -\mu(O, X)$ se $X \in p'^*$ (quindi $X \neq O$).

Diremo O l'**origine** e p' la **semiretta positiva** del **sistema di riferimento** C su p . Si ottiene così un'applicazione $C: p \rightarrow \mathbf{R}$, che come sappiamo è anche una biiezione, nonché un isomorfismo rispetto agli ordini $<$ (su p) e $<$ (su \mathbf{R}). Questa applicazione risulta poi un omeomorfismo rispetto alle topologie indotte da μ su p (μ soddisfa infatti agli assiomi metrici) e rispettivamente dalla **distanza pitagorica** su \mathbf{R} (valore assoluto della differenza tra i due reali considerati); e una isometria se si identifica μ con d .

Nel caso di un piano π , fissiamone una retta p corredata di un sistema di riferimento C come sopra descritto (di origine O e di semiretta positiva p'), e uno dei due semipiani di bordo p di π , diciamo π' (ovvero, un punto Y_0 di π , mediante il quale si ponga $\pi' =: \mathcal{SP}(p, Y_0)$). Sia X il generico punto di π e X_p la sua proiezione ortogonale su p ; oltre alla coordinata $C_1(X) =: C(X_p)$ di X , definiremo la sua **coordinata assoluta** $C_2(X)$ secondo le

$C_2(X) =: \mu(X, X_p)$ se $X \in \pi'$;

$C_2(X) =: 0$ se $X \in p$;

$C_2(X) =: -\mu(X, X_p)$ se $X \in \pi'^*$.

Diremo O l'**origine**, p' la **semiretta positiva** e π' il **semipiano positivo del riferimento assoluto** $\langle C_1, C_2 \rangle$ del piano π .

La generalizzazione allo spazio S è ovvia. Si fissa un piano π di S e uno dei semispazi di bordo π (o un suo punto Z_0), diciamo S' , come semispazio positivo. Alla proiezione ortogonale su π del generico punto $X \in S$, X_π , si associano le coordinate $C_1(X)$ e $C_2(X)$ come sopra descritte, e inoltre la **coordinata assoluta** $C_3(X)$ secondo le

$C_3(X) =: \mu(X, X_\pi)$ se $X \in S'$;

$C_3(X) =: 0$ se $X \in \pi$;

$C_3(X) =: -\mu(X, X_\pi)$ se $X \in S'^*$.

Definiti al solito modo, O sarà ancora l'**origine**, p' la **semiretta positiva**, π' il **semipiano positivo**, e S' il **semispazio positivo del riferimento assoluto** $\langle C_1, C_2, C_3 \rangle$ di S .

Anche nei casi del piano π e dello spazio S si ottengono evidentemente delle biiezioni (di π su \mathbf{R}^2 , e rispettivamente di S su \mathbf{R}^3), che sono omeomorfismi rispetto alle topologie indotte da μ (su π e rispettivamente su S) e dalla distanza pitagorica (su \mathbf{R}^2 e rispettivamente su \mathbf{R}^3). Con le dovute modifiche degli assiomi dimensionali, si potrebbe analogamente continuare per spazi assoluti di dimensione maggiore di 3.

Passiamo ora alle **coordinate rettangolari**, e cominciamo dal caso di un piano π . Esse saranno definite analogamente alle coordinate cartesiane ortogonali della S.sez. 1.1.2: vale a dire, per dato $X \in \pi$, le coordinate rettangolari sono le coordinate $\xi_1 =: C(X_1)$ e $\xi_2 =: C(X_2)$ delle proiezioni ortogonali X_1 e X_2 di X su due rette orientate p_1 e p_2 di π , tra loro perpendicolari e incidenti in O , origine comune delle corrispondenti semirette positive. In questo caso, tuttavia, *non* si può più parlare di biiezione di π su \mathbf{R}^2 , ossia non si può affermare che una generica coppia ordinata di reali $\langle \eta_1, \eta_2 \rangle$ si identifichi con quella delle coordinate rettangolari di qualche $X \in \pi$. Ciò in quanto gli assiomi della geometria assoluta *non ci consentono* di affermare che la perpendicolare a p_1 passante per $C^{-1}(\eta_1)$ e la perpendicolare a p_2 passante per $C^{-1}(\eta_2)$ si intersechino (in un punto di π , come sarebbe vero in un piano euclideo). In conclusione un riferimento rettangolare *non è una suriezione* di π su \mathbf{R}^2 . Il caso dello spazio S si tratta in modo analogo, facendo uso dei piani passanti per X e perpendicolari alle tre rette coordinate p_1, p_2, p_3 .⁵¹

Benché non suriettive, le coordinate rettangolari sono manifestamente “simmetriche”, nel senso che sono definite tutte allo stesso modo. Viceversa le coordinate assolute sono suriettive ma non simmetriche. Se cioè (riferendoci al caso del piano π) si prende la perpendicolare q alla retta p passante per la sua origine O , e la sua semiretta positiva q' nel precedente semipiano π' , e poi si definisce $C_{*2}(X)$ come la coordinata lungo q della proiezione ortogonale X_q (di $X \in (\pi \setminus q)$ su q), e la coordinata $C_{*1}(X)$ mediante la solita regola – usando come semipiano positivo, dei due semipiani in cui q divide $\pi \setminus q$, quello che contiene la precedente semiretta p' –, non si ottengono in generale le *precedenti* coordinate assolute di π , ossia non è necessariamente $C_{*i}(X) = C_i(X)$, $i = 1, 2$. Entrambe queste discrepanze (coordinate rettangolari simmetriche ma non suriettive, coordinate assolute suriettive ma non simmetriche), rispetto alle analoghe e familiari situazioni della geometria euclidea, sono riconducibili alla seguente circostanza. In un piano assoluto π sia data la coppia di punti distinti $\{X, Y\}$; si prendano le perpendicolari $r \subset \pi$ per X e $s \subset \pi$ per Y alla retta p passante per X e Y , e su queste r ed s si prendano – dalla stessa parte rispetto a p – punti $U \neq X$ e rispettivamente $V \neq Y$ con $U|X \approx V|Y$. In generale non risulta allora $X|Y \approx U|V$; né la retta q per U e V risulta perpendicolare alla r e/o alla s . Il quadrilatero di vertici X, Y, V, U appena descritto, che dunque *non è* in generale un rettangolo, è noto come **quadrilatero di Saccheri**, e gode di proprietà caratteristiche.

Abbandoniamo finalmente la geometria assoluta aggiungendo ai suoi assiomi il quinto postulato euclideo nella forma (E). Lo spazio euclideo 3-dimensionale così ottenuto sarà ormai

⁵¹ Sia le coordinate assolute che quelle rettangolari possono generalizzarsi adottando unità di misura diverse per le diverse coordinate. In questo caso parleremo di coordinate assolute, o rispettivamente rettangolari, “generalizzate”.

denotato con $H \equiv H_3$, e in generale con $H_{n \geq 2}$ nel caso ($n \geq 2$)-dimensionale. Poiché i teoremi derivabili nella nuova situazione sono tutti e soli quelli della familiare geometria elementare, non occorrerà menzionarne una gran scelta, essendo sufficiente limitarsi a quelli (in parte già anticipati) che appaiono più significativi nel presente contesto. Innanzitutto il teorema di Saccheri-Legendre viene modificato nel familiare «la somma di due angoli di un triangolo è *uguale* al supplementare del terzo angolo», ovvero, usando la misura naturale per gli angoli, «la somma dei tre angoli di un triangolo è *uguale* a π ». In secondo luogo, «l'essere parallele, per le rette di un piano, è una relazione *anche* transitiva, e quindi una equivalenza». Le classi di equivalenza corrispondenti si dicono **direzioni** del piano. Lo stesso fatto, e la stessa definizione, valgono ovviamente anche nello spazio. In forza dell'ultimo teorema di transitività, date due rette p, q incidenti, ed un punto qualsiasi X del loro piano, la retta passante per X e parallela alla p interseca la retta q in un (unico) punto $X_{q|p}$ che si dirà la **proiezione parallela di X su q secondo p** . Se in particolare p e q sono perpendicolari, le proiezioni parallele del generico punto X (del loro piano π) sull'una secondo l'altra, cioè $X_{p|q}$ e $X_{q|p}$, coincidono con le sue proiezioni ortogonali X_p su p e rispettivamente X_q su q . Ad ogni coppia ordinata di coordinate *rettangolari* della geometria assoluta del piano π corrisponde adesso un punto di π , perché la perpendicolare a p e la perpendicolare a q per loro punti arbitrari si intersecano in esattamente un punto di π (il quadrilatero di Saccheri è un rettangolo!); quelle coordinate rettangolari, ormai suriettive, si diranno **coordinate cartesiane ortogonali** di X . È infine evidente che anche le coordinate assolute di $X \in \pi$, così come sono state definite più sopra, coincidono con le sue coordinate cartesiane ortogonali. Quanto precede si generalizza immediatamente ad uno spazio euclideo ($n > 2$)-dimensionale.

Tornando al caso di rette p e q incidenti (ma non perpendicolari in generale), sia ora r una terza retta del loro piano non parallela a p (in particolare r può dunque essere parallela a q), e siano X, Y, Z tre suoi generici punti distinti tra loro, e $X_{q|p}, Y_{q|p}, Z_{q|p}$ le loro proiezioni parallele su q secondo p . In analogia con il caso delle proiezioni ortogonali, vale il teorema « $[X, Y, Z] \Rightarrow \Rightarrow [X_{q|p}, Y_{q|p}, Z_{q|p}]$; inoltre, risulta $d(X, Y)/d(X_{q|p}, Y_{q|p}) = d(Y, Z)/d(Y_{q|p}, Z_{q|p}) = d(X, Z)/d(X_{q|p}, Z_{q|p})$ ». Come sappiamo, quando in una biiezione di un sistema A su un sistema B di punti le distanze tra tutte le coppie di punti distinti di B sono proporzionali secondo uno stesso fattore $\sigma > 0$ a quelle tra le coppie corrispondenti di A , quella biiezione si dice una similitudine di B ad A . Quindi il precedente teorema può enunciarsi dicendo che «la proiezione parallela di una retta su una retta secondo una terza retta non parallela alle prime due, è una similitudine»; risultato noto come **teorema di Talete**. (Questa similitudine è una isometria se r e q sono parallele.) Una conseguenza di questo teorema è che «due triangoli sono simili sse i loro angoli sono uguali». A questo punto diventa immediata la dimostrazione del **teorema di Pitagora**: «se nel triangolo di vertici (distinti)

$\{X,Y,Z\}$ l'angolo in Y è retto, $d(X,Y)^2 + d(Y,Z)^2 = d(X,Z)^2$): basta infatti osservare che, se V è la proiezione ortogonale di Y sul lato opposto, i triangoli di vertici $\{X,Y,Z\}$, $\{X,V,Y\}$, $\{Y,V,Z\}$ sono simili a due a due.

In uno spazio euclideo «esiste esattamente un piano ρ parallelo ad un piano π dato passante per un punto X dato; se $X \in \pi$, allora $\rho = \pi$.» La relazione “ ρ è parallelo a π ”, oltre che riflessiva e simmetrica, «è transitiva, e quindi è una equivalenza». La relativa classe di equivalenza di piani si dice **giacitura**; ma questa nozione non è realmente nuova, nel senso che una giacitura è in corrispondenza biunivoca con la direzione ad essa perpendicolare.

La corrispondenza biunivoca tra il piano π e \mathbf{R}^2 si conserva se al posto delle coordinate cartesiane ortogonali si usano le coordinate (generalmente con unità diverse) delle proiezioni parallele $X_{p|q}$ e $X_{q|p}$ di $X \in \pi$ (su p secondo q e rispettivamente su q secondo p) avendo supposto tali rette coordinate p , q di π semplicemente incidenti nella comune origine O delle prescelte corrispondenti semirette positive p' e q' .⁵² Le coordinate ottenute si diranno **coordinate cartesiane (generalmente) oblique del piano**. In modo analogo si procede per le **coordinate cartesiane oblique dello spazio**, servendosi di piani per X paralleli ai piani coordinati, ecc. I teoremi di Talete e di Pitagora permettono allora di giustificare le affermazioni anticipate senza dimostrazione nella S.sez. 1.1.2 circa le coordinate cartesiane oblique. Vale a dire, se x_i (con $i = 1,2,3$, o in generale $i = 1, \dots, n$) sono coordinate cartesiane ortogonali di X , e ξ_i ne sono generiche coordinate oblique, per il teorema di Talete deve essere

$$(1) \quad x_i = \sum_j a_{ij} \xi_j + x_i^o,$$

dove a_{ij} sono n^2 costanti soddisfacenti a $\det\{a_{ij}\} \neq 0$, e x_i^o sono n altre costanti; quindi se Δ denota differenza,

$$(1') \quad \Delta x_i = \sum_j a_{ij} \Delta \xi_j.$$

In forza poi del teorema di Pitagora, facendo $\mathcal{M} = d$ nella definizione delle coordinate cartesiane ortogonali risulta $d^2(X,Y) = \sum_i (\Delta x_i)^2$, ove con Δx_i si è indicata la differenza tra la i -ma coordinata ortogonale di Y e quella di X . Sostituendo in questa la (1'), abbiamo $d^2(X,Y) = \sum_{i,j,k} a_{ij} \Delta \xi_j a_{ik} \Delta \xi_k \equiv \sum_{j,k} g_{jk} \Delta \xi_j \Delta \xi_k$, avendo posto $g_{jk} = \sum_i a_{ij} a_{ik} \equiv g_{kj}$; ossia, la tesi (cfr. S.sez 1.1.2) che d^2 è una forma quadratica definita positiva nelle differenze delle coordinate cartesiane oblique omologhe. In particolare, questa forma risulta pitagorica sse

$$(2) \quad \sum_i a_{ij} a_{ik} = \delta_{jk} \text{ (simbolo di Kronecker).}$$

Una trasformazione affine del tipo (1) per cui vale la (2) si dice **ortonormale**, ed è automaticamente non singolare, in quanto (come subito si verifica) $\det^2\{a_{ij}\} = 1$; essa corrisponde ad una generica rototraslazione del sistema degli assi coordinati ortogonali.

Se una $(n \times n)$ -matrice (matrice di ordine n) ad elementi a_{ij} in generale complessi soddisfa il sistema di $n(n+1)/2$ equazioni indipendenti (2), il quadrato del suo determinante vale comunque 1, ed essa continua a dirsi “ortonormale”. Si dimostra poi⁵³ che il sistema (2) equivale al sistema

$$(2') \quad \sum_i a_{ji} a_{ki} = \delta_{jk},$$

in cui la somma è fatta sugli indici di colonna invece che di riga come nella (2).

⁵² Inoltre $O|X_{p|q} \approx X_{q|p}|X$ e $O|X_{q|p} \approx X_{p|q}|X$, cioè il parallelogramma di vertici $O, X_{p|q}, X, X_{q|p}$ ha i lati opposti uguali. Naturalmente si può anche dire che se $p \subset \pi$ e $q \subset \pi$ sono incidenti in O e si manda la parallela a q per $X \in p$ e la parallela a p per $Y \in \pi$, tali parallele si intersecano in un punto $Z \subset \pi$ per il quale $O|X \approx Y|Z$ e $O|Y \approx X|Z$.

⁵³ Trascurando il segno di sommatoria, partiamo dalle (2), $a_{ij} a_{ik} = \delta_{jk}$, e moltiplichiamole per a_{hk} sottintendendo la somma sugli indici ripetuti: si ottiene $a_{ij} a_{ik} a_{hk} = a_{hj}$. Questa è moltiplicata a sua volta per la matrice inversa della precedente (che esiste unica perché il relativo determinante vale ± 1), di elementi b_{jp} . Allora a 1° membro abbiamo $a_{ij} b_{jp} a_{ik} a_{hk}$ e a 2° membro $a_{hj} b_{jp}$. Possiamo dunque scrivere $\delta_{ip} a_{ik} a_{hk} = a_{pk} a_{hk} = \delta_{hp}$, che è quanto volevamo dimostrare.

1. 3) L'ISOMORFISMO TRA H_n E LO SPAZIO CARTESIANO REALE n -DIMENSIONALE \mathbf{R}^n

1.3.1) \mathbf{R}^n COME MODELLO NORMALE DI H_n : PARTE PRIMA

Al costo di un piccolo tour de force, abbiamo sostanzialmente concluso l'illustrazione della nostra formalizzazione "metrica" della geometria euclidea. Restano da trattare due importanti questioni, quella dell'"estensione" o "misura" di certe figure degli spazi euclidei $H_{n \geq 1}$, e quella del loro "orientamento", questioni che rinviemo tuttavia alla Sez. 1.4, allo scopo dedicata.

In questa sezione ci proponiamo invece di approfondire l'esame dell'isomorfismo (rispetto a *tutte* le strutture di interesse) tra lo spazio H_n (in particolare per $1 \leq n \leq 3$) e lo spazio delle sue n coordinate cartesiane ortogonali corrispondenti ad un prefissato sistema di riferimento con misura $\mu = d$. Fin qui $\mathbf{R}^{n>1}$, potenza cartesiana n -ma di base \mathbf{R} , è stato pensato come insieme di n -ple ordinate privo di struttura (\mathbf{R} ha invece la sua struttura standard completa). Ma $\mathbf{R}^{n>1}$ possiede in realtà una struttura canonica piuttosto ricca: ad esempio, esso è uno spazio metrico assegnando come distanza tra due suoi elementi la distanza pitagorica d_{pit} ; e come sappiamo, H_n e \mathbf{R}^n così metrizzato sono isometrici (ciò vale evidentemente anche per $n = 1$). La struttura canonica di $\mathbf{R}^{n>1}$ è quella di uno "spazio lineare (su \mathbf{R}) n -dimensionale dotato di prodotto interno euclideo" (\equiv simmetrico e definito-positivo se i due fattori sono uguali). Nel seguito di questo capitolo sottintenderemo di regola la specificazione di "euclideo" riferita al prodotto interno. Tale struttura canonica è ben nota, ma vale la pena di ricordare brevemente di cosa si tratta.

Per cominciare, ricordiamo che uno spazio lineare su \mathbf{R}^1 , diciamo E , è un insieme sul quale è definita un'applicazione **somma** di $E \times E$ in E (denotata $+$, e da scrivere "tra"), un'applicazione **prodotto** di $\mathbf{R} \times E$ in E (denotata mediante semplice giustapposizione, diciamo con il reale a sinistra dell'elemento di E), e un'applicazione $-$ di E in E , soggette alle seguenti proprietà (per ogni x, y, \dots di E e ogni α, β, \dots di \mathbf{R}):

- (a) $x + y = y + x$ (**simmetria di $+$**),
- (b) $(x+y) + z = x + (y+z)$ (**associatività di $+$**),
- (c) $x + 0 = x$ (**definizione dello zero (0) additivo**),
- (d) $x + (-x) = 0$ (**definizione dell'inverso $-x$ di x**)

¹ Salvo avvertenza in senso contrario, ogni spazio lineare di cui ci occuperemo in questo libro sarà supposto su \mathbf{R} , e quindi non menzioneremo questo fatto volta per volta.

(quindi E è un gruppo additivo (rispetto a $+$) abeliano (cioè simmetrico) con zero che abbiamo denotato 0 e inverso che abbiamo denotato con l'anteposizione di $-$)²; e inoltre,

- (e) $(\alpha\beta)x = \alpha(\beta x)$ (**associatività del prodotto**)
- (f) $(\alpha+\beta)x = \alpha x + \beta x$ (**distributività del primo fattore**)
- (g) $\alpha(x+y) = \alpha x + \alpha y$ (**distributività del secondo fattore**)
- (h) $1x = x$ (**ruolo dell'unità 1 di \mathbf{R}**).

E si dice poi dotato di **prodotto interno** (euclideo) se è definita un'applicazione di $E \times E$ in \mathbf{R} (denotata con (\cdot) , e da scrivere "tra") soggetta alle proprietà:

- (i) $x \cdot y = y \cdot x$ (**simmetria di (\cdot)**);
- (l) $(x+y) \cdot z = (x \cdot z) + (y \cdot z)$ (**distributività di (\cdot) rispetto al primo fattore**);
- (m) $(\alpha x) \cdot y = \alpha(x \cdot y)$ (**associatività di (\cdot) rispetto al primo fattore**);

(Si noti che (l) e (m) insieme affermano la **linearità di (\cdot) rispetto al primo fattore**, e quindi per (i) anche rispetto al secondo, o la **bilinearità di (\cdot)** .)

- (n) $x \cdot x \geq 0$, con $x \cdot x = 0$ solo se $x = 0$ (**positività definita di (\cdot)**).

Al prodotto interno si associa un **modulo** (o **norma**³) del generico x di E , da denotare $|x|$, definito da $|x| =: (x \cdot x)^{1/2}$. Il prodotto interno $x \cdot y$ può esprimersi per mezzo di moduli, e precisamente mediante la formula di immediata giustificazione $2(x \cdot y) = |x+y|^2 - (|x|^2 + |y|^2)$; per cui esso si sarebbe potuto *definire* mediante quest'ultima, imponendo a $| \cdot |$ le proprietà di norma (vedi la precedente nota a piè pagina). Se $x \cdot y = 0$, x e y si dicono tra loro **ortogonali**.

E (spazio lineare) si dice "**n-dimensionale**" (o **n-dim**); e in tal caso si denota comunemente con E_n se esiste una famiglia $\{e_i\}_{i=1, \dots, n}$ di suoi elementi al contempo **libera** (o **linearmente indipendente**) (\equiv se α_i sono n reali qualsiasi, $\sum_i \alpha_i e_i = 0$ (somma su i da 1 a n) $\Rightarrow \alpha_i = 0 \forall i = 1, \dots, n$) e **completa** (\equiv qualsiasi elemento x di E può porsi nella forma $x = \sum_i x_i e_i$, dove gli x_i sono n reali, unicamente determinati se la famiglia $\{e_i\}_{i=1, \dots, n}$ è libera). La definizione è consistente, perché se esistessero due famiglie di E , diciamo $\{e_i\}_{i=1, \dots, n}$ e $\{f_j\}_{j=1, \dots, m}$, entrambe libere e complete, sarebbe $n = m$. Una famiglia $\{e_i\}_{i=1, \dots, n}$ libera e completa di $E \equiv E_n$ si dice una sua **base**; se in

² Confidiamo che la scelta dei simboli $+$, 0 , $-$, che sono gli stessi che si usano per denotare la somma, lo zero e l'inverso in \mathbf{R} , non sia fonte di confusione.

³ Non a caso tale modulo soddisfa gli **assiomi di norma** (denoteremo una norma con $\| \cdot \|$) in uno spazio lineare, che sono: 1) $\|x\| \geq 0$, con $\|x\| = 0$ solo se $x = 0$; 2) $\|\alpha x\| = |\alpha| \|x\|$; 3) $\|x+y\| \leq \|x\| + \|y\|$. Le prime due proprietà di $\| \cdot \|$ sono evidenti, mentre la terza si dimostra a partire dalla $|x \cdot y| \leq |x| |y|$ (nota come **disuguaglianza di Cauchy-Schwarz**), che discende a sua volta dalle definizioni di \cdot e di $| \cdot |$. Se definiamo la distanza tra due elementi di E come il modulo della loro differenza, E diventa automaticamente uno spazio lineare metrico. Così E è anche uno spazio topologico con la topologia indotta da questa metrica (vedi App. Gen. B); e poiché le operazioni di somma e di prodotto per un reale risultano continue rispetto alla topologia di E ed a quella standard di \mathbf{R} (come è facile verificare), in conclusione E è uno **spazio lineare topologico**.

particolare E_n è dotato di prodotto interno, una sua base $\{e_i\}_{i=1, \dots, n}$ si dice **ortonormale** se $e_i \cdot e_j = \delta_{ij}$ per ogni $i, j = 1, \dots, n$. E_n ha *comunque* una base ortonormale, che può ottenersi mediante una conveniente procedura (vedi Sez. 3.2) partendo da una sua base qualsiasi.⁴

Un'applicazione A di uno spazio lineare E in E stesso si dice **lineare** se, per ogni x, y di E e ogni α di \mathbf{R} , risulta:

$$(o) \quad A(x+y) = A(x) + A(y),$$

$$(p) \quad A(\alpha x) = \alpha A(x).$$

Se questo è il caso, $A(x)$ si scrive anche, comunemente, Ax . Se è una *biiezione*, A (lineare) si dice un **automorfismo** (o **sostituzione lineare**) di E . Se $E \equiv E_n$ è n -dim, $\{e_i\}_{i=1, \dots, n} \equiv \{e\}$ è una sua base e A è un suo automorfismo, allora la famiglia $\{Ae_i \equiv e'_i\}_{i=1, \dots, n} \equiv \{Ae \equiv e'\}$ è anch'essa una sua base; la $(n \times n)$ -matrice reale $\{A_{jh}\}$ per cui

$$(1) \quad e'_h \equiv Ae_h = \sum_j A_{jh} e_j \quad (h = 1, \dots, n, \text{ somma da } 1 \text{ a } n)$$

si dice allora **matrice di A nella base $\{e\}$** , e risulta non singolare. Viceversa, per una $(n \times n)$ -matrice non singolare $\{A_{jh}\}$ ed una base $\{e\}$ di E_n , avendo posto $x = \sum_i x_i e_i$ si definisce un corrispondente automorfismo A di E_n secondo $Ax =: \sum_{jh} x_h A_{jh} e_j \equiv \sum_j x'_j e_j$ (**automorfismo della matrice $\{A_{jh}\}$ nella base $\{e\}$**), ove

$$(2) \quad x'_j =: \sum_h x_h A_{jh} \quad (j = 1, \dots, n; \text{ tutte le somme sono da } 1 \text{ a } n).$$

Si noti che nella (1) la somma è fatta rispetto al primo indice della matrice, e nella (2) rispetto al secondo. In conclusione, per una data base di E_n vi è corrispondenza biunivoca tra automorfismi e $(n \times n)$ -matrici non singolari. Il fatto che un automorfismo A di E_n possibilmente trasformi una sua base ortonormale in una sua base anch'essa ortonormale è una proprietà intrinseca di A (vale infatti partendo da qualsiasi base ortonormale), e corrisponde alla proprietà intrinseca della matrice $\{A_{jh}\}$, valida in una *qualsiasi* base ortonormale di E_n , che sia

$$(3) \quad \sum_j A_{jh} A_{jk} = \delta_{hk}$$

cioè alla (2) della S.sez. 1.2.4, come giustifica un momento di riflessione.

Automorfismi e relative matrici, aventi la proprietà di trasformare una base ortonormale (di uno spazio n -dim) in una sua base ortonormale, quindi le seconde soggette alla (3), si dicono **ortogonali** (o più precisamente **ortonormali**).⁵ In conclusione gli automorfismi ortogonali di E_n

⁴ Lo spazio descritto fornisce un esempio elementare di **spazio di Hilbert** (su \mathbf{R}). Come presumiamo noto al lettore, uno spazio di Hilbert è in generale su \mathbf{C} (campo dei numeri complessi), il suo prodotto interno non è simmetrico ma simmetrico-hermitiano (restando tuttavia positivo-definito), la sua dimensione è in generale numerabilmente infinita, ed occorrono altri due assiomi per definirlo completamente.

⁵ Come abbiamo già segnalato (v. S.sez. 1.2.4), matrici soddisfacenti alla (3) sono automaticamente non-singolari in quanto dalla (3) consegue $\det^2 \{A_{jh}\} = 1$. In senso stretto, una matrice ortogonale è quella per cui il 1° membro della (3) è nullo per $h \neq k$.

conservano i prodotti interni, perché $x' \cdot y' = \sum_i x'_i e'_i \cdot \sum_j y'_j e'_j = \sum_i x'_i y'_i$, e similmente $x \cdot y = \sum_i x_i y_i$, e $\sum_i x'_i y'_i = \sum_i x_i y_i$ in forza delle (2) e (3).⁶ Le $(n \times n)$ matrici ortogonali si dicono comunemente **rotazioni**; **proprie** (o **rotazioni tout court**) aut **improprie** (o **invertenti**) a seconda che sia $\det\{A_{ij}\} = +1$ aut $\det\{A_{ij}\} = -1$.

Venendo alla struttura canonica di $\mathbf{R}^{n>1}$, insieme delle n -ple ordinate di reali $x \equiv \langle x_1, x_2, \dots, x_n \rangle$, $y \equiv \langle y_1, y_2, \dots, y_n \rangle$, $z \equiv \text{ecc.}$, esso è uno spazio lineare (su \mathbf{R}) del quale denoteremo nel modo standard la somma, lo zero, l'inverso additivo e la moltiplicazione per un reale se si danno le seguenti definizioni (in termini della analoga struttura di \mathbf{R}):

$$(4) \quad x + y =: \langle x_1 + y_1, \dots, x_n + y_n \rangle,$$

$$(5) \quad 0 =: \langle 0, 0, \dots, 0 \rangle \quad (n\text{-pla ordinata di zeri}),$$

$$(6) \quad \alpha x =: \langle \alpha x_1, \dots, \alpha x_n \rangle,$$

$$(7) \quad -x =: (-1)x \quad (\equiv \langle -x_1, \dots, -x_n \rangle \text{ per la (6)}).$$

Inoltre $\mathbf{R}^{n>1}$ è anche dotato di prodotto interno euclideo, che denoteremo ancora nel modo standard, mediante la definizione:

$$(8) \quad x \cdot y =: \sum_i x_i y_i \quad (\text{somma su } i \text{ da } 1 \text{ a } n).$$

Nei prossimi quattro paragrafi n sarà supposto > 1 , salvo osservare alla fine che il loro contenuto vale banalmente anche per $n = 1$. Si verifica dunque immediatamente che gli assiomi (a) ÷ (n) di spazio lineare con prodotto interno euclideo sono soddisfatti da \mathbf{R}^n con le definizioni (4÷8). Inoltre il modulo della differenza $y - x$ è ora la distanza pitagorica tra y e x , e quindi \mathbf{R}^n è uno spazio metrico con questa distanza. Infine, con queste definizioni \mathbf{R}^n risulta n -dim, perché la famiglia delle n n -ple ordinate $\langle \delta_{i1}, \dots, \delta_{in} \rangle$ (ove $i = 1, \dots, n$) ne è una base, e precisamente una base ortonormale (anche questo si verifica immediatamente) che si dice **base canonica**. D'altra parte, qualunque spazio lineare (su \mathbf{R}) con prodotto interno euclideo e n -dim E_n è in corrispondenza biunivoca con \mathbf{R}^n sotto le definizioni (4÷8), e ad esso isomorfo. Basta allo scopo fissare una base ortonormale $\{e\}$ di E_n (come è *sempre* possibile); con questa, il generico x di E_n è unicamente

⁶ Se $n = 2$, la più generale matrice ortogonale con $\det\{A_{ij}\} = 1$ è $A_{11} = \cos\varphi$, $A_{12} = -\sin\varphi$, $A_{21} = \sin\varphi$ e $A_{22} = \cos\varphi$, dove φ è un reale qualsiasi. Lo scambio di $_1$ con $_2$ non muta sostanzialmente la situazione, perché equivale ad invertire il segno di φ . Sempre per $n = 2$, la più generale matrice ortogonale con $\det\{A_{ij}\} = -1$ si ottiene dalla precedente invertendo il segno di *una* sua riga o colonna, ciò che non si può mai ricondurre ad una conveniente alterazione di φ . Se $n > 2$, la più generale matrice ortogonale con $\det\{A_{ij}\} = 1$ si ottiene applicando la stessa regola soltanto a prefissati indici r, s , con $1 \leq r < s \leq n$, secondo le precedenti formule in cui si ponga r in luogo di 1 e s in luogo di 2 , e inoltre $A_{hk} = \delta_{hk}$ per $h \neq r, s$ e $k \neq r, s$; e infine moltiplicando tra loro le matrici così ottenute per ogni scelta della coppia $r < s$, ciascuna col suo *specifico* parametro $\varphi \equiv \varphi_{rs}$ (se $\varphi_{rs} = 0$, la corrispondente matrice è ovviamente unitaria). Ad es. per $n = 3$ abbiamo 3 siffatte matrici "fattori" (o **rotazioni elementari**), e quindi altrettanti parametri indipendenti; per n generico ne abbiamo $n(n-1)/2$. Da ultimo, per $n > 2$ la più generale matrice ortogonale con $\det\{A_{ij}\} = -1$ si ottiene dalla precedente invertendovi il segno di un numero *dispari* di righe e/o colonne. Nel caso $n = 3$, i tre sopravvisti parametri indipendenti possono ad esempio identificarsi con i noti "angoli di Eulero" della geometria elementare.

rappresentabile come $\sum_i x_i e_i$, e la corrispondenza biunivoca $x \leftrightarrow \langle x_1, \dots, x_n \rangle$ rende i due spazi isomorfi, ed anche isometrici se la distanza in E_n è il modulo della differenza e quella in \mathbf{R}^n è la distanza pitagorica. In conclusione, \mathbf{R}^n sotto le (4 ÷ 8) e E_n coincidono a meno di uno specifico isomorfismo.

Come sappiamo, l'automorfismo ortogonale A di \mathbf{R}^n che porta x in $x' = Ax$ ne conserva il prodotto interno (e quindi il modulo, o viceversa). Si conclude che un automorfismo ortogonale di \mathbf{R}^n ne conserva la distanza (pitagorica), e quindi è una isometria (lineare) di \mathbf{R}^n . Vale anche il viceversa: se cioè si richiede ad un automorfismo A di \mathbf{R}^n di essere una isometria (rispetto alla distanza pitagorica), allora A è ortogonale.

Generalizzeremo adesso la (2) passando ad una particolare trasformazione affine, cioè alla

$$(9) \quad x'_j = \sum_h x_h A_{jh} + c_j,$$

dove $\{A_{jh}\}$ è ortogonale come nella (2), e dove c_j sono n reali arbitrari e non tutti nulli. Anche questa definisce una biiezione $x \leftrightarrow x'$, che evidentemente, tuttavia, a differenza della (2) *non* conserva i moduli. Mostriamo ora che si può definire uno spazio lineare in tutto simile a \mathbf{R}^n (quindi n -dim e con prodotto interno), e legato in modo particolare a \mathbf{R}^n , nel quale i moduli sono conservati a fronte della trasformazione affine (9). Basta considerare la relazione binaria (che denoteremo \approx e scriveremo “tra”) tra coppie ordinate di elementi di \mathbf{R}^n espressa da $\langle \langle x, y \rangle \approx \langle z, u \rangle \Leftrightarrow \langle y - x = u - z \rangle$. La relazione \approx è manifestamente una equivalenza, e quindi decompone $\mathbf{R}^n \times \mathbf{R}^n$ in classi di equivalenza (disgiunte) che si dicono **vettori associati a \mathbf{R}^n** e si denoteranno u, v, w, \dots . Se $\langle x, y \rangle \in u$, se cioè $\langle x, y \rangle$ è un rappresentante di u , a u corrisponde l' n -pla ordinata $u_{\#} =: y - x$, che *non* dipende dal rappresentante $\langle x, y \rangle$ di u prescelto. In altre parole, *pur essendo oggetti concettualmente distinti*, u e $u_{\#}$ sono in corrispondenza biunivoca (\leftrightarrow), perché qualunque rappresentante di u corrisponde alla stessa n -pla ordinata di reali $u_{\#}$ (n -pla ordinata delle **componenti di u**), e viceversa. Lo spazio dei vettori così definiti riceve la stessa struttura di \mathbf{R}^n nel modo standard: se cioè $u \leftrightarrow u_{\#}$, $v \leftrightarrow v_{\#}$, la somma $u + v$ si definisce mediante la $u + v \leftrightarrow u_{\#} + v_{\#}$, il prodotto αu mediante la $\alpha u \leftrightarrow \alpha u_{\#}$, e il prodotto interno $u \cdot v$ mediante la $u \cdot v = u_{\#} \cdot v_{\#}$ (quindi il modulo $|u|$ come $|u_{\#}|$; o viceversa). Lo spazio dei vettori associato a \mathbf{R}^n , che sarà denotato con \mathbf{R}^n , è dunque n -dim con prodotto interno, e il suo zero corrisponde (\leftrightarrow) allo zero di \mathbf{R}^n . È poi chiaro che il generico elemento u di \mathbf{R}^n *non* risente del componente “traslatorio” c_j della trasformazione affine (9): infatti se $\langle x, y \rangle \in u$ e $\langle x', y' \rangle \in u'$, e se, secondo la (9), $x' = Ax + c$, $y' = Ay + c$, $y' - x' = A(y - x)$; ossia, essendo $u'_{\#} =: y' - x'$ e $u_{\#} =: y - x$, risulta $u'_{\#} = Au_{\#}$. Inoltre l'ortogonalità di A equivale a $|u'_{\#}| = |u_{\#}|$ (e quindi a $u'_{\#} \cdot v'_{\#} = u_{\#} \cdot v_{\#}$, o viceversa) e per definizione a $|u'| = |u|$ (e quindi a $u' \cdot v' = u \cdot v$, o viceversa). Un vettore di \mathbf{R}^n di modulo unitario (\equiv uguale a 1) si dice **vettore unitario** o **versore**.

Seguendo la convenzione (già adottata in situazioni analoghe) di denotare una classe di equivalenza mediante un suo rappresentante posto tra $[\]$, sia adesso $u = [\langle x, y \rangle]$ e $v = [\langle z, t \rangle] = [\langle y, w \rangle]$, avendo posto $w =: t + y - z$. Con questa v , risulta dunque $u + v \leftrightarrow u_{\#} + v_{\#} = [y - x + w - y] = [w - x]$, cioè $u + v = [w - x]$. Quest'ultima relazione descrive il vettore $u + v$ mediante il **trasporto per parallelismo** del rappresentante $\langle z, t \rangle$ di v (che viene così sostituito dal rappresentante $\langle y, w \rangle$) secondo la ben nota “regola del parallelogramma” per la somma di due vettori. Come preannunciato, tutto ciò vale banalmente anche per $n = 1$.

Mentre lo spazio $\mathbf{R}^{n \geq 1}$ associato a $\mathbf{R}^{n \geq 1}$ è univocamente definito a partire da quest'ultimo, evidentemente non vale il contrario: infiniti spazi di n -ple ordinate, legati l'uno all'altro da una **traslazione**, hanno per associato spazio vettoriale lo stesso \mathbf{R}^n . Questi spazi di n -ple ordinate sono i cosiddetti **spazi affini associati a \mathbf{R}^n** , e risultano tra loro “distinti” in un senso particolare che illustriamo qui appresso. Partiamo ancora da \mathbf{R}^n , fissiamone arbitrariamente un' n -pla ordinata o , e consideriamo l'applicazione di \mathbf{R}^n in se stesso definita da $x \mapsto x' =: x - o$. Questa applicazione è evidentemente biiettiva, e come ogni biiezione induce la struttura di uno spazio sull'altro, cioè in questo caso la struttura di \mathbf{R}^n sullo spazio delle x' , che denoteremo ${}_o\mathbf{R}^n$. Quest'ultimo risulta così uno spazio lineare n -dim con prodotto interno, “con origine o ” (infatti $x' = 0 \Leftrightarrow x = o$). Se da ${}_o\mathbf{R}^n$ passiamo all'associato spazio vettoriale \mathbf{R}^n con le modalità in precedenza descritte, otteniamo lo *stesso* spazio vettoriale associato a \mathbf{R}^n , indipendentemente dalla scelta di o , perché banalmente $y' - x' = (y - o) - (x - o) = y - x$, ovvero $\langle x', y' \rangle \approx \langle x, y \rangle$. La biiezione di \mathbf{R}^n su ${}_o\mathbf{R}^n$ si dice **o -traslazione di \mathbf{R}^n** . È chiaro che \mathbf{R}^n e ${}_o\mathbf{R}^n$ sono distinti come spazi dotati di struttura, pur essendo identici come insiemi (senza struttura). Per ogni $o \in \mathbf{R}^n$, ${}_o\mathbf{R}^n$ si dice **spazio affine associato a \mathbf{R}^n , di origine o** . Il fatto che \mathbf{R}^n e ${}_o\mathbf{R}^n$ sono lo stesso insieme di n -ple ordinate può oscurare l'argomento testé illustrato, ma confidiamo che quanto segue chiarisca del tutto la situazione.

La nozione algebrica di **spazio affine associato a uno spazio lineare** è completamente generale, e per la sua importanza in geometria vale la pena di darne la definizione. Sia E uno spazio lineare, ed \mathcal{A} un insieme qualsiasi. Sia poi $+ \bullet: E \times \mathcal{A} \rightarrow \mathcal{A}$ un'applicazione che scriveremo tra l'elemento di E , a sinistra, e l'elemento di \mathcal{A} , a destra, soggetta alle seguenti proprietà assiomatiche.

$$(\alpha) \quad \forall (t, s \in E, x \in \mathcal{A}) \{ (t+s) + \bullet x = t + \bullet (s + \bullet x) \}$$

$$(\beta) \quad \forall (x \in \mathcal{A}) \{ t + \bullet x = x \} \Leftrightarrow t = 0$$

$$(\gamma) \quad \forall (x, y \in \mathcal{A}) \exists (t \in E) \{ y = t + \bullet x \}.$$

Convorrà scrivere (β_1) per la parte \Leftarrow della (β) e (β_2) per la parte \Rightarrow . In forza della (α) e della (β_1) , per $t \in E$ fisso l'applicazione $x \mapsto t + \bullet x$ di \mathcal{A} in se stesso è una biiezione (di \mathcal{A} su \mathcal{A}), o sostituzione di \mathcal{A} , della quale la $x \mapsto -t + \bullet x$ è l'inversa. Questa sostituzione di \mathcal{A} si dice

t-traslazione di \mathcal{A} , e sarà denotata \mathcal{T}_t . L'applicazione $t \mapsto \mathcal{T}_t$ è un omomorfismo del gruppo additivo di E nel gruppo delle sostituzioni di \mathcal{A} (cioè $\mathcal{T}_{t+s} = \mathcal{T}_t \circ \mathcal{T}_s = \mathcal{T}_s \circ \mathcal{T}_t$, ove \circ denota la composizione gruppale), che è iniettiva in forza della (β_2) , e permette di identificare E additivo con il **gruppo delle traslazioni** di \mathcal{A} , evidentemente abeliano. Per $x \in \mathcal{A}$ fisso, l'applicazione $t \mapsto t + \bullet x$ è suriettiva in forza della (γ) ; ma essa è anche *iniettiva*⁷, e quindi è una *biiezione* di E su \mathcal{A} . Fissiamo ora un arbitrario elemento o di \mathcal{A} . Per quanto abbiamo appena provato, $\forall x \in \mathcal{A}$ esiste un *unico* elemento di E che sommato a sinistra $(+ \bullet)$ a o dà x . Denoteremo con ox questo elemento di E ⁸ (in particolare, risulta $oo = 0$ ⁹). La biiezione $ox \mapsto x$ di E su \mathcal{A} induce su \mathcal{A} , nel modo standard, la struttura di E , che si dice “ottenuta prendendo o come zero di \mathcal{A} ”. Dotato di questa struttura, \mathcal{A} si dice lo **spazio affine associato a E con origine o** , e si denota ${}_o\mathcal{A}$. Scrivendo (\leftrightarrow) per la biiezione in oggetto, e usando in ${}_o\mathcal{A}$ gli stessi simboli della struttura di E , si ha cioè, $\forall (x, y \in {}_o\mathcal{A})$, $x + y \leftrightarrow \leftrightarrow ox + oy$, e $\forall (\alpha \in \mathbf{R})$, $\alpha x \leftrightarrow \alpha(ox)$ (queste definizioni soddisfano banalmente gli assiomi lineari). Se poi E è dotato di prodotto interno, questo è trasferito in ${}_o\mathcal{A}$ in modo analogo, mediante la $x \cdot y = ox \cdot oy$ (ancora, questa definizione soddisfa banalmente gli assiomi di prodotto interno). Infine, se dallo spazio lineare affine ${}_o\mathcal{A}$ associato ad E si risale all'associato spazio vettoriale secondo le già definite modalità, quest'ultimo ed E coincidono a meno di un isomorfismo, indipendentemente dalla scelta di o . È chiaro che ${}_o\mathcal{A} \neq {}_{o'}\mathcal{A}$ se $o' \neq o$, anche se ${}_o\mathcal{A}$ e ${}_{o'}\mathcal{A}$ hanno lo stesso insieme-supporto. Si verifica immediatamente, infine, che ${}_{o'}\mathcal{A}$ risulta dalla oo' -traslazione di ${}_o\mathcal{A}$.

Un caso particolare di grande importanza si verifica quando $\mathcal{A} = E$. Allora la $+ \bullet$ diventa la somma $+$ di E , gli assiomi $(\alpha + \gamma)$ sono banalmente soddisfatti, e si ottiene lo **spazio affine ${}_oE$ associato a E , di origine $o \in E$** . (${}_oE$, ove il pedice o è lo zero di E , coincide quindi con E a tutti gli effetti.) Il caso dianzi considerato dello spazio affine associato a \mathbf{R}^n , di origine o , è ancora più

⁷ Infatti, sia t tale che $t + \bullet x = x$; vogliamo provare che $t = 0$. A questo scopo, sia y un qualunque elemento di \mathcal{A} ; per la (γ) , esiste un $u \in E$ per cui $y = u + \bullet x$, ossia per cui $t + \bullet y = t + \bullet (u + \bullet x) = (t+u) + \bullet x$ (per la (α)) $= u + \bullet x = y$. Per la supposta arbitrarietà di y , e in forza della (β_1) , $t = 0$, qed.

⁸ In luogo di ox , si può usare la notazione più suggestiva $x - \bullet o$.

⁹ Questa eguaglianza (o piuttosto la $xx = 0 \forall x \in \mathcal{A}$) si generalizza facilmente, facendo uso dell'assioma (α) , nella seguente **formula di Chasles** (Michel, 1793-1880), valida per ogni $m \geq 2$: $(*) x_1x_2 + x_2x_3 + \dots + x_mx_1 = 0$ (zero di E), dove x_1, x_2, \dots, x_m sono punti qualsiasi di \mathcal{A} . In particolare la $(*)$ dà $m(xx) = 0$, e quindi ancora $xx = 0$, facendovi $x_1 = x_2 = \dots = x_m = x$. La nozione di spazio affine ${}_o\mathcal{A}$ associato a E si può adesso introdurre anche partendo dalla formula di Chasles con $m = 3$ e usando un conveniente “assioma di biiezione”. Precisamente, si consideri un'applicazione di $\mathcal{A} \times \mathcal{A}$ in E , della quale denotiamo con xy l'immagine (in E) della coppia ordinata $\langle x, y \rangle \in \mathcal{A} \times \mathcal{A}$, soggetta alle seguenti proprietà assiomatiche: (i) “ $xy + yz + zx = 0 \forall (x, y, z \in \mathcal{A})$ ”, e (ii) “fissato un o di \mathcal{A} , l'applicazione di \mathcal{A} in E descritta dalla $x (\in \mathcal{A}) \mapsto ox \equiv t (\in E)$ è una biiezione.” Facendo $x = y = z$ nella (i), si ha $xx = 0$ ($\forall x \in \mathcal{A}$), e facendovi $z = x$, si ha $xy + yx = 0$ ($\forall x, y \in \mathcal{A}$). La (ii) ci assicura poi che per ogni $t \in E$ esiste un *unico* x di \mathcal{A} per cui $ox = t$. Posto quindi $x =: {}_oA(t)$, la biiezione ${}_oA$ trasferisce su \mathcal{A} l'algebra di E secondo le ovvie $x + y = {}_oA(t+s)$ (con $oy=s$) e $\alpha x = {}_oA(\alpha t)$. Alla luce di queste, sia ${}_oA$ che ${}_oA^{-1}$ sono banalmente lineari. In conclusione ${}_oA^{-1}(-x) = -ox = xo$, e ${}_oA^{-1}(y-x) = oy + xo = xy$. Per il prefissato o , possiamo ormai denotare con ${}_o\mathcal{A}$ lo spazio lineare ${}_oAE$, che è lo spazio affine associato a E con l'algebra di origine o .

particolare, e consiste nel fare $\mathcal{A} = E = \mathbf{R}^n$, traslando poi \mathbf{R}^n di o . Il risultante spazio ${}_o\mathbf{R}^n$ si dice **o-traslato di \mathbf{R}^n** .

Torniamo finalmente, dopo questo lungo preambolo, allo spazio euclideo H_n . Come sappiamo, fissato in esso un sistema di riferimento di assi cartesiani ortogonali, H_n è in corrispondenza biunivoca con lo spazio \mathbf{R}^n delle corrispondenti coordinate; e usando su quegli assi la misura $\mu = d$, sotto questa corrispondenza la distanza d tra due punti in H_n uguaglia la distanza pitagorica, che per convenienza tipografica scriveremo anche $*d$, tra le n -ple ordinate di coordinate. Una rototraslazione del sistema di riferimento corrisponde ad una trasformazione di tipo (9) delle coordinate, e lo spazio vettoriale associato \mathbf{R}^n non è influenzato dalla componente traslatoria della trasformazione. La tesi centrale che si dimostrerà nel seguito di questa sottosezione è che lo spazio \mathbf{R}^n delle coordinate cartesiane ortogonali di H_n con l'associata distanza pitagorica $*d$, e per il dato riferimento (in cui si sia fatto $\mu = d$), è un **modello normale**¹⁰ (dove la parola “modello” ha qui il senso che ad esso si dà in logica, e quindi andrebbe scritta “Modello”, giusta la convenzione stabilita in S.sez. 0.1.1) di H_n con la distanza d . Con ciò si intende che la corrispondenza biunivoca tra H_n e \mathbf{R}^n è tale che ogni enunciato P_H in H_n (una relazione tra distanze d tra sue coppie di punti) e il corrispondente enunciato P_R in \mathbf{R}^n (una relazione tra le distanze $*d$ tra le coppie di n -ple ordinate corrispondenti di \mathbf{R}^n) hanno lo stesso valore di verità (ovvero $P_R \Leftrightarrow P_H$). Risulta anche che ogni altro modello (della stessa cardinalità) di H_n è isomorfo a \mathbf{R}^n , ossia che la teoria formalizzata “ H_n ” è **categorica**.

Per provare la prima tesi, occorrerà anzitutto (i) prendere cognizione della interpretazione in \mathbf{R}^n dei termini e delle relazioni derivate che occorrono negli assiomi di H_n , e (ii) verificare poi che ogni assioma di H_n , nella interpretazione (o modello) “ \mathbf{R}^n ”, è un enunciato vero. Specificamente, a noi interessano qui i valori di n compresi tra 1 e 3; ma se disponessimo (come *non* disponiamo, avendovi soltanto accennato) della formalizzazione di $H_{n>3}$, non sarebbe difficile estendere, su analoga base, la stessa tesi a qualunque $n > 3$. Esamineremo dunque con questo programma gli assiomi di $H_{1 \leq n \leq 3}$ uno ad uno nel loro ordine di presentazione; si tratterà cioè di dimostrare come veri (nel modello “ \mathbf{R}^n ”) diciannove asserti della geometria analitica classica, o euclidea. Diremo al solito $X, Y, Z \dots$ i punti di H_n e x, y, z, \dots le corrispondenti n -ple ordinate di \mathbf{R}^n *che nel dato riferimento* ne sono le coordinate cartesiane ortogonali.

I tre primi assiomi ($\mathbf{A}_0 \div \mathbf{A}_2$, v. S.sez. 1.2.2)¹¹ non meritano commenti: la loro traduzione e il loro valore di verità “vero” nel modello \mathbf{R}^n sono del tutto evidenti. Passiamo dunque all'assioma

¹⁰ Ricordiamo che un modello normale di una teoria del 1° ordine con uguaglianza è un modello in cui l'uguaglianza nella teoria corrisponde all'uguaglianza nel modello.

¹¹ Per maggiore comodità di consultazione, evidenziamo qui col grassetto i singoli assiomi di cui verifichiamo la verità nel modello \mathbf{R}^n quando sono nominati per la prima volta.

(A'₃). Nel modello \mathbf{R}^n , esso si traduce nella forma:

$$\forall(x,y)\{x \neq y \Rightarrow \exists(z)\{ *d(x,y) = *d(x,z) + *d(z,y) \wedge z \neq x \wedge z \neq y \}\},$$

e risulta vero perché basta scegliere $z = (1-t)x + ty$ (cioè $z_i = (1-t)x_i + ty_i$ per $i = 1, \dots, n$), con $t \in (0,1)$. Infatti in tal caso $z \neq x$ e $z \neq y$, e inoltre, tenendo presente che $*d(u,v) = |u-v|$, $|z-x| + |z-y| = |(1-t)x+ty-x| + |(1-t)x+ty-y| = t|y-x| + (1-t)|y-x| = |y-x|$. Passando a

(A''₃), il suo conseguente si traduce in \mathbf{R}^n nella forma:

$\exists(z)\{ *d(z,y) = *d(z,x) + *d(x,y) \wedge z \neq x \wedge z \neq y \}$, che è vero ponendo $z = (1-t)^{-1}(x-ty)$, ancora con $t \in (0,1)$. Infatti $|z-y| = (1-t)^{-1}|x-y|$, e $|z-x| = t(1-t)^{-1}|x-y|$, da cui si hanno immediatamente la seconda e la terza tesi sotto l'ipotesi $x \neq y$; e inoltre, $|z-x| + |x-y| = t(1-t)^{-1}|x-y| + |x-y| = (1-t)^{-1}|x-y| = |z-y|$. Per discutere

(A₄), occorre prima sapere in cosa si traduce il termine derivato $\mathcal{R}(X,Y)$ per dati X e Y distinti (retta passante per X,Y). La risposta è che, in \mathbf{R}^n , $\mathcal{R}(X,Y)$ si traduce nell'insieme delle n -ple ordinate

$$(10) \quad z = ax + by \text{ (con } x \neq y \text{ e } a, b \text{ reali qualsiasi sotto il vincolo } a + b = 1).$$

La (10) è l'equazione parametrica della retta passante per x e y (distinti) nel modello \mathbf{R}^n , ed esprime la collinearità di x , y e z se $a \neq 0$ e $b \neq 0$. Questo fatto si verifica senza difficoltà alla luce della definizione di $\mathcal{R}(X,Y)$ (v. S.sez. 1.2.2) e della (10) stessa. L'assioma (A₄) è allora banalmente vero, perché la (10) definisce la sopraddetta retta di \mathbf{R}^n in modo unico ("per x e y distinti passa esattamente una retta"). Ci occupiamo ora dell'assioma

(A'₅), l'ipotesi del quale, nel modello \mathbf{R}^n , diventa: " $y = (1-t)x + tz$ per qualche $t \in (0,1)$ e $z = (1-s)y + su$ per qualche $s \in (0,1)$, sotto $x \neq z$ e $y \neq u$ ", mentre la tesi diventa " $y = (1-r)x + ru$ per qualche $r \in (0,1)$ e $x \neq u$ ". Eliminando z nell'ipotesi, troviamo:

$$(11) \quad y = (1-t(1-s))^{-1}((1-t)x + tsu),$$

ove evidentemente $1 - t(1-s) \neq 0$ per $t \in (0,1)$ e $s \in (0,1)$, e la somma dei coefficienti di x e di u a 2° membro vale 1. Infine la funzione $ts(1-t(1-s))$ delle due variabili t e s è in $(0,1)$ per t e s in $(0,1)$ (la semplice verifica è lasciata al lettore). Si può dunque porre r uguale a questa funzione, e la tesi è dimostrata. Simile è la discussione dell'assioma

(A''₅). Nel modello \mathbf{R}^n , l'ipotesi diventa infatti " $y = (1-t)x + tu$ per qualche $t \in (0,1)$ e $z = (1-s)y + su$, sotto $x \neq u$ e $y \neq u$ ", mentre la tesi diventa " $y = (1-r)x + rz$ per qualche $r \in (0,1)$, e $x \neq z$ ". Eliminando u dall'ipotesi si ottiene

$$(12) \quad y = (1+(1-s)t/s)^{-1}((1-t)x + tz/s),$$

ove $1+(1-s)t/s \neq 0$ se t e s sono in $(0,1)$, e la somma dei coefficienti di x e z a 2° membro vale 1. La funzione $t(s+t(1-s))^{-1}$ di t e s è in $(0,1)$ se t e s lo sono (la verifica è lasciata al lettore). La tesi è

dimostrata ponendo r uguale a questa funzione. Infine la discussione nella stessa chiave dell'ultimo assioma

(A₆) della S.sez. 1.2.2 è quasi immediata: l'assegnazione della “distanza con segno” di y da x su una “semiretta” data con origine in x , determina y unicamente. Sia infatti $u = (1-t)x + ty_0$, sotto $x \neq y_0$, e con $t > 0$, l'insieme delle n -ple ordinate che costituisce la semiretta in questione nel modello \mathbf{R}^n . La facile verifica di questo fatto, alla luce della definizione di $\mathcal{SR}(X|Y_0)$, è lasciata al lettore. Allora $|u-x| = t|y_0-x|$, e la prescrizione di $|u-x| > 0$ determina unicamente $t_u > 0$ al quale corrisponde l' y richiesto, diciamo $y_u = (1-t_u)x + t_u y_0$. Se $t_u = 1$, $y_u = y_0$. Il primo assioma

(A₇) della S.sez 1.2.3 afferma che “se $n \geq 2$ esistono tre punti non collineari in H_n ”. Basta limitarsi a $n = 2$ per dimostrare la tesi in \mathbf{R}^n . Senza limitare la generalità, possiamo infatti assumere $x = \langle 0,0 \rangle$, $y = \langle 1,0 \rangle$ e $z = \langle 0,1 \rangle$, quindi $ax + by = \langle b,0 \rangle$; la richiesta di “collinearità” di x , y e z richiederebbe $\langle 0,1 \rangle = \langle b,0 \rangle$ per qualche b , che non può essere soddisfatta. In definitiva le tre coppie ordinate x , y , e z del “piano” \mathbf{R}^2 non sono collineari (ai sensi della definizione di “retta” in \mathbf{R}^2 , cfr. eq. (10)). La dimostrazione vale anche per $n > 2$ nel senso che ci si può sempre ridurre nel piano $x_3 = \dots = x_n = 0$. Per discutere gli assiomi

(A₈) e successivi, dobbiamo avere il significato in \mathbf{R}^n del termine derivato $\mathcal{P}(X,Y,Z)$ (S.sez. 1.2.3), cioè del “piano passante per X , Y , Z distinti e non collineari”. Con facile generalizzazione dell'analogo caso della retta per X , Y distinti, $\mathcal{P}(X,Y,Z)$ si traduce nell'insieme delle n -ple ordinate del tipo (*) $u = ax + by + cz$, ove a, b, c sono reali soggetti al vincolo $a + b + c = 1$, sotto la $\neq(x,y,z)$ e sotto la condizione che x, y, z non siano collineari, cioè che *non* esistano reali a' e b' sotto $a' + b' = 1$ per cui $z = a'x + b'y$ ¹². Sotto le condizioni menzionate, la (*) si dice “equazione parametrica del piano per x, y, z (distinti e non collineari)”, nel modello \mathbf{R}^n . L'assioma (A₈) è quindi valido nel modello, perché tale equazione parametrica definisce l'insieme delle u in modo unico (“per x, y e z distinti e non collineari passa esattamente un piano”). Arriviamo così all'assioma

(A₉). Siano \underline{X} e \underline{Y} i punti distinti comuni alla retta p e al piano π . Come sappiamo, nel modello \mathbf{R}^n la retta p diventa l'insieme delle n -ple ordinate $u = (1-t)\underline{x} + t\underline{y}$ per $t \in (0,1)$; e il piano π , che per ipotesi contiene \underline{X} e \underline{Y} oltre ad un terzo punto \underline{Z} distinto dai primi e con essi non collineare, diventa l'insieme delle n -ple ordinate $w = a\underline{x} + b\underline{y} + c\underline{z}$ con a, b, c reali sotto $a + b + c = 1$. È chiaro che la “retta” delle u giace per intero nel “piano” delle w ; basta prendere $a = 1-t$, $b = t$, e $c = 0$ affinché w sia uguale a $(1-t)\underline{x} + t\underline{y}$ e quindi sia un' n -pla ordinata di tipo u . È la volta dell'assioma

¹² In caso contrario avremmo $u = (a+va')x + (b+vb')y$, con la somma dei coefficienti di x e di y uguale a 1 in forza delle $a' + b' = 1$ e $a + b + c = 1$; e quindi l'insieme delle u sarebbe su una retta nel modello \mathbf{R}^n .

(A₁₀). Ancora senza limitare la generalità, possiamo qui fare $n = 2$ e supporre che la retta p si rappresenti in \mathbf{R}^2 come l'insieme delle coppie ordinate $u = \langle u_1, u_2 \rangle$ per cui $u_2 = 0$.¹³ I tre punti distinti X, Y e Z si rappresentano quindi in \mathbf{R}^2 come le coppie ordinate $x = \langle x_1, x_2 \rangle$, $y = \langle y_1, y_2 \rangle$ e $z = \langle z_1, z_2 \rangle$, con $x_2 \neq 0$, $y_2 \neq 0$ e $z_2 \neq 0$, e per ipotesi $x_2 y_2 < 0$. Sia $x_2 > 0$ e $y_2 < 0$; quindi, se $z_2 > 0$, risulta $y_2 z_2 < 0$, e se $z_2 < 0$, risulta $y_2 z_2 > 0$. La prima possibilità ($z_2 > 0$) verifica l'assioma, perché la "retta" $x_2 = 0$ interseca il "segmento" di "estremi" y, z (ma non quello di estremi x, z). Se invece $x_2 < 0$ e $y_2 > 0$, basta scambiare i ruoli di x e y ; allora è la seconda possibilità ($z_2 < 0$) a verificare l'assioma, perché la stessa retta interseca il segmento di estremi x, z (ma non quello di estremi y, z). Ci occupiamo ora dell'assioma

(A₁₁). Anche in questo caso possiamo limitarci a $n = 2$. Inoltre senza limitare la generalità possiamo prendere $x = \langle 0, 0 \rangle$ per la coppia ordinata corrispondente al punto X dell'assioma e $y = \langle 1, 0 \rangle$ per quella corrispondente al punto Y .¹⁴ In forza della disuguaglianza triangolare in \mathbf{R}^2 , posto $m =: |z-x|$ e $n =: |z-y|$ (dove la coppia ordinata z corrisponde al punto Z dell'assioma), risulta

$$(13) \quad m + n > 1 \quad \text{e} \quad |m - n| < 1.$$

Per il teorema di Pitagora in \mathbf{R}^2 , è allora $z_1^2 + z_2^2 = m^2$ e $(1-z_1)^2 + z_2^2 = n^2$, ossia

$$(14) \quad z_2^2 = m^2 - (1+m^2-n^2)^2/4,$$

ove il segno di z_2 è prefissato, ad es. è positivo. Sotto le condizioni (13), il 2° membro della (14) risulta strettamente positivo (la verifica è lasciata al lettore). Si conclude così che esiste esattamente una coppia ordinata z come richiesto dall'assioma, della quale $z_1 = (1+m^2-n^2)/2$, e z_2 è la radice positiva del 2° membro della (14). Connessa a quanto appena dimostrato è l'esame dell'assioma (A₁₂), a proposito del quale sarà ancora sufficiente limitarci al caso $n = 2$. In sostanza (A₁₁) afferma che, se $[X, Y, Z]$ e V non è sulla retta $\mathcal{R}(X, Z)$, essendo assegnate le distanze (necessariamente positive) $d(X, V)$ e $d(Y, V)$, è unicamente determinata la distanza $d(Z, V)$. Dicendo al solito x, y, z e v le coppie ordinate corrispondenti a X, Y, Z e V , e supponendo (senza limitare la generalità) $x_2 = y_2 = z_2 = 0$, abbiamo per ipotesi $|x_1 - z_1| = |x_1 - y_1| + |y_1 - z_1|$. Alla luce della precedente verifica di validità dell'assioma (A₁₁), la coppia ordinata $v = \langle v_1, v_2 \rangle$ è determinata a meno del segno di v_2 . Poniamo allora $v' =: \langle v_1, |v_2| \rangle$, $v'' =: \langle v_1, -|v_2| \rangle$ e $\alpha =: |v_1 - z_1|$. Abbiamo $|v' - z|^2 = |v_2|^2 + \alpha^2 = |v'' - z|^2$, qed. Della S.sez. 1.2.3, restano da esaminare l'assioma

(A₁₃), il che è ormai banale, potendosi procedere, con generalizzazione ovvia, come per (A₇); e l'assioma

¹³ A questa situazione ci si può sempre ridurre con opportuna rototraslazione (9), rispetto alla quale è indifferente quanto asserito da (A₁₀).

¹⁴ Vale anche qui il commento a piè pagina alla analoga situazione a proposito di (A₁₀); ed è da aggiungere che alla scelta $y_1 = 1$ ci si può sempre ridurre con un cambiamento di unità, rispetto al quale l'asserto (A₁₁) è indifferente. Del resto, l'intera dimostrazione può ripetersi senza difficoltà sostituendo un valore qualsiasi al valore 1 di y_1 .

(A₁₄). Dalla geometria analitica “cartesiana ortogonale”, ricordiamo che una “retta” di $\mathbf{R}^{n \geq 2}$ è l’insieme delle sue n-ple ordinate $x \equiv \langle x_1, \dots, x_n \rangle$ per le quali è soddisfatto il sistema di n-1 equazioni

$$(15) \quad \sum_j C_{ij}x_j + C_i = 0 \quad (i = 1, 2, \dots, n-1),$$

dove $\{C_{ij}\}$ è una ((n-1)×n)-matrice e $\{C_i\}$ è una (n-1)-colonna, entrambe definite a meno di uno stesso fattore $\neq 0$, sotto la condizione che $\{C_{ij}\}$ abbia rango ¹⁵ n - 1 (cioè, che le (15) siano indipendenti). Nel caso n = 2, si è ridotti ad un’unica equazione del tipo

$$(15') \quad C + C_1x_1 + C_2x_2 = 0,$$

sotto la condizione che C₁ e C₂ non siano entrambi nulli. È facile accertare che la (15') rappresenta una “retta” nel senso della precedente definizione, cioè di “insieme delle u del tipo (1-t) \underline{x} + t \underline{y} per qualche \underline{x} e qualche \underline{y} (tra loro distinte) di \mathbf{R}^2 ”. ¹⁶ In modo del tutto analogo, un “piano” di $\mathbf{R}^{n > 2}$ è l’insieme delle sue n-ple ordinate x per cui è soddisfatto il sistema di n-2 equazioni (15) con i = 1, .., n - 2, sotto la condizione che la relativa ((n-2)×n)-matrice abbia rango n - 2. ¹⁷ Se n = 3, siamo ridotti ad un’unica equazione del tipo

$$(15'') \quad C + C_1x_1 + C_2x_2 + C_3x_3 = 0,$$

con C₁, C₂, C₃ non simultaneamente nulli. Come nel precedente caso della retta, è facile provare che la definizione di piano di \mathbf{R}^3 come “insieme delle terne ordinate u del tipo a \underline{x} + b \underline{y} + c \underline{z} per certi \underline{x} , \underline{y} , \underline{z} distinti e non collineari, sotto a + b + c = 1”, e quella data dalla (15'') coincidono. A questo punto la verifica di (A₁₄) è quasi immediata. Senza limitare la generalità, possiamo far corrispondere al primo dei due piani dell’assioma, diciamo π, quello di equazione x₃ = 0, e al secondo, diciamo ρ, quello di equazione (15'') (sotto la solita condizione della non nullità simultanea di C₁, C₂, C₃). In questo caso si deve tuttavia escludere la possibilità che sia C₁ = C₂ = 0; se così fosse, infatti, l’equazione di ρ si ridurrebbe a C + C₃x₃ = 0 (con C₃ ≠ 0), e quindi ρ non

¹⁵ Il rango di una matrice si dice o si diceva anche spesso, equivalentemente, sua “caratteristica”.

¹⁶ Sostituendo infatti questa definizione nella (14'), si ottiene “Esistono \underline{x} e \underline{y} distinti, per cui $C + C_1[(1-t)\underline{x}_1 + t\underline{y}_1] + C_2[(1-t)\underline{x}_2 + t\underline{y}_2] = 0$ identicamente rispetto a t, quindi per cui $C + C_1\underline{x}_1 + C_2\underline{x}_2 = 0$ e $C_1(\underline{y}_1 - \underline{x}_1) + C_2(\underline{y}_2 - \underline{x}_2) = 0$. Poiché almeno una delle due differenze $(\underline{y}_1 - \underline{x}_1)$, $(\underline{y}_2 - \underline{x}_2)$ è $\neq 0$, la seconda di queste definisce C₁ e C₂ (non simultaneamente nulli) a meno di un fattore, e la prima definisce C a meno dello stesso fattore. Viceversa, per dati C, C₁, C₂ (con C₁ e C₂ non simultaneamente nulli), la $C + C_1x_1 + C_2x_2 = 0$ ha la soluzione $\underline{x}_1 = 0$, $\underline{x}_2 = -C/C_2$ (se C₂ ≠ 0), e la soluzione $\underline{y}_2 = 0$ e $\underline{y}_1 = -C/C_1$ (se C₁ ≠ 0). L’insieme $u = (1-t)\underline{x} + t\underline{y}$ soddisfa alla $C + C_1u_1 + C_2u_2 = 0$. Se poi ad es. C₁ = 0 (A₂ ≠ 0), allora u soddisfa alla $C + C_2u_2 = 0$ con u₁ arbitraria e u₂ = -C/C₂. Questo caso degenerare corrisponde a $\underline{x} = \langle 0, -C/C_2 \rangle$ e (ad es.) $\underline{y} = \langle 1, -C/C_2 \rangle$. In conclusione le due definizioni di retta si equivalgono.

¹⁷ Più in generale, per n > 3, un “m-piano” (con 2 < m < n) di \mathbf{R}^n è l’insieme delle sue n-ple ordinate per cui è soddisfatto il sistema (14) di n - m equazioni per i = 1, .., n - m, supposte indipendenti. Così definiti, gli “m-piani” (rette per m = 1, piani per m = 2, m-piani per m > 2) si dicono anche “varietà affini m-dim” di \mathbf{R}^n ; ed è facile capire che m è in tal caso il numero dei parametri liberi delle loro equazioni parametriche, come mostrano gli esempi già visti per m = 1 e m = 2. A partire da ciò, una **dimensione lineare** m, con 0 ≤ m < n, può attribuirsi a qualunque sottoinsieme A di \mathbf{R}^n , definendola come la dimensione della più piccola varietà affine di \mathbf{R}^n includente A. Se poi una tale varietà affine non esiste, allora la dimensione lineare di A è per definizione n. La nozione generale di “dimensione” di una “varietà immersa in \mathbf{R}^n ” (sorvoliamo qui sulla definizione precisa di questo concetto) è data soltanto nell’ambito della “teoria delle varietà”, in particolare con le relative nozioni di “carta” e di “atlante”, vedi Cap 4.

potrebbe avere un punto in comune con π ed esserne distinto. In definitiva l'equazione di ρ è la (15'') con C_1 e C_2 non simultaneamente nulli. Combinando questa con l'equazione di π , si ha $C + C_1x_1 + C_2x_2 = 0$, che è l'equazione di una intera retta, corrispondente all'intersezione di π e ρ . L'esame di (A_{14}) è con ciò terminato.

Dobbiamo infine verificare la validità dei due assiomi (C) e (E) nel modello \mathbf{R}^n ; ma quella di (C) è ovvia alla luce della definizione standard di "retta reale". Concludiamo quindi con la verifica (della validità) dell'assioma

(E). Evidentemente basta limitarsi al caso $n = 2$. Proviamo innanzitutto che se x, y, z sono distinte e "non collineari" allora una qualsiasi "retta" di \mathbf{R}^2 , che nomineremo come p , interseca quella passante per x e y e/o quella passante per x e z . Ancora senza limitare la generalità, possiamo porre $x =: \langle 0,0 \rangle$ e $y =: \langle y_1,0 \rangle$; allora la retta passante per x e y ha equazione $u_2 = 0$, mentre quella passante per x, z ha equazione del tipo $(\dagger) C_1u_1 - C_2u_2 = 0$, con $C_1 \neq 0$. Se la retta p non interseca la retta per x e y , essa ha equazione $u_2 = \text{cost} = B \neq 0$ ¹⁸; sostituendo questa nella (\dagger) , si ha $C_1u_1 - A_2B = 0$. Quindi p e la retta per x e z si intersecano in $B\langle z_1/z_2, 1 \rangle$ (z_2 è certamente $\neq 0$, perché $C_2z_2 = C_1z_1$). Supponiamo ora che due rette distinte q e r siano entrambe disgiunte da una terza retta p passando per un x fuori di p . Prendendo arbitrarie $y \in (q \setminus \{x\})$ e $z \in (r \setminus \{x\})$, x, y e z risultano distinte e non collineari, e quindi almeno una delle q e r dovrebbe intersecare p , contrariamente all'ipotesi che ne siano entrambe disgiunte. In conclusione per $x \notin p$ passa *al più* una retta disgiunta da p . Contando gli assiomi esaminati, troviamo appunto diciannove, come anticipato.

Abbiamo così concluso la verifica del fatto che \mathbf{R}^n è un modello normale di H_n se l' n -pla ordinata delle coordinate del generico punto di H_n , rispetto ad un arbitrario sistema di riferimento cartesiano ortogonale in cui la distanza d sia usata come misura di segmenti liberi, è il corrispondente elemento di \mathbf{R}^n .

¹⁸ In generale, le due rette di equazioni $C_i + \sum_j C_{ij}x_j = 0$ ($i, j = 1, 2$) (sotto le solite condizioni " $C_{i1} \neq 0 \vee C_{i2} \neq 0$ " ($i = 1, 2$)), si intersecano sse la (2×2) -matrice $C =: \{C_{ij}\}$ e la (2×3) -matrice $\{C_{ij}, C_i\}$ hanno lo stesso rango (teorema di Capelli (Alfredo, 1855-1910)). Ora il rango di C è per definizione 1 aut 2. Se è 2, le due rette si intersecano perché il rango di $\{C_{ij}, C_i\}$ è comunque 2. Se il rango di C è invece 1, cioè se $\det C = 0$, la richiesta di intersezione risulta essere " $(C_{11}C_2 = C_{21}C_1) \wedge (C_{12}C_2 = C_{22}C_1)$ ". Volendo che le rette siano disgiunte, dobbiamo richiedere (i) $\det C = 0$ e (ii) " $(C_{11}C_2 \neq C_{21}C_1) \vee (C_{12}C_2 \neq C_{22}C_1)$ ". Di fatto, la (i) equivale al parallelismo delle due rette, e $(i) \wedge \neg(ii)$ equivale alla loro identità; quindi $(i) \wedge (ii)$ equivale alla loro disgiunzione. La (ii) si interpreta facilmente se l'origine del sistema di riferimento è posta su una delle due rette, ad es. sulla prima di esse ($i=1$); allora $C_1=0$, e la (ii) diventa " $(C_{11}C_2 \neq 0) \vee (C_{12}C_2 \neq 0)$ ", cioè $C_2 \neq 0$, perché " $(C_{11} \neq 0) \vee (C_{12} \neq 0)$ ". In conclusione due rette di equazioni $C_i + \sum_j C_{ij}x_j = 0$, di cui una passante per l'origine e l'altra no (cioè $(C_1=0) \wedge (C_2 \neq 0)$), sono disgiunte sse $\det A = 0$.

1.3.2) \mathbf{R}^n COME MODELLO NORMALE DI H_n : PARTE SECONDA

Ciò che rende di grande interesse la corrispondenza biunivoca (\leftrightarrow) tra H_n e \mathbf{R}^n , fondata sulla scelta di un riferimento cartesiano ortogonale in H_n , non si esaurisce nel riconoscimento appena accertato di \mathbf{R}^n come modello normale di H_n . Sappiamo infatti che \mathbf{R}^n possiede una struttura “naturale” ben più ricca di quella di H_n (almeno come l’abbiamo fin qui descritta), la quale ultima è soltanto quella di uno spazio metrico con certe relazioni d’ordine. Ovviamente questa struttura di \mathbf{R}^n può essere trasmessa secondo le modalità standard su H_n mediante la corrispondenza (\leftrightarrow); ma non ci si può attendere che, così acquisita, essa struttura risulti di necessità altrettanto “naturalmente” significativa da un punto di vista intrinseco, *interno* alla teoria formalizzata di H_n (nella versione delle S.sez. 1.2.2 ÷ 1.4.4, o in qualunque versione equivalente). Ad esempio, ci si può chiedere se la nozione di “somma” dei punti X e Y di H_n , $X + Y$, introdotta con le convenienti modalità “autonome”, e quindi prescindendo dalla definizione standard espressa mediante l’implicazione $(X \leftrightarrow x \wedge Y \leftrightarrow y) \Rightarrow (X + Y \leftrightarrow x + y)$, ove ovviamente $x, y \in \mathbf{R}^n$, risulti di per sé abbastanza naturale. Il nostro prossimo programma è quello di esplorare in questa chiave l’intera struttura di \mathbf{R}^n così come è stata precedentemente descritta.

Cominciando con la predetta somma, il modello \mathbf{R}^n ci suggerisce la seguente definizione. Si fissi una volta per tutte un punto O di H_n ; per i dati X e Y di H_n , supponendo O, X, Y distinti e non collineari, si consideri la parallela per Y alla retta $\mathcal{R}(O, X)$ e la parallela per X alla $\mathcal{R}(O, Y)$. Nella teoria H_n , sappiamo che tali parallele si incontrano in un punto Z unicamente determinato. Se invece O, X, Y sono collineari sotto $X \neq O$ e $Y \neq O$, sulla loro retta esiste un *unico* punto Z tale che il segmento orientato XZ sia congruente ed equiverso al segmento orientato OY (o equivalentemente tale che YZ sia congruente ed equiverso a OX). Infine se $X = O$ [$Y = O$] si pone $Z = Y$ [$Z = X$]. Così Z è sempre definito, e in ogni caso si pone $X + Y =: Z$ (**regola generalizzata del parallelogramma**). È chiaro che questa “somma” dipende dalla scelta di O , e quindi andrebbe denotata con un simbolo ad hoc che ne recasse traccia; ma questo può essere qui sottinteso per brevità. Con questa precisazione, H_n risulta un gruppo abeliano sotto l’operazione di somma testé definita (la verifica che gli assiomi di gruppo additivo a) ÷ d) di S.sez. 1.3.1 sono soddisfatti è lasciata al lettore). Passiamo ora al “prodotto” di un reale α per $X \in H_n$ (da denotare con giustapposizione a sinistra di α a X). Per lo *stesso* punto O scelto nella definizione della somma, se $X \neq O$ esiste un *unico* punto Z sulla retta $\mathcal{R}(O, X)$ per il quale $d(O, Z) = |\alpha|d(O, X)$ e, se $\alpha \neq 0$, il segmento orientato OZ sia equiverso aut antiverso a OX a seconda che α sia positivo aut negativo. Se poi $X = O$, allora si pone $Z = O$. Così Z è sempre definito, e si pone $\alpha X =: Z$. Anche questa operazione prodotto dipende dalla scelta di O , e andrebbe denotata con un simbolo ad hoc. La

verifica che gli assiomi di prodotto e) \div h) di S.sez. 1.3.1 sono soddisfatti è ancora lasciata al lettore. In conclusione H_n diventa così uno spazio lineare; non solo, ma essendo anche topologico con la topologia indotta dalla sua metrica d , si potrebbe verificare che le operazioni di somma e prodotto appena introdotte risultano continue rispetto a questa topologia di H_n e a quella standard di \mathbf{R} ; e quindi, che H_n diventa con le date definizioni uno spazio lineare topologico. Questo spazio si rapporta nel modo più semplice al modello \mathbf{R}^n , considerato anch'esso come spazio lineare topologico (rispetto alla metrica $*d$ e alla topologia indotta, oltre alla topologia standard di \mathbf{R}): allo scopo, occorre e basta che $O \in H_n$ coincida con l'origine del sistema di riferimento cartesiano ortogonale con il quale si passa a \mathbf{R}^n . Così \mathbf{R}^n induce la sua struttura su H_n nel modo standard (come esemplificato nel paragrafo precedente per la somma), e H_n e \mathbf{R}^n risultano isomorfi come spazi lineari topologici.

Meno semplice è la definizione autonoma, in H_n , di un “prodotto interno” (\cdot) che soddisfi alla richiesta $X \cdot Y = x \cdot y$ sotto la corrispondenza canonica (\leftrightarrow) , cioè per cui $(X \leftrightarrow x \wedge Y \leftrightarrow y) \Rightarrow (X \cdot Y = x \cdot y)$. Fissato O nel solito ruolo, conviene limitare la nostra attenzione ad un prefissato piano $\rho \subset H_n$ cui appartengono O (fisso) e i due generici punti X e Y tra i quali vogliamo definire il prodotto interno. Quindi $\rho = \mathcal{P}(O, X, Y)$ se O, X e Y sono distinti e non collineari; ma ρ continua a essere definito anche nei casi degeneri in cui questa condizione non è soddisfatta. Riferendoci per cominciare alla situazione non-degenere, sia $p =: \mathcal{R}(O, X)$ [$q =: \mathcal{R}(O, Y)$], che converremo di orientare secondo $O \rightarrow X$ [$O \rightarrow Y$]. Sia X_q [Y_p] la proiezione ortogonale di X su q [di Y su p], e OX_q [OY_p] la coordinata di X_q [di Y_p] su q [su p] rispetto all'origine O (uguale a $\pm d(O, X_q)$ [a $\pm d(O, Y_p)$] a seconda che OX_q [OY_p] sia equiverso o antiverso a q [a p], e a zero se $O = X_q$ [$O = Y_p$]). Se invece O, X e Y sono collineari ma ancora distinti, e quindi p e q coincidono a meno dell'orientamento, distingueremo due casi: quello in cui $[O, X, Y]$ (oppure, equivalentemente, con lo scambio di X con Y , $[O, Y, X]$), e quello in cui $[X, O, Y]$. Le definizioni di OX_q e di OY_p sono ancora valide facendo $X_q = X$ e $Y_p = Y$, e risulta $OX_q > 0$ e $OY_p > 0$ nel primo caso e $OX_q < 0$ e $OY_p < 0$ nel secondo; e continuano ad esserlo, banalmente, se $X = Y \neq O$. Se poi $X = O$, si continuerà a dare alla retta p un orientamento convenzionale, in modo che OY_p continui ad avere senso come distanza con segno, o eventualmente come distanza nulla; e similmente se $Y = O$, si continuerà a dare alla retta q un orientamento convenzionale per dare senso a OX_q . In definitiva OX_q e OY_p sono numeri reali ben definiti comunque si scelgano i punti X e Y correnti nel piano ρ . Vale allora la seguente uguaglianza:

$$(1) \quad d(O,X)OY_p = d(O,Y)OX_q, \quad ^{19}$$

la quale riflette tra l'altro la simmetria di tutto l'argomento rispetto allo scambio di X con Y. In particolare se $O = X$, sia $d(O,X)$ che OX_q sono nulli indipendentemente dall'orientamento di q; e similmente se $Y = O$ lo sono $d(O,Y)$ e OY_p indipendentemente dall'orientamento di p.

Consideriamo adesso il primo membro della (1). Si può facilmente accertare che se in esso si sostituisce Y con $Y + Y'$ (secondo la definizione di $Y + Y'$ mediante la regola generalizzata del parallelogramma), il risultato è la somma dei valori che gli competono separatamente per Y e Y' . Similmente, si verifica che se in esso si sostituisce Y con αY (secondo la definizione di αY data nel secondo paragrafo di questa sezione), il risultato è uguale al prodotto di α per il valore che gli compete per Y. In altre parole il primo membro della (1) è lineare rispetto a Y. Per simmetria, il secondo membro della (1) è lineare rispetto a X; per cui, visto che i due oggetti considerati sono uguali, il numero reale nei due membri della (1) è lineare sia rispetto a X che a Y. Infine se $X = Y \neq O$, $d(O,X)OY_p = d(O,X)^2 > 0$, mentre se $X = Y = O$, $d(O,X)OY_p = 0$. In conclusione, abbiamo provato che il reale nei due membri della (1) soddisfa agli assiomi i) ÷ m) di S.sez. 1.3.1, ed è quindi naturale assumerlo come prodotto interno euclideo di X per Y, ossia $(^\circ) X \cdot Y =: d(O,X)OY_p = d(O,Y)OX_q$. Al solito, la definizione dipende dalla scelta di O, e quindi tale prodotto interno andrebbe denotato con un simbolo ad hoc. Va da sé che ponendo in O l'origine delle coordinate di H_n vale l'induzione canonica (da parte di \mathbf{R}^n) del prodotto interno euclideo definito da $X \cdot Y = x \cdot y$.

La disponibilità in H_n di un prodotto interno (sottintenderemo ancora l'attributo "euclideo" qui di seguito) pone una legittima questione, quella di esaminare qual'è la distanza d' tra X e Y definita attraverso di esso, cioè come modulo della differenza $Y - X$ (cioè $d'(X,Y) =: |Y - X|$). Usando la regola del parallelogramma, si dimostra facilmente che $|Y - X|$ non dipende da O, ed inoltre *che è uguale a* $d(X,Y)$, $d' \equiv d$. Vi è dunque consistenza tra H_n come spazio metrico (e quindi topologico) con la distanza d e come spazio metrico (e quindi topologico) con la distanza d' che gli deriva dall'essere dotato di prodotto interno secondo la definizione $(^\circ)$ del precedente paragrafo. In conclusione, fissato O, H_n è completamente definito in modo autonomo come spazio lineare con prodotto interno, consistentemente metrico e topologico.

Ricordiamo ora che in uno spazio topologico T con topologia indotta da una metrica d, quest'ultima è una funzione (≥ 0) *continua* rispetto ai suoi due argomenti e alla topologia standard di \mathbf{R} . Basta usare la disuguaglianza triangolare: se x, y, x', y' sono punti di T per cui $d(x,x') < \varepsilon/2$ e $d(y,y') < \varepsilon/2$, con $\varepsilon < 0$ prefissato ad arbitrio, allora $|d(x',y') - d(x,y)| < \varepsilon$; cioè il criterio $(\varepsilon, \delta_\varepsilon)$ di

¹⁹ La dimostrazione è immediata se O, X e Y sono distinti e non collineari, perché i triangoli di vertici ordinati $\langle X_q, O, X \rangle$ e $\langle Y_p, O, Y \rangle$ sono simili, e i segni di OX_q e di OY_p sono comunque gli stessi. E anche facile provare la (1) quando O, X e Y sono collineari, esaminando separatamente i casi degeneri.

Weierstrass è valido con $\delta_\varepsilon = \varepsilon/2$.²⁰ In H_n , inoltre, la proiezione ortogonale di un punto X su una retta p (con $X_p = X$ se $X \in p$) è funzione continua rispetto al suo argomento X , perché $d(X'_p, X_p) \leq d(X, X')$ in forza del teorema di Pitagora. Segue che il reale nei due membri della (1) è funzione continua dei suoi argomenti X e Y .

Se $X \neq O$ e $Y \neq O$, la (1) può risciversi nella forma:

$$(2) \quad OY_p/d(O, Y) = OX_q/d(O, X).$$

Diremo per il momento M il reale nei due membri della (2). Quindi M è funzione continua dei suoi argomenti X, Y finché questi punti sono fuori da una palla chiusa di raggio arbitrario > 0 e centro in O . Inoltre (teorema di Talete), M resta costante se a X [a Y] si sostituisce un X' [un Y'] arbitrario sul raggio $p' =: \mathcal{SR}(O|X)$ [sul raggio $q' =: \mathcal{SR}(O|Y)$]; e ciò implica che M dipenda in realtà soltanto dai raggi p' e q' , $M = M(p', q')$, risultando in particolare $M = 1$ per $q' = p'$ e $M = -1$ per $q' = p'^*$. Esclusi questi due casi, sempre in forza del teorema di Talete risulta $M \in (-1, 1)$. Inoltre, supposti p' e q' (di un fascio di origine O) distinti e non complementari, se la coppia di raggi r' e s' dello stesso fascio è congruente alla coppia p' e q' , cioè $p'|q' \approx r'|s'$ (questo implica che anche r' e s' siano distinti e non complementari), allora $M(p'|q') = M(r'|s')$ (la facile verifica è lasciata al lettore). In definitiva, M dipende soltanto dall'angolo libero $[p'|q']$; ovvero, in forza della biiezione esistente tra angoli liberi e loro misure normalizzate (diciamo, secondo $\mu(\mathbb{R}) = \pi/2$, o misure naturali), M deve dipendere soltanto dalla misura naturale (μ_{nat}) di tale angolo libero $[p'|q']$. Insomma,

$$(3) \quad M = M(\mu_{\text{nat}}([p'|q'])).$$

Questa misura è stata a suo tempo definita come elemento di separazione di una specifica D -sezione dell'insieme dei numeri diadici positivi e minori di π , sottoinsieme dei razionali nello stesso intervallo aperto $(0, \pi)$. Nel seguito, sarà conveniente scrivere $|\mathfrak{A}|, |\mathfrak{B}|, |\mathfrak{C}|, \dots$ in luogo di $\mu_{\text{nat}}(\mathfrak{A}), \mu_{\text{nat}}(\mathfrak{B}), \mu_{\text{nat}}(\mathfrak{C}), \dots$. Posto allora $\mathfrak{A} =: [p'|q']$, semplificheremo la scrittura della (3) nella più maneggevole

$$(3') \quad M = M(|\mathfrak{A}|).$$

Tale M è una funzione continua a valori in $(0, 1)$ ²¹ del suo argomento $|\mathfrak{A}| \in (0, \pi)$, e inoltre $M(|\mathfrak{A}| \rightarrow 0+) = 1$, $M(|\mathfrak{A}| \rightarrow \pi-) = -1$, $M(\pi/2) = 0$. Nel seguito denoteremo questa funzione per ora

²⁰ Infatti, poiché $d(x, x') + d(x', y') + d(y', y) \geq d(x, y)$ e $d(x', x) + d(y, y') + d(x, y) \geq d(x', y')$, la tesi segue sottraendo membro a membro e prendendo il valore assoluto.

²¹ Tornando alla definizione di M come $OY_p/d(OY)$, se si vincola $Y (\in p)$ a essere sul cerchio $d(O, Y) = 1$, abbiamo $M = OY_p$; e in questa forma, M è a maggior ragione (rispetto a quanto provato per Y comunque variabile in p) una funzione continua di Y , per p (orientata) fissa. D'altra parte, se si vincola Y ad essere in uno dei semipiani di p di bordo p e sul cerchio $d(O, Y) = 1$, allora Y è in corrispondenza biunivoca con $|\mathfrak{A}|$, e $Y(|\mathfrak{A}|)$, e quindi $OY_p(|\mathfrak{A}|)$, è continua.

incognita di $|\mathfrak{A}| \in (0,\pi)$ con Cos ; ²² se essa fosse nota e invertibile, la (3') ci fornirebbe un'interpretazione alternativa (ma equivalente) di $|\mathfrak{A}|$, secondo la $|\mathfrak{A}| = \text{Cos}^{-1}(M)$. Nell'App. Spec. 1.B è dimostrato che in realtà la teoria H_n provvede informazioni sufficienti ad identificare completamente Cos in $(0,\pi)$, e che risulta ivi $\text{Cos} \equiv \cos$; e quindi, assegnando a $\text{Cos}(0)$ e $\text{Cos}(\pi)$ i corrispondenti valori limite (in forza della continuità), che $\text{Cos} \equiv \cos$ anche in $[0,\pi]$. Concludendo, potremo scrivere:

$$(4) \quad X \cdot Y = |X||Y|\cos(|\mathfrak{A}|),$$

ove \mathfrak{A} è l'angolo libero $[\mathcal{SR}(O|X)|\mathcal{SR}(O|Y)]$ (se i due raggi sono distinti e non complementari). Nei casi limite $|\mathfrak{A}| \rightarrow 0+$ e $|\mathfrak{A}| \rightarrow \pi-$, la (4) diventa $X \cdot Y = \pm|X||Y|$. Se poi X e/o Y sono nulli ($= O$), e quindi \mathfrak{A} non è definito, porremo $X \cdot Y = 0$, in conformità alla già data definizione.

Nel modello \mathbf{R}^n , abbiamo per definizione

$$(5) \quad x \cdot y = \sum_i x_i y_i;$$

ma possiamo subito calcolare il secondo membro, supposti x e y non nulli, servendoci di un riferimento cartesiano ortogonale (di origine O in H_n) in cui $x_{i>2} = 0$ e $y_{i>2} = 0$, e prendendo come asse coordinato (1) la retta orientata da O a $X \leftrightarrow x$, per cui $x_1 = |x|$. Allora la somma nella (5), che come sappiamo è invariante rispetto a rotazioni, si riduce a $|x|y_1$, ove y_1 è la coordinata sull'asse (1) della proiezione ortogonale di $Y \leftrightarrow y$ su di esso; e quindi $y_1 = |y|\cos\alpha$ se α è l'"angolo" tra i "raggi" Ox e Oy nella definizione standard della trigonometria (lunghezza dell'arco di cerchio di centro nell'origine e raggio 1 sotteso da Ox e Oy , nel "piano" di O, x e y). Questo angolo α è in $(0,\pi)$, o al limite in $[0,\pi]$ volendo includere i casi di x e y proporzionali secondo un fattore positivo o negativo. Quindi, finché x e y non sono nulli,

$$(6) \quad x \cdot y = |x||y|\cos\alpha, \quad ^{23}$$

per $\alpha \in [0,\pi]$; se poi $x = 0$ e/o $y = 0$, e quindi α non è definito, allora $x \cdot y = 0$ in conformità alla (5). In definitiva la definizione di $X \cdot Y$ da una parte, e quella di $x \cdot y$ dall'altra assicurano che $X \cdot Y = x \cdot y$, secondo la definizione standard di prodotto interno indotto da \mathbf{R}^n su H_n ; e poiché la corrispondenza canonica (\leftrightarrow) tra H_n e \mathbf{R}^n è una isometria, cioè $|X| = d(O,X) = *d(0,x) = |x|$ per ogni X di H_n corrispondente a x di \mathbf{R}^n , se $|X| = |x| \neq 0$ e $|Y| = |y| \neq 0$ concludiamo che $\cos|\mathfrak{A}| = \cos\alpha$, cioè che $|\mathfrak{A}| = \alpha$. È precisamente questa la preannunciata interpretazione alternativa di \mathfrak{A} offerta dal modello \mathbf{R}^n .

²² Abbiamo scritto Cos con la maiuscola per evitare che venga confusa con la funzione "coseno", ovunque denotata con "cos".

²³ La (6) è un teorema di \mathbf{R}^n se α è definito secondo la modalità standard della trigonometria. Ma se si rinunciava per un momento a tale definizione, sotto $x \neq 0$ e $y \neq 0$ essa fornirebbe una *definizione* di α sfruttando l'invertibilità (rispetto ad α) della $\cos\alpha = x \cdot y / |x||y|$ per $\alpha \in [0,\pi]$, tenuto conto che il 2° membro è in $[-1,1]$ (disuguaglianza di Schwarz).

L'algebra "di origine O" testé istituita su H_n (somma di due punti, moltiplicazione di un reale per un punto, prodotto interno di due punti) è certamente naturale e significativa in ordine agli interrogativi sollevati nel primo paragrafo di questa sottosezione, ma ha la proprietà (o piuttosto il limite) di *dipendere* dall'origine O prescelta, perché i risultati della generica operazione dipendono da O. C'è tuttavia una parziale eccezione: la sottrazione di X da $Y \neq X$ (definita da $Y - X \equiv Y + + (-1)X$) *non dipende*, seppur in un senso un po' particolare, da quella scelta. Denotiamo per un momento, per maggior chiarezza, con $-_O$ la sottrazione nell'algebra di origine O, e con $-_{O'}$ la sottrazione nell'algebra di origine O'. Si verifica allora senza difficoltà che i punti $Z_O =: Y -_O X$, e rispettivamente $Z_{O'} = Y -_{O'} X$ sono tali che i segmenti orientati OZ_O e $O'Z_{O'}$ sono paralleli-equiversi e congruenti. In generale, due segmenti orientati XY ($X \neq Y$) e ZU ($Z \neq U$) di H_n si dicono paralleli-equiversi e congruenti se, essendo coplanari, sono lati opposti di un parallelogramma, oppure se $XY = ZU$ (cioè se $X = Z$ e $Y = U$). La sopradescritta relazione di $\langle X, Y \rangle$ verso $\langle Z, U \rangle$ è manifestamente un'equivalenza; l'insieme delle coppie ordinate di $H_n \times H_n$ è quindi ripartito in corrispondenti classi di equivalenza (disgiunte) che si dicono Vettori²⁴ di H_n , e sarà qui denotato come H_n ("spazio Vettoriale associato a H_n ").

Questo insieme H_n riceve una naturale struttura di spazio lineare con prodotto interno fissando un'origine O di H_n , facendo corrispondere al generico Vettore $X (\in H_n)$ il punto $X (\in H_n)$, unicamente determinato, per il quale $\langle O, X \rangle \in X$ (e similmente al Vettore Y il punto Y per il quale $\langle O, Y \rangle \in Y$, ecc.), e infine definendo $X + Y$ come il corrispondente (sotto \leftrightarrow , o meglio \leftrightarrow_O) di $X + Y$, αX come il corrispondente di αX e $X \cdot Y$ come $X \cdot Y$; ove ovviamente le operazioni in H_n sono quelle dell'algebra di origine O. Si verifica facilmente che l'algebra così indotta su H_n gode della proprietà *di essere indipendente da O*. Con quest'algebra O-indipendente indotta da H_n , H_n si dice **spazio Vettoriale associato a H_n** . La completa analogia tra il rapporto " H_n vs. H_n " e il rapporto " \mathbf{R}^n vs. \mathbf{R}^n " è evidente. Infine, se partendo da H_n con la sua algebra O-indipendente se ne costruisce lo spazio affine associato (secondo le modalità generali illustrate in questa sezione) avente come supporto l'insieme (senza struttura) H_n e come origine O, si riottiene esattamente lo spazio H_n con l'algebra di origine O, diciamo ${}_O H_n$.

Come sappiamo, convenendo di porre in $O \in H_n$ l'origine del sistema di riferimento (cartesiano ortogonale) con cui si passa allo spazio \mathbf{R}^n delle relative coordinate, ${}_O H_n$ e \mathbf{R}^n (quest'ultimo con la sua algebra standard) sono in corrispondenza biunivoca e isomorfi. Ma ${}_O H_n$ è anche in corrispondenza biunivoca con H_n (con $(X \leftrightarrow X) \Leftrightarrow (\langle O, X \rangle \in X)$) e ad esso isomorfo; e d'altra parte \mathbf{R}^n è in corrispondenza biunivoca con \mathbf{R}^n (con $(x_{\#} = x) \Leftrightarrow (\langle 0, x \rangle \in x)$), e ad esso

²⁴ La maiuscola serve a distinguere i "Vettori" di H_n dai "vettori" di \mathbf{R}^n .

isomorfo. Sotto queste condizioni abbiamo dunque le corrispondenze biunivoche $x \leftrightarrow x \leftrightarrow X \leftrightarrow X$ tra \mathbf{R}^n , \mathbf{R}^n , ${}_0H_n$ e H_n , tutti isomorfi tra loro. Se più in generale (e con maggior simmetria) si pone o al posto di 0 nella precedente corrispondenza (cioè, se $(x_{\#} = x) \Leftrightarrow (\langle o, x \rangle \in x)$), l'isomorfismo sussiste tra \mathbf{R}^n , ${}_o\mathbf{R}^n$, ${}_oH_n$ e H_n .

1.4) QUESTIONI DI ESTENSIONE E ORIENTAMENTO, CONCLUSIONI

1.4.1) ESTENSIONE

Ci occuperemo ora della “estensione”, o “misura”, di regioni geometriche di diversa “dimensionalità” come le “linee”, le “superficie” e i “solidi”, ossia rispettivamente della loro “lunghezza”, “area” e “volume”. Premettiamo una definizione. Due insiemi I_1 e I_2 di punti di uno spazio metrico si dicono **congruenti** (cioè simili con fattore 1) se esiste una biiezione di I_1 su I_2 , diciamo $\mathcal{J}: I_1 \rightarrow I_2$, che preserva la distanza tra coppie di punti qualsiasi, cioè tale che $\forall (P, Q) \in I_1$, $d(P, Q) = d(\mathcal{J}(P), \mathcal{J}(Q))$ (se questo è vero, lo stesso vale ovviamente scambiando I_1 con I_2 , e \mathcal{J} con \mathcal{J}^{-1}). La definizione è in accordo con quella data a suo tempo (v. S.sez 1.2.3) quando I_1 e I_2 si riducevano entrambi ad una coppia di punti distinti. L’essere congruenti, per due insiemi di uno spazio metrico, è una relazione di equivalenza, e precisamente una isometria. La definizione di congruenza sarà qui riferita a insiemi di una retta H_1 , di un piano H_2 , dello spazio H_3 , e in generale dello spazio $H_{n>3}$ posto che gli assiomi dimensionali ne prevedano l’esistenza¹.

Nella S.sez. 1.2.4 abbiamo definito la misura, che conviene ora chiamare “lunghezza” e denotare L , di segmenti h, k, \dots , in particolare inclusi in una retta che consideriamo come lo spazio base H_1 al quale facciamo ora riferimento. Questa lunghezza è un numero reale con le seguenti proprietà:

1. per ogni h , $L(h) > 0$ (**positività**)² ;
2. se h, k sono internamente disgiunti e tali che $h \cup k$ è un segmento (cioè se essi hanno in comune esattamente un estremo), allora $L(h \cup k) = L(h) + L(k)$ (**additività**);
3. $L(h) = L(k)$ se h e k sono congruenti (**invarianza rispetto alla congruenza**);
4. $L(h) = 1$ se h è il segmento campione unitario (**normalizzazione**).

Sulla base degli assiomi di H_1 si può dimostrare che le 1. ÷ 4. sono caratteristiche, cioè che non possono esistere due funzioni-lunghezza *distinte* con le proprietà 1. ÷ 4. D’altra parte, un dato segmento può sempre decomporre in un numero finito di segmenti internamente disgiunti, e allora la lunghezza del segmento è uguale alla somma dei segmenti componenti. Come vedremo, questo fatto colloca il presente caso di H_1 in una posizione singolare rispetto a quella degli $H_{n>1}$, perché la

¹ Salvo avviso in senso contrario, questa possibilità sarà per lo più tacitamente sottintesa nel seguito.

² Nella S.sez. 1.2.4 abbiamo definito la lunghezza (misura) di un segmento dello spazio metrico S come il valore di un’applicazione di $S^2 \setminus \Delta_S$ in $\mathbb{R}_{>0}$. Volendo, si potrebbero includere tra i segmenti di S segmenti degeneri con estremi coincidenti e lunghezza nulla. Questo corrisponderebbe a sostituire $S^2 \setminus \Delta_S$ con S^2 e $\mathbb{R}_{>0}$ con $\mathbb{R}_{\geq 0}$. In questo caso il punto 1. diventerebbe 1’. “per ogni h , $L(h) \geq 0$ (**non-negatività**), con $L(h) = 0$ sse h è degenere”. Analoghe modifiche sarebbero possibili più oltre in casi analoghi.

decomposizione di un segmento non dà luogo ad insiemi-componenti di tipo diverso da quello di partenza.

Poiché H_n è munito della topologia indotta dalla metrica, *sotto convenienti condizioni* che ne assicurino l'esistenza possiamo considerare il limite di una famiglia di spezzate tra due punti fissi, quando la lunghezza del segmento più lungo, tra quelli che compongono la singola spezzata, tende a zero (e quindi il numero dei suoi segmenti tende a infinito). Questo insieme-limite si dice **linea** o **curva**³ di H_n , e il limite della somma delle lunghezze dei segmenti componenti, che esiste finito sotto le condizioni assunte, sua **lunghezza (rettificazione di una curva)**. Tuttavia questa nozione appartiene già a pieno titolo alla teoria delle *varietà* immerse in H_n , e non sarà approfondita in questo capitolo.

Una nozione simile a quella della lunghezza dei segmenti di una retta può essere convenientemente estesa a certi insiemi di punti, o figure, del piano H_2 , che si dicono **poligoni** o meglio **poligoni semplici** (cioè senza buchi, o semplicemente connessi), o **s-poligoni** (s come semplice). In H_2 , un **bordo poligonale** è definito come una spezzata finita *chiusa e semplice*, cioè costituita da un numero finito di segmenti (internamente) disgiunti tali che ogni loro estremo sia estremo di esattamente due segmenti (quindi il numero degli estremi è uguale a quello dei segmenti).⁴ Secondo un teorema dimostrato da Jordan (Camille, Lione 1838, Milano 1922), questo bordo decompone il piano in esattamente due regioni connesse e internamente disgiunte, delle quali quella limitata (\equiv di diametro finito, o **regione interna**) si dice **s-poligono associato** a quel bordo⁵. I segmenti del bordo si dicono **lati** del s-poligono. Gli s-poligoni più elementari (cioè con numero minimo (3) di lati) sono evidentemente i triangoli (non degeneri).

È allora naturale definire un'area del generico s-poligono P di H_2 , diciamo $A(P)$, con proprietà strettamente analoghe alle 1. ÷ 4., e cioè:

1'. per ogni P , $A(P) > 0$ (positività);

³ Per quanto strano possa apparire, in geometria non esiste una chiara distinzione tra "linea" e "curva": la terminologia appare cioè più dettata dall'uso che da una classificazione razionale. A lume di buon senso, una linea dovrebbe essere un oggetto più generale di una curva: perché esiste una linea retta, ma certamente non una curva retta. Il problema nasce nel momento in cui si conviene che una retta sia una particolare curva degenera, una "curva senza curvatura". A questo punto la scelta più corretta è quella di ignorare il problema lessicale per concentrarsi sulle distinzioni concrete, usando così le due parole come sinonimi. Si parlerà così di "linee \approx curve nel senso di ...": ad esempio "nel senso di Jordan", "nel senso di Hilbert", "nel senso di Cantor", "nel senso di Urisohn", e così via. Per i nostri presenti fini, basterà menzionare il prevalente significato di linea \approx curva come "varietà astratta unidimensionale" (vedi Cap 4), e in sott'ordine, come "immersione, di classe $C^{r \geq 1}$, di un intervallo (tipicamente aperto) in uno spazio euclideo o "pseudoeuclideo".

⁴ Alternativamente, si può anche definire un bordo poligonale assegnando punti distinti $A_0, A_1, \dots, A_m \equiv A_0$ di H_2 , con $m \geq 3$, e tali che gli m segmenti $A_0A_1, \dots, A_{m-1}A_m$ non abbiano punti interni comuni. Si può anche richiedere, come alcuni fanno, che non vi siano coppie di segmenti consecutivi collineari, ma ciò è inessenziale ai nostri presenti scopi.

⁵ Un modo equivalente di distinguere le due regioni è che ogni retta per due punti della regione limitata ha almeno due punti sul suo bordo, mentre esistono rette per due punti della regione illimitata disgiunte dal bordo.

2'. se P e Q sono s -poligoni internamente disgiunti e tali che $P \cup Q$ è un s -poligono, $A(P \cup Q) = A(P) + A(Q)$ (additività);

3'. $A(P) = A(Q)$ se P e Q sono congruenti (invarianza rispetto alla congruenza);

4'. $A(P) = 1$ se P è un quadrato di lato unitario (normalizzazione).

Ancora, sulla base degli assiomi di H_2 si dimostra che queste proprietà sono caratteristiche: esiste al più una funzione A definita sugli s -poligoni di H_2 e con valori in $\mathbb{R}_{>0}$ con le proprietà richieste. Un s -poligono può sempre decomporre in un numero finito di s -poligoni internamente disgiunti; e in particolare di triangoli, cioè di s -poligoni generalmente più elementari del s -poligono di partenza. Questo processo corrisponde alla decomposizione finita di un segmento in segmenti internamente disgiunti, e si dice **triangolazione** del s -poligono dato. In H_2 si ha tuttavia una situazione diversa da quella in H_1 ; per avere l'area di un s -poligono basta definire quella del generico triangolo e applicare la proprietà 2'. Naturalmente occorrerà assicurarsi che questa definizione sia autoconsistente; e ciò può effettivamente dimostrarsi, provando che l'area così definita di un s -poligono è indipendente dalla sua decomposizione in triangoli, in accordo con la sua sopravvissuta unicità. Si noti che anche un triangolo può decomporre in triangoli; in questo caso la prova dell'indipendenza dell'area dalla decomposizione è elementare. L'area di un triangolo si definisce poi come il semiprodotto della lunghezza di un suo lato, che diciamo convenzionalmente "base" del triangolo, per la distanza del vertice opposto dalla sua proiezione ortogonale sulla retta della base, che diciamo convenzionalmente "altezza" del triangolo, e il risultato non dipende dal lato scelto come base. Equivalentemente, l'area del triangolo si può anche calcolare mediante la classica formula di Erone (Alessandria, I sec a.C. – I sec. d.C.), in realtà dovuta ad Archimede, in termini delle lunghezze dei tre lati.⁶ Secondo questa formula, l'area di un triangolo rettangolo è uguale al semiprodotto delle lunghezze dei suoi cateti, quella di un rettangolo è uguale al prodotto delle lunghezze di due suoi lati contigui, e quella di un quadrato è uguale al quadrato della lunghezza di un suo lato, in accordo con la normalizzazione richiesta al punto 4'.

Il processo di triangolazione (finita) permette di definire l'area di regioni piane connesse più generali degli s -poligoni. Queste sono le regioni connesse che si lasciano decomporre *finitamente* in unioni di triangoli (non degeneri) internamente disgiunti, o "poligoni" in generale non semplicemente connessi. Secondo l'importante teorema di **Wallace-Bolyai-Gerwein** (vide), la condizione che due poligoni siano **equidecomponibili**, cioè che esista una decomposizione finita internamente disgiunta dell'uno e dell'altro i cui componenti siano due a due congruenti, è

⁶ Il polinomio sotto radice nella formula di Erone è di grado 4 nelle lunghezze dei lati ed è strettamente positivo salvo il caso di un triangolo degenere, con i tre vertici collineari. L'esistenza di una formula come quella di Erone ci conferma che l'area non dipende dal lato scelto come base.

necessaria, oltre che sufficiente, per l'eguaglianza delle loro aree. *Mutatis mutandis*, *non* sussiste un analogo teorema in $H_{n \geq 3}$.

Muniti della definizione di area dei poligoni, possiamo passare a quella di certe regioni piane più generali dei poligoni mediante una procedura indicata da Peano (1883, 1887). Per un sottoinsieme M limitato (\equiv contenuto in un dato rettangolo) di H_2 che “sia quadrabile (o abbia un'area) secondo Peano”, si suppone che esistano due classi di poligoni, la classe $(i)_M$ e la classe $(e)_M$, per le quali $P \subset M \forall P \in (i)_M$ e $M \subset Q \forall Q \in (e)_M$; e quindi $A(P) \leq A(Q)$ per P e Q comunque scelti nelle due classi (la differenza $Q \setminus P$ è un poligono, possibilmente degenere). Detto allora $A_i(M)$ l'estremo superiore (sup) di $A(P)$ al variare di P in $(i)_M$, e $A_e(M)$ l'estremo inferiore (inf) di $A(Q)$ al variare di Q in $(e)_M$, risulta evidentemente $A_i(M) \leq A_e(M)$; se in particolare risulta $A_i(M) = A_e(M)$, allora M si dice **quadrabile secondo Peano**, e la sua **area secondo Peano** si definisce come il valore comune dei due estremi. Alla classe delle regioni Peano-quadrabili appartengono evidentemente i poligoni, e la loro area (come già definita) coincide banalmente con quella secondo Peano.

Equivalentemente (si dimostra), ci si può limitare a considerare classi di poligoni $P \subset M \subset Q$ che siano unioni finite di rettangoli internamente disgiunti con lati paralleli a due direzioni (ortogonali) prestabilite di H_2 , diciamo x^\wedge e y^\wedge . Tali particolari poligoni si possono a loro volta decomporre in rettangoli internamente disgiunti ottenuti mediante una conveniente **reticolazione** (o **quadrettatura**) finita, con rette parallele a x^\wedge e y^\wedge , del rettangolo-base E (diciamo, chiuso) con lati paralleli a x^\wedge e y^\wedge e includente M , $M \subset E$. Questa possibilità fu indicata da Jordan nel 1892, per cui si parla anche di **insiemi di H_2 quadrabili secondo Jordan**, o **J-quadrabili** o **jordaniani**. La definizione si trova in ogni testo istituzionale di Analisi, e la riproduciamo qui in breve per comodità del lettore. Vale a dire, se δ_i è il generico rettangolo della quadrettatura di E *incluso* in M ($\delta_i \subset M$, quindi $\cup_i \delta_i \subset M$), e Δ_j il generico rettangolo della quadrettatura *congiunto* con M ($\Delta_j \cap M \neq \emptyset$, quindi $M \subset \cup_j \Delta_j$), si ha $A_i(M) =: \sup \sum_i A(\delta_i)$ e $A_e(M) =: \inf \sum_j A(\Delta_j)$. Si noti che i rettangoli di cui parliamo sono “completi” (includono cioè i loro lati) e quindi la reticolazione di E *non* è una sua decomposizione in senso stretto (che presuppone la disgiunzione dei componenti, fermo restando che essi formino un ricoprimento giusto). I rettangoli della reticolazione si possono disgiungere in vario modo (sempre ricoprendo E); così facendo, le somme $\sum_i A(\delta_i)$ e $\sum_j A(\Delta_j)$ – assegnando ad un rettangolo incompleto la stessa area del suo associato completo – possono cambiare, ma i loro due estremi (il sup della prima e l'inf della seconda), cioè $A_i(M)$ e rispettivamente $A_e(M)$, restano gli stessi. Nella teoria della misura (vedi App. Gen. D), è norma riferirsi sempre a decomposizioni in senso stretto (ricoprimenti giusti a componenti *disgiunti*);

quindi converrà presupporre tali le reticolazioni (o le n -reticolazioni, passando a $n > 2$ dimensioni) del rettangolo (o n -rettangolo) base. La J -quadrabilità di M , espressa ancora dalla (*) $A_i(M) = A_e(M)$, può equivalentemente scriversi nella forma (**) $A(E) = A_e(M) + A_e(E \setminus M)$, oppure (***) $A(E) = A_i(M) + A_i(E \setminus M)$, ove $A(E)$ è l'area standard di E .⁷ Si può provare che: «se M è J -quadrabile, anche la sua chiusura M^- (rispetto alla metrica standard, vedi App. Gen. C) lo è, e le due aree sono uguali». È anche evidente che «un insieme M di H_2 è J -quadrabile sse esistono unioni del tipo $\cup_j \Delta_j$ e unioni del tipo $\cup_i \delta_i$ per le quali è $0 \leq A(\cup_j \Delta_j) - A(\cup_i \delta_i) \equiv A((\cup_j \Delta_j) \setminus (\cup_i \delta_i)) < \varepsilon$, con $\varepsilon > 0$ arbitrario; il che è come dire sse il suo contorno (vedi ancora App. Gen. C) è J -quadrabile e ha J -area nulla». Infine, «la J -area di M , se esiste, *non* dipende dalla coppia di direzioni ortogonali x^\wedge, y^\wedge prescelte (quindi la nozione di J -area è in tal senso intrinseca)».

La situazione si complica, sotto uno specifico aspetto, passando dal piano (2-dim) allo spazio (3-dim). Anche in quest'ultimo si può definire un **bordo poliedrale** come unione finita, internamente disgiunta e connessa, di s -poligoni, tali che ogni lato di s -poligono (\equiv spigolo del poliedro) è lato di esattamente due s -poligoni. Questo bordo, si dimostra, decompone lo spazio in due regioni connesse e internamente disgiunte, delle quali quella interna (di diametro finito) si dice **poliedro**, o meglio **poliedro semplice** (senza buchi, o semplicemente connesso), o **s -poliedro**. Gli s -poligoni del bordo di un s -poliedro si dicono sue **facce**. Ancora, si vuole che il **volume** di un s -poliedro sia un numero reale soddisfacente a proprietà (diciamo $1'' \div 4''$) in tutto analoghe a quelle dell'area di un poligono (naturalmente deve ora valere 1 il volume di un *cubo* di lato 1). Sotto queste proprietà, il volume si dimostra esistere unico (ciò è alquanto meno semplice che nell'analogo caso piano). Un s -poliedro può decomporsi in un numero finito di s -poliedri – e in particolare in **tetraedri**, che sono gli s -poliedri più elementari (a 4 facce) – internamente disgiunti; per semplicità, questo processo di decomposizione in tetraedri si dice ancora **triangolazione** del dato s -poliedro. Il volume del s -poliedro si definisce allora come la somma dei volumi dei tetraedri componenti, e non dipende dalla triangolazione effettuata. Quanto al volume del tetraedro, esso è uguale a $1/3$ del prodotto dell'area di una sua faccia, convenzionalmente scelta come “base”, per l’“altezza” associata, cioè per la distanza del vertice opposto alla base dal piano della base stessa, e non dipende dalla faccia prescelta (Democrito, 460-370 a.C.); oppure, equivalentemente, è dato da una tarda (XIX secolo) generalizzazione della formula di Erone, omogenea di grado 3 nelle

⁷ Infatti $A_e(E \setminus M) = A(E) - A_i(M)$ e $A_i(E \setminus M) = A(E) - A_e(M)$ (quindi $A_e(E \setminus M) - A_i(E \setminus M) = A_e(M) - A_i(M) \geq 0$). Se vale la (*), le (**) e la (***) si riducono entrambe ad identità del tipo $0 \equiv 0$; se viceversa vale la (**) o la (***), allora vale la (*). Si noti anche che la prima delle precedenti uguaglianze consentirebbe di *definire* $A_i(M)$ come $A(E) \setminus A_e(E \setminus M)$, piuttosto che come $\sup \sum_i A(\delta_i)$.

lunghezze dei sei spigoli ⁸. Formando unioni finite e connesse di tetraedri internamente disgiunti si possono ottenere regioni spaziali, a bordo composto di triangoli, più generali degli s-poliedri, in quanto non-semplicemente connesse, o “poliedri”; il loro volume è allora definito al solito modo, mediante applicazione di 2’.

Tuttavia a questo punto sorgono difficoltà, rispetto al caso piano, con la definizione dell’uguaglianza dei volumi di generici poliedri per loro “equidecomposizione” *finita*, in quanto nello spazio non sussiste, come già detto, un analogo del teorema di Wallace-Bolyai-Gerwein. Poliedri aventi di fatto uguale volume (come accertabile con mezzi diversi dalla equidecomposizione finita ⁹) non sono sempre finitamente equidecomponibili: ovvero, la equidecomponibilità finita di poliedri è condizione *sufficiente*, ma non *necessaria* all’uguaglianza dei loro volumi. Ad esempio un tetraedro regolare ed un cubo di uguale volume non sono finitamente equidecomponibili. ¹⁰ Questo fatto rende innaturale una generalizzazione dell’idea di Peano per la definizione del volume di regioni spaziali volumetricamente approssimabili mediante poliedri interni ed esterni. È invece sempre possibile, e in modo ovvio, estendere allo spazio il metodo di Jordan, includendo il sottoinsieme S (limitato) di H_3 in oggetto in un parallelepipedo

⁸ Secondo tale formula, il volume del tetraedro è uguale a $(12 \cdot \sqrt{2})^{-1}$ volte la radice quadrata del modulo del **determinante di Cayley-Menger** (vide), che è strettamente positivo salvo il caso degenerare in cui i quattro vertici sono coplanari. Ciò conferma l’indipendenza del risultato dalla scelta della faccia usata come base nella prima definizione. Il determinante di Cayley-Menger (C.M.) vale 4 per il tetraedro *regolare* di lato unitario, per cui il volume di questo risulta uguale a $\sqrt{2}/12$. Esiste anche una versione generalizzata della formula di cui sopra, che dà il volume dell’(n+1)-edro di $H_{n \geq 1}$ in termini delle lunghezze dei suoi $n(n+1)/2$ spigoli, e che naturalmente riproduce la formula di Erone per $n = 2$ e quella appena illustrata per $n = 3$ (oltre a dare 1 per $n = 1$). Ricordiamo qui che i (m+1)-edri *regolari* immersi in H_n , con $1 \leq m \leq n$, sono quelli i cui m+1 vertici sono tutti equidistanti l’uno dall’altro. Per $m = 1$ abbiamo un segmento, per $m = 2$ un triangolo equilatero (o regolare), per $m = 3$ un tetraedro regolare, .. e così via. È significativo che m *non possa superare* n; cioè, che in H_n *non esista* un insieme di più di n+1 punti equidistanti l’uno dall’altro. In H_n , un (n+1)-edro non degenerare è il più semplice n-solido (solido n-dim), ed è definito come l’“inviluppo convesso”, o “n-simplesso”, degli n+1 punti che ne sono i vertici non co-(n-1)-planari (vedi la S.sez. 5.1.2, dove gli n-simplessi sono introdotti e discussi). Se si nominano con 0, 1, .. n, gli n+1 vertici di un n-simplesso, e si denotano con d_{ij} ($i, j = 0, .. n$) le loro distanze (per cui $d_{ii} = 0$), la (n+2)×(n+2)-**matrice di C.M.** si ottiene aggiungendo alla (n+1)×(n+1)-matrice simmetrica $\{d_{ij}^2\}$ una (n+2)-ma riga di 1 e una (n+2)-ma colonna di 1, salvo il loro comune ultimo elemento 0. Il risultato è ancora una matrice simmetrica con la diagonale principale fatta di zeri. Si dimostra inoltre che, se tutti i d_{ij} sono uguali a 1 (caso dell’n-simplesso regolare unitario), il determinante della matrice di C.M. vale n+1 se n è dispari e -(n+1) se n è pari. Secondo la sopraccennata formula generale di C.M., detto C_n il determinante della matrice di C.M., il volume dell’n-simplesso è $|_nS| = [2^{-n}|C_n|]^{1/2}/n!$; quindi, nel caso dell’n-simplesso regolare unitario ($d_{ij} = 1$), abbiamo $|_nS| = [2^{-n}(n+1)]^{1/2}/n!$, e dunque $|_1S| = 1$, $|_2S| = \sqrt{3}/4 = 0,433$, $|_3S| = \sqrt{2}/12 = 0,118$, $|_4S| = \sqrt{5}/96 = 2,33 \cdot 10^{-2}$, .. e così via. La verifica *diretta* della seconda e della terza di queste è immediata. Precisamente, l’altezza del triangolo equilatero di lato 1 è $\sqrt{3}/2$, la lunghezza della sua base è 1, e la sua area è $|_2S| = (1/2) \cdot \text{base} \cdot \text{altezza} = \sqrt{3}/4$. L’altezza del tetraedro regolare di spigolo 1 è $\sqrt{(2/3)}$, l’area $|_2S|$ della sua base è $\sqrt{3}/4$, e il suo volume è $|_3S| = (1/3) \cdot \text{base} \cdot \text{altezza} = (1/3) \cdot \sqrt{3}/4 \cdot \sqrt{(2/3)} = \sqrt{2}/12$.

⁹ Antesignano di questi mezzi è quel “metodo di esaustione” dei matematici greci, ripreso da alcuni matematici italiani del XV secolo (soprattutto da Cavalieri (Bonaventura, 1598-1647)), che a buon diritto può considerarsi come precursore della teoria della misura e quindi del calcolo integrale. Naturalmente si verifica che il detto metodo è in accordo con la definizione generale di volume attraverso le proprietà 1’’ ÷ 4’’, nonché con la definizione di J-volume, vedi appresso. Torneremo diffusamente sul metodo di esaustione nella Sez. 5.1.

¹⁰ La ragione è che l’angolo-diedro di un tetraedro regolare non è in rapporto razionale con l’angolo retto. La non equidecomponibilità finita di poliedri di ugual volume indusse Euclide a servirsi di decomposizioni non-finite nel trattare i volumi di certi solidi.

base (ad es. chiuso) a facce parallele a tre giaciture mutuamente ortogonali prestabilite, e analogamente 3-reticolando questo parallelepipedo, mediante piani paralleli alle sue facce, in parallelepipedi componenti *disgiunti*. Ancora denotando con δ_i i parallelepipedi della 3-reticolazione inclusi in S e con Δ_j quelli congiunti con S (gli uni e gli altri generalmente incompleti), la possibile **J-cubabilità** (o in generale **J-misurabilità**) di S segue sulla falsariga della J-quadrabilità del sottoinsieme M di H_2 . Sussistono asserti analoghi a quelli enunciati nel caso bidimensionale: ad esempio, il **volume** di regioni J-cubabili non dipende dalla scelta delle tre giaciture ortogonali cui si associano il parallelepipedo base e la sua 3-quadrettatura. Infine l'**n-volume** di sottoinsiemi limitati di $H_{n>3}$ si definisce senza difficoltà in modo analogo a quello illustrato per $n = 2, 3$. In generale, la n -reticolazione del n -rettangolo base includente la regione in oggetto si dice anche una sua **decomposizione cartesiana ortogonale**.

Come la nozione di lunghezza di una curva non è limitata alle curve rettificabili piane, ma si estende a una corrispondente classe di curve dello spazio, anche quella di area non è limitata alle regioni piane quadrabili, ma si estende in modo opportuno a una corrispondente classe di “superficie” dello spazio; e così via passando a spazi di dimensione maggiore; ma come abbiamo già fatto a proposito delle curve, non approfondiremo qui questo tipo di generalizzazioni ¹¹.

La nozione di J-misura di regioni limitate di H_n (non degeneri, cioè non incluse in un suo $(m < n)$ -iperpiano), e quindi del suo modello canonico \mathbf{R}^n , è sufficiente alla maggior parte delle applicazioni pratiche, ma è lontana dalla generalità che sarebbe desiderabile da un punto di vista matematico, in quanto esiste una moltitudine di sottoinsiemi di \mathbf{R}^n di rilevanza analitico-geometrica che *non* sono J-misurabili. A cavallo tra i due ultimi secoli, questa situazione fu correttamente vista come un crescente “stato di sofferenza” dell’analisi su \mathbf{R}^n , e condusse, per opera di Borel (Émile, Saint-Affrique Fr. 1871, Parigi 1956), di Baire (René, Parigi 1874, Chambery Fr. 1932) e di Lebesgue (Henri, Beauvais Fr. 1875, Parigi 1941), alla elaborazione di nozioni di misura più sofisticate e potenti, e alla fine ad una Teoria della Misura, nonché dell’Integrazione, fondate su basi completamente assiomatiche. Una sintetica rassegna di tali questioni (che sono significativamente rilevanti alla geometria) è offerta nell’App. Gen. C. Altre generalizzazioni, riguardanti la misura e l’integrazione su varietà (astratte o immerse), saranno trattate nel Cap. 4 e più oltre.

Il problema di precisare, mediante convenienti operazioni ideali, a cosa corrisponda l’area *empirica* di una regione piana, per semplicità connessa [il volume *empirico* di una regione spaziale

¹¹ È tuttavia opportuno segnalare che, a differenza dalla lunghezza delle curve piane, l’“area” di una superficie data in H_3 non si può definire, *in generale*, come limite delle aree di poliedri inscritti quando il contorno di questi tenda alla superficie, come ha dimostrato Schwarz (Hermann, 1843-1921) con un noto e sorprendente controesempio. La teoria rigorosa delle aree superficiali fu iniziata da Gauss e completata da Jordan (in generale con riferimento a varietà k -dimensionali immerse in H_n , $1 \leq k \leq n$, sotto convenienti condizioni) nel suo Cours d’Analyse.

connessa] non è banale, perché, almeno a prima vista, in questi casi non disponiamo di qualcosa di analogo al regolo rigido graduato che usavamo per la misura (della lunghezza) dei segmenti rettilinei. D'altra parte, in accordo con le definizioni date, il protocollo operativo che ci porta alla misura della lunghezza di un segmento si estende in modo naturale a quella della lunghezza di una generica linea empirica tra due punti dati approssimando quella linea mediante una spezzata con i vertici sulla linea stessa, riducendo la lunghezza del suo segmento più lungo ad un limite ragionevole, e identificando la lunghezza in oggetto come quella della spezzata (\equiv somma delle lunghezze dei segmenti componenti) così ottenuta.

I protocolli ideali per la misura di regioni empiriche piane/spaziali si ispirano ad un metodo alternativo (ma equivalente) a quello appena ricordato per la misura della lunghezza di una data linea tra due punti. Vale a dire, possiamo ottenere quella lunghezza sovrapponendo alla linea un filo graduato, inestendibile e incomprimibile (e abbastanza lungo), e leggendo poi su questo la lunghezza desiderata (oppure utilizzando allo stesso modo un analogo filo non graduato, segnando gli estremi della parte sovrapposta alla linea da misurare e distendendolo poi su un regolo graduato per stabilire così la lunghezza di quella parte). Riferendoci dapprima ad una regione spaziale, per misurarne il volume empirico possediamo di fatto uno strumento in tutto analogo (in senso concettuale) al filo non graduato usato nel caso della linea. Questo strumento è semplicemente un "liquido", cioè un fluido con ottima approssimazione incomprimibile e indilatabile, con il quale possiamo riempire la regione spaziale in oggetto (si supporrà di averne sempre a sufficienza). Basterà dunque misurare la quantità di liquido usato allo scopo mediante un recipiente graduato (tipicamente cilindrico) per ottenere il volume desiderato.

Passando alle regioni (empiriche) piane e alle loro aree, osserveremo innanzitutto che di fatto non esistono piani reali nello spazio empirico, ma soltanto strati piani di spessore molto piccolo, che per comodità supporremo costante. Basterà allora riempire la porzione di strato reale corrispondente alla regione data con un liquido, misurare il volume di quest'ultimo e dividere il risultato per lo spessore (piccolo ma misurabile) dello strato. Facendo idealmente tendere a zero lo spessore dello strato si ottiene la nozione (ideale) di "liquido bidimensionale". Dopo tutto, il fatto che per verniciare due m^2 di superficie piana occorra il doppio della vernice necessaria per verniciarne, *con le medesime modalità*, uno soltanto, fa parte della più elementare esperienza quotidiana. In modo analogo si provvederebbe alla misura dell'area di una superficie non piana. Come si accennerà nella S.sez. 5.1.3, questa idea addirittura suggerisce una definizione formale *alternativa* dell'area secondo Jordan (se esiste) di una regione piana, e si generalizza alle superfici di $H_3 \equiv \mathbf{R}^3$, nonché alle varietà $(m < n)$ -dimensionali di $H_n \equiv \mathbf{R}^n$.

1.4.2) ORIENTAMENTO

Ci spostiamo ora su una materia a prima vista estranea a quella dell'estensione di regioni geometriche, ma di fatto, come vedremo, ad essa collegata: quella dell'**orientamento** del generico spazio H_n . La questione è stata esaurientemente illustrata per quanto riguarda la retta H_1 . Come sappiamo (v. S.sez.1.2.2), assegnare un orientamento a una retta equivale a stabilire un ordine su una coppia di suoi punti distinti $\{X_0, Y_0\}$, ad es. l'ordine $X_0 \rightarrow Y_0$, ossia ad assegnare una coppia *ordinata* $\langle X_0, Y_0 \rangle$ di suoi punti distinti. Che una qualsiasi coppia ordinata di punti distinti $\langle X, Y \rangle$ della retta sia equiversa o antiversa alla coppia di riferimento $\langle X_0, Y_0 \rangle$ può decidersi mediante il seguente criterio, che è suscettibile di un'immediata generalizzazione ai piani, allo spazio, e in generale a spazi $(n > 3)$ -dimensionali. Se i due punti X, Y coincidono con i due punti X_0, Y_0 a meno di una permutazione (se cioè risulta $X = X_0$ e $Y = Y_0$, oppure $X = Y_0$ e $Y = X_0$), allora la coppia ordinata $\langle X, Y \rangle$ è per definizione equiversa, o rispettivamente antiversa, alla $\langle X_0, Y_0 \rangle$ a seconda che la permutazione sia pari o dispari. Escluso questo caso banale, almeno uno dei punti X, Y è distinto da X_0 . Si sposta allora X_0 in quel punto ponendo un (+) in un contatore se lo spostamento avviene "restando dalla stessa parte rispetto a Y_0 " e un (-) in caso contrario, e si chiama X'_0 la nuova posizione di X_0 . Si sposta poi analogamente Y_0 nel restante punto della coppia $\langle X, Y \rangle$ seguendo la stessa procedura, cioè ponendo un (+) nel contatore se ciò avviene restando dalla stessa parte rispetto a X'_0 , e un (-) in caso contrario, e si chiama Y'_0 la nuova posizione di Y_0 . Si è così ridotti al caso banale considerato all'inizio, e il giudizio di equiversità/antiversità di $\langle X, Y \rangle$ rispetto a $\langle X'_0, Y'_0 \rangle$ è immediato. Si pone infine un ultimo (+) nel contatore se $\langle X, Y \rangle$ è equiversa a $\langle X'_0, Y'_0 \rangle$, e un (-) in caso contrario. A questo punto la procedura è conclusa, e risulta che $\langle X, Y \rangle$ è equiversa a $\langle X_0, Y_0 \rangle$ sse il *prodotto* dei tre segni nel contatore è (+), ed antiversa sse è (-).¹²

Passando al caso di un piano, consideriamo una terna ordinata $\langle X, Y, Z \rangle$ di suoi punti distinti e non collineari, il cui orientamento è da confrontare con quello di un'analogha terna di riferimento $\langle X_0, Y_0, Z_0 \rangle$. Se i punti X, Y, Z coincidono con X_0, Y_0, Z_0 a meno di una permutazione, $\langle X, Y, Z \rangle$ è per definizione equiversa a $\langle X_0, Y_0, Z_0 \rangle$ se la permutazione è pari, e antiversa se è dispari. Escluso questo caso banale, evidentemente almeno uno dei punti X, Y, Z *non* è sulla retta p passante per X_0 e Y_0 ; si trasporta allora Z_0 in quel punto, chiamandolo Z'_0 , e si pone un (+) nel contatore se ciò avviene "restando dalla stessa parte rispetto alla retta p ", e un (-) in caso contrario. Si ripete la procedura con la nuova terna $\langle X_0, Y_0, Z'_0 \rangle$ in luogo della $\langle X_0, Y_0, Z_0 \rangle$, e così via fino a che i punti X'_0, Y'_0, Z'_0

¹² È chiaro che il segno-prodotto permane scambiando X_0 con Y_0 et X con Y , mentre si inverte scambiando X_0 con Y_0 aut X con Y .

coincidono con X, Y, Z a meno di una permutazione, come nel caso considerato all'inizio. Se la permutazione in oggetto è pari si pone un ultimo (+) nel contatore, e un (-) nel caso contrario. Per definizione, $\langle X, Y, Z \rangle$ è **equivversa** o **antiversa** a $\langle X_0, Y_0, Z_0 \rangle$ a seconda che il prodotto dei quattro segni nel contatore è (+) o (-). Non è difficile proseguire analogamente nello spazio 3-dim, considerandovi una quaterna ordinata di punti distinti non coplanari $\langle X, Y, Z, U \rangle$, il cui orientamento è da confrontare con quello di una analoga quaterna di riferimento $\langle X_0, Y_0, Z_0, U_0 \rangle$. Escluso il caso banale, in cui è ormai ovvio come comportarsi, si sposta U_0 in uno dei punti X, Y, Z, U che non giacciono nel piano π passante per X_0, Y_0, Z_0 (ve ne è certamente uno), lo si chiama U'_0 e si pone un (+) o un (-) nel contatore con il solito criterio, ora riferito al piano π . Dopo aver similmente operato sugli altri punti Z_0, Y_0 e X_0 , si è ricondotti al caso banale con X, Y, Z, U coincidenti con X'_0, Y'_0, Z'_0, U'_0 a meno di una permutazione; si pone allora un ultimo segno (+) o (-) nel contatore a seconda che questa permutazione sia pari o dispari. Per definizione, $\langle X, Y, Z, U \rangle$ è **equivversa** o **antiversa** a $\langle X_0, Y_0, Z_0, U_0 \rangle$ a seconda che il prodotto dei cinque segni nel contatore sia (+) o (-). Si continua analogamente nel caso di uno spazio ($n > 3$)-dim .¹³

In conclusione, la nozione di equivversità/antiversità di una coppia di punti (distinti) di una retta rispetto ad una analoga coppia, di una terna di punti (distinti non collineari) di un piano rispetto ad una analoga terna, di una quaterna di punti (distinti non coplanari) di uno spazio 3-dim rispetto ad una analoga quaterna, ... e via dicendo, è delegata a quella di "essere dalla stessa parte": di due punti di una retta, di un piano, di uno spazio 3-dim, ...ecc., rispetto ad un suo punto, ad una sua retta, ad un suo piano, ...ecc. La relazione di *equivversità* di una $(n+1)$ -pla ordinata (di punti distinti e non $\text{co-}(n-1)$ -planari di H_n) rispetto ad un'analoga $(n+1)$ -pla ordinata risulta essere un'equivalenza, che decompone l'insieme di tali $(n+1)$ -ple ordinate in esattamente due classi di equivalenza (disgiunte) che si dicono **orientamenti** di H_n . La coppia costituita da H_n e da uno dei suoi due orientamenti (cioè da una $(n+1)$ -pla ordinata rappresentante di tale orientamento) si dice **spazio orientato di supporto** H_n . Questa rappresentante si sceglie di norma come l' $(n+1)$ -pla ordinata dei vertici di un $(n+1)$ -edro *regolare* unitario; oppure, come l' $(n+1)$ -pla ordinata $\langle O, U_1, \dots, U_n \rangle$, dove O è un punto-origine arbitrario, e $OU_1 \equiv e_1, \dots, OU_n \equiv e_n$ sono segmenti orientati unitari e mutuamente ortogonali. L' n -pla *ordinata* di questi segmenti $\langle e_1, \dots, e_n \rangle$ si dice allora una **n -pla coordinata cartesiana ortogonale ordinata di origine O** .¹⁴

¹³ Si può verificare che la definizione è consistente, cioè *non* dipende dalle arbitrarietà presenti nell'operazione complessiva. Naturalmente gli "o" del testo che precede sono esclusivi (aut).

¹⁴ È chiaro che "orientare H_n " equivale dunque ad assegnare una delle due possibili parità all' n -pla coordinata cartesiana ortogonale $\{e_1, \dots, e_n\}$ (con origine arbitraria). Da un punto di vista più generale ed astratto, l'orientamento di uno spazio lineare n -dim E_n consiste nello scegliere un ordine su una sua base $\{e_i\}_{i=1 \dots n}$ (se $n = 1$ questa operazione è vuota), considerando equivalenti due ordini della stessa parità. Si ottiene così una **base ordinata** $\langle e_i \rangle_{i=1 \dots n}$. Due basi

Si consideri in particolare l'insieme dei segmenti unitari orientati di origine O del piano H_2 orientato, e siano OX_1 e OX_2 due di essi. La nozione di **angolo con segno** φ (misura standard) dal segmento OX_1 al segmento OX_2 , nell'intervallo $[0, 2\pi)$, diventa allora univoca. Ciò si ottiene convenendo che sia $0 < \varphi < \pi$ [$\pi < \varphi < 2\pi$] se il triangolo orientato $\langle O, X_1, X_2 \rangle$ è equiverso [antiverso] al triangolo di riferimento; e inoltre, che sia $\varphi = 0$ se $X_2 = X_1$ e $\varphi = \pi$ se $OX_2 = -OX_1$. Viceversa, se è dato l'angolo con segno φ (in $[0, 2\pi)$ dal segmento OX_1 al segmento OX_2 , quest'ultimo è unicamente determinato. Prefissato allora un segmento orientato unitario **origine** OX_0 (avente per definizione angolo con segno zero), si stabilisce una corrispondenza biunivoca (\leftrightarrow) tra segmenti unitari orientati OX (o punti X del cerchio unitario di centro O) e angoli con segno in $[0, 2\pi)$, o equivalentemente in $\mathbf{R} \bmod 2\pi$. Rispetto a questa corrispondenza, il sopraddetto cerchio unitario e $\mathbf{R} \bmod 2\pi$ sono isomorfi come gruppi additivi abeliani se la somma di X_1 ($\leftrightarrow \varphi_1$) e di X_2 ($\leftrightarrow \varphi_2$) si *definisce* come $(X_1 + X_2) \leftrightarrow (\varphi_1 + \varphi_2)$. In un riferimento cartesiano ortogonale di origine O , con asse (orientato) x coincidente con la retta orientata OX_0 , e asse (orientato) y coincidente con la retta orientata OY per cui $Y \leftrightarrow \pi/2$, le coordinate di $X_{1,2}$ ($\leftrightarrow \varphi_{1,2}$) sono $x_{1,2} = \cos\varphi_{1,2}$ e $y_{1,2} = \sin\varphi_{1,2}$; allora le coordinate di $Z \equiv X_1 + X_2$ sono $x_Z = \cos(\varphi_1 + \varphi_2)$ e $y_Z = \sin(\varphi_1 + \varphi_2)$, e sono legate a quelle di $X_{1,2}$, cioè $\cos\varphi_{1,2}$ e $\sin\varphi_{1,2}$, dalle usuali formule di addizione per \cos e \sin , da usare ormai con $\varphi_{1,2}$ *arbitrari* in $\mathbf{R} \bmod 2\pi$.

La nozione di orientamento non è legata all'assioma (E), perché tutto quello che serve per definirla è la relazione “essere tra” di un punto rispetto ad altri due, e relative generalizzazioni (“essere tra” di una retta, “essere tra” di un piano, .. sempre rispetto a due punti non appartenenti a quella retta, a quel piano, ecc.). La disponibilità di (E) permette tuttavia di trasportare “per parallelismo” l'orientamento di un piano su un piano ad esso parallelo, esattamente come tale orientamento si trasporta per parallelismo da una retta ad una sua parallela. La relazione “il piano π è parallelo ed equiverso al piano ρ ” è una equivalenza e la classe di equivalenza relativa si dice **giacitura orientata**. Anche questa situazione è facilmente generalizzabile al caso di un generico spazio euclideo n -dim.¹⁵

ordinate si considerano **equiverse** (tra loro) se la matrice di transizione dall'una all'altra (non singolare) ha determinante positivo. L'equiorientamento tra basi è una relazione di equivalenza, ed esistono esattamente due classi (disgiunte) ad essa relative. Uno spazio lineare n -dim orientato è la coppia costituita da E_n e da una delle due classi di equivalenza delle basi.

¹⁵ Si comprende bene perché sia necessario il parallelismo delle varietà affini orientate di cui si vuole giudicare l'equi- o l'antiversità. In un piano, abbiamo soltanto il caso delle rette parallele; in uno spazio 3-dim quello delle rette e dei piani, e in generale in uno spazio ($n > 3$)-dim quello delle rette, dei piani, ... degli ($n-1$)-piani. Evidentemente, l'orientamento di una retta immersa in H_2 orientato, di un piano immerso in H_3 orientato, ... di un ($n-1$)-piano immerso in H_n orientato, può anche definirsi prescrivendo l'orientamento di una retta perpendicolare (o comunque trasversa) a quella retta, a quel piano, ... a quel ($n-1$)-piano.

La fondamentale nozione di **spostamento rigido** (o **rototraslazione**), in H_n , di una figura arbitraria di dimensione lineare n , è legata all'orientamento: oltre alla distanza tra due qualsiasi punti della figura, per definizione il suo spostamento rigido deve infatti preservarne anche l'orientamento, ossia l'orientamento di una $(n+1)$ -pla ordinata di suoi punti distinti e non co- $(n-1)$ -planari, che deve restare equiversa alla stessa $(n+1)$ -pla nella sua posizione iniziale. Una congruenza non è necessariamente uno spostamento rigido, mentre è vero il contrario.

Nel seguito, farà comodo denotare con \hat{H}_n uno dei due spazi orientati di supporto H_n , senza specificare esplicitamente il suo orientamento. Essendo strutturalmente più ricco di H_n , \hat{H}_n permette l'introduzione di nozioni/operazioni più specifiche che non H_n , e in conclusione di arricchirne l'algebra. Per cominciare, in \hat{H}_2 una semplice convenzione determina unicamente uno dei due semipiani aventi per bordo una data retta *orientata* p : basta prendere su p una coppia ordinata di suoi punti distinti $\langle X, Y \rangle$ equiversa a p , e considerare una qualunque terna ordinata $\langle X, Y, Z \rangle$ di punti distinti non collineari (di \hat{H}_2) equiversa alla terna ordinata di riferimento $\langle X_0, Y_0, Z_0 \rangle$ di \hat{H}_2 stesso. Allora Z fissa il semipiano in oggetto come suo rappresentante. Invertendo l'orientamento di p , si passa automaticamente al semipiano complementare.¹⁶ In modo del tutto simile, in \hat{H}_3 è unicamente determinato uno dei due semispazi aventi per bordo un dato piano *orientato* π : una quaterna ordinata di punti distinti non coplanari $\langle X, Y, Z, U \rangle$ equiversa alla quaterna di riferimento di \hat{H}_3 determina il semispazio in oggetto mediante il suo rappresentante U , posto che $\langle X, Y, Z \rangle$ appartenga a π e gli sia equiversa. Si potrebbe continuare in modo ovvio per spazi $H_{n>3}$.

Consideriamo ora un poligono (semplice) di \hat{H}_2 e una sua triangolazione i cui triangoli siano tutti equiversi [antiversi] al triangolo di riferimento (di \hat{H}_2). È immediato rendersi conto che i lati comuni di triangoli della triangolazione adiacenti risultano antiversi tra loro, mentre i lati dei triangoli che formano il contorno del poligono sono equiversi lungo di esso, e assegnano quindi un **orientamento** al contorno stesso. Un tale poligono si dice **orientato**, ed **equiverso** [antiverso] ad \hat{H}_2 . La situazione si estende facilmente ad un poliedro (semplice) di \hat{H}_3 decomposto in tetraedri tutti equiversi [antiversi] al tetraedro di riferimento. Le facce comuni e gli spigoli comuni di tetraedri adiacenti sono antiversi, e lo stesso si verifica per i lati comuni dei triangoli adiacenti che formano il contorno del poliedro, che ne riceve così un orientamento. Il poliedro si dice allora **poliedro orientato** ed **equiverso** [antiverso] ad \hat{H}_3 . Si potrebbe analogamente continuare passando a poliedri di $\hat{H}_{n>3}$.

Insieme ad altri, i concetti esposti in modo semiintuitivo nel precedente paragrafo sono alla base della (Teoria della) “Omologia” (o “Teoria Omologica”), una branca della Topologia che si

¹⁶ Dopotutto, l'inequivocità della indicazione stradale “voltare a sinistra” è legata al fatto che camminiamo sui piedi e non sulla testa, il che fissa l'orientamento del piano sul quale ci spostiamo. Inoltre, percorrendo la stessa strada in senso opposto la stessa indicazione ci porta nel semipiano complementare al precedente.

occupa appunto delle strutture algebriche associate alla topologia delle cosiddette “regioni geometriche astratte”. I pochi cenni che ne abbiamo riportato permettono comunque di stabilire un naturale e semplicissimo rapporto tra l’orientamento di un generico poliedro (che è Jordan-n-misurabile) di \hat{H}_n e il suo n-volume, nel senso che questo n-volume riceve un *segno*, positivo o negativo a seconda che il poliedro in oggetto sia equiverso o antiverso a \hat{H}_n . Un altro naturale collegamento si stabilisce poi tra l’orientamento di H_n e un corrispondente orientamento del suo modello \mathbf{R}^n . Ciò si ottiene convenendo di scegliere l’n-pla ordinata degli assi coordinati cartesiani ortogonali di origine O con cui si fissa la biiezione tra H_n e \mathbf{R}^n in modo che i punti U_i ($i = 1, \dots, n$) di coordinate $U_{ij} =: \delta_{ij}$ ($j = 1, \dots, n$) costituiscano con O una (n+1)-pla ordinata $\langle O, U_1, \dots, U_n \rangle$ equiversa a \hat{H}_n (cioè equiversa all’(n+1)-edro di riferimento). Questa convenzione fissa il *verso* degli assi coordinati (altrimenti arbitrario) – e quindi il segno delle coordinate del generico punto di \hat{H}_n – a meno di una permutazione pari. Con ciò anche \mathbf{R}^n diventa **orientato** (congruentemente a \hat{H}_n), e come tale può denotarsi $\hat{\mathbf{R}}^n$. Autonomamente rispetto a \hat{H}_n , **orientare \mathbf{R}^n** , cioè passare da \mathbf{R}^n a $\hat{\mathbf{R}}^n$, significa infatti scegliere una permutazione delle n+1 n-ple ordinate $\{0$ (zero di \mathbf{R}^n), $e_1, e_2, \dots, e_n\}$, dove $e_{1 \leq i \leq n}$ è l’n-pla ordinata $\langle 0, \dots, \delta_{ij}, \dots, 0 \rangle$ con δ_{ij} in posizione j-ma, come fondamentale, assegnandole quindi parità “pari”. (La famiglia $\{e_i\}_{1 \leq i \leq n}$ si dice la **base canonica ordinata di \mathbf{R}^n** .) Di regola si sceglie la permutazione $\langle 0, e_1, \dots, e_n \rangle$ come pari. È allora immediato comprendere che la procedura più sopra descritta col riferire \mathbf{R}^n a \hat{H}_n orienta \mathbf{R}^n in questo preciso senso. Tutto ciò vale anche per $n = 1$, e risulta allora che $\hat{\mathbf{R}}$ è l’usuale retta reale orientata dal – al +.

Se si pone O (origine del riferimento cartesiano ortogonale) in uno dei vertici del triangolo orientato di cui si vuole l’area con segno (in \hat{H}_2), del tetraedro orientato di cui si vuole il volume con segno (in \hat{H}_3), ... del (n+1)-edro orientato di cui si vuole l’n-volume con segno (in \hat{H}_n), si vede dunque che quell’area, quel volume, ... quell’n-volume, sono espressi, volta per volta, da una forma omogenea di grado 2, 3, ... n, e completamente antisimmetrica nelle 4, 9, ... n^2 coordinate dei rimanenti 2, 3, ... n vertici. Denotiamo con $Y_{i_1 i_2 \dots i_n}$ (qui i pedici i_1, i_2, \dots, i_n , stanno al solito per i_1, i_2, \dots, i_n , cfr. la Pres. 0.0.3) il **simbolo unitario completamente antisimmetrico di ordine n**, cioè la funzione numerica degli n pedici ordinati $\langle i_1, i_2, \dots, i_n \rangle$, variabili ciascuno sull’insieme numerico $\{1, 2, \dots, n\}$, il cui valore è +1 o –1 a seconda che $\langle i_1, \dots, i_n \rangle$ sia una permutazione pari o dispari di $\langle 1, 2, \dots, n \rangle$, e zero altrimenti. Per $n \geq 2$, la forma omogenea sopra descritta deve quindi esprimersi, a meno di un fattore di normalizzazione > 0 dipendente da n, come $\sum Y_{i_1 \dots i_n} x_{i_1}^1 x_{i_2}^2 \dots x_{i_n}^n$ (somma su i_1, \dots, i_n da 1 a n), dove x_j ($i, j = 1, \dots, n$) è la coordinata j-ma del vertice i-mo del (n+1)-edro in questione. È facile verificare che questa forma è invariante rispetto a una rotazione *propria* elementare del sistema di riferimento attorno all’origine O, e quindi rispetto ad un prodotto

arbitrario di tali rotazioni proprie elementari. Per provarlo è sufficiente considerare il caso $n = 2$, quando la forma in oggetto si riduce a ${}^1x_1 {}^2x_2 - {}^1x_2 {}^2x_1$. Si trova subito, allora, che sotto la rotazione propria elementare $x'_1 = x_1 \cos \varphi - x_2 \sin \varphi$, $x'_2 = x_1 \sin \varphi + x_2 \cos \varphi$ (che in questo caso è anche la più generale rotazione propria) risulta: ${}^1x'_1 {}^2x'_2 - {}^1x'_2 {}^2x'_1 \equiv {}^1x_1 {}^2x_2 - {}^1x_2 {}^2x_1$ indipendentemente da φ ; cioè, che per $n = 2$, la forma è invariante sotto una rotazione propria arbitraria. L'estensione a rotazioni proprie composte, per $n > 2$, è a questo punto banale.¹⁷

Nel caso di un triangolo, avendo posto l'origine O nel vertice (0) , e facendo passare l'asse x_1 per il vertice (1) , si ha ${}^1x_2 = 0$; mentre la relativa area con segno è evidentemente ${}^1x_1 {}^2x_2/2$.¹⁸ Tornando per rotazione propria ad un generico riferimento equiverso, si ha dunque per l'area \mathcal{A} con segno di un triangolo orientato con vertice (0) in O :

$$(1_2) \quad \mathcal{A} \equiv {}_2\mathcal{A} = (1/2) \sum \Upsilon_{ik} {}^1x_i {}^2x_k \quad (\text{somma su } i, k \text{ da } 1 \text{ a } 2).$$

Questo risultato mostra che il fattore di normalizzazione è $1/2$ per $n = 2$. Invertendo l'orientamento di \hat{H}_2 si inverte il segno di *una* delle coordinate di entrambi i vertici (1) e (2) . \mathcal{A} cambia allora segno; e lo stesso succede scambiando tra loro i vertici (1) e (2) (con il che è il triangolo che cambia orientamento).

In modo completamente analogo si tratta il calcolo del volume con segno ${}_3\mathcal{A}$ di un tetraedro orientato di \hat{H}_3 avente il vertice (0) in O . Il risultato, facilmente verificabile, è

$$(1_3) \quad {}_3\mathcal{A} = (1/3!) \sum \Upsilon_{ijk} {}^1x_i {}^2x_j {}^3x_k \quad (\text{somma su } i, j, k \text{ da } 1 \text{ a } 3),$$

e mostra che il fattore di normalizzazione è $1/6$ per $n = 3$. Infine il n -volume di un $(n+1)$ -edro orientato di \hat{H}_n avente il vertice (0) in O è

$$(1_n) \quad {}_n\mathcal{A} = (1/n!) \Upsilon_{i_1 \dots i_n} {}^1x_{i_1} \dots {}^n x_{i_n} \equiv {}_n\mathcal{V}/n! \quad (\text{somma su } i_1, \dots, i_n \text{ da } 1 \text{ a } n),$$

dove si è ormai abolito il segno \sum , bastando a denotare la sommatoria la ripetizione degli indici i_1, \dots, i_n ("convenzione di Einstein"), e che mostra un fattore di normalizzazione generale pari a $1/n!$. Il reale ${}_n\mathcal{V}$ fornito dalla (1_n) è il **n -volume** (o se si preferisce **n -area**, o **n -estensione**) **con segno dell' n -parallelogramma orientato di \hat{H}_n** (l'analogo del parallelogramma in $n > 2$ dimensioni, e quindi un particolare poliedro), con un vertice in O e n spigoli (ordinati) orientati uscenti da O aventi secondi estremi in ${}^1X, \dots, {}^nX$ (nell'ordine). Naturalmente il segno di ${}_n\mathcal{V}$ è $(+)$ o

¹⁷ Ovviamente, se la rotazione è di tipo generico (propria aut invertente) la forma è invariante sse il numero delle rotazioni elementari invertenti presenti nella composizione grupale è pari.

¹⁸ Se poi il vertice (1) è sul semiasse x_1 *positivo*, ${}^1x_1 = |{}^1X|$ e ${}^2x_2 = |{}^2X| \sin \theta$, dove iX ($i = 1, 2$) è il vertice i -mo e θ è l'angolo, in $(0, \pi)$, tra i lati orientati O^1X e O^2X . Quindi l'area (assoluta) del triangolo è $|A| = (1/2) |{}^1X| |{}^2X| \sin \theta$, secondo la ben nota formula elementare.

(-) a seconda che $\langle O, {}^1X, \dots, {}^nX \rangle$ sia equiversa o antiversa a \hat{H}_n . Si verifica anche che ${}_n\mathcal{V}$ non è altro che il determinante della $(n \times n)$ -matrice il cui elemento sulla riga i -ma e sulla colonna j -ma è ${}^i x_j$.¹⁹

Se – con le modalità a suo tempo definite – da H_n passiamo all'associato spazio dei Vettori H_n , e orientiamo entrambi in modo congruente, l' n -pla ordinata di punti $\langle {}^1X, \dots, {}^nX \rangle$ corrisponde all' n -pla ordinata di Vettori $\langle {}^1X, \dots, {}^nX \rangle$, e ${}^i x_j$ ($i, j = 1, \dots, n$) è la componente cartesiana ortogonale j -ma del Vettore ${}^i X$ (cioè quella che abbiamo denotato in corsivo minuscolo affetto da #). Per una data n -pla ordinata $\langle {}^1X, \dots, {}^nX \rangle$ di Vettori di H_n , il reale ${}_n\mathcal{V}$ della (1_n) (in cui si faccia ${}^i x_j \equiv {}^i x_{j\#}$) definisce, nel senso della $\langle {}^1X, \dots, {}^nX \rangle \mapsto {}_n\mathcal{V}$, un'applicazione di $(\mathbf{R}^n)^n$ in \mathbf{R} , n -lineare e completamente antisimmetrica, che si dice **prodotto misto** degli n Vettori considerati, presi nell'ordine, il cui valore (in \mathbf{R}) è il n -volume con segno del corrispondente n -parallelogramma orientato.²⁰

Il **prodotto-vettore** di una coppia ordinata $\langle X, Y \rangle$ di Vettori di \hat{H}_3 è invece il Vettore $Z =: X \times Y$ (spesso denotato con $X \wedge Y$ nella letteratura, in particolare italiana, di qualche anno fa) di componenti cartesiane ortogonali $z_i =: \epsilon_{ijk} x_j y_k$ ($i = 1, 2, 3$, somma da 1 a 3 sugli indici ripetuti), dove x_j e y_k sono le componenti cartesiane ortogonali di X e rispettivamente di Y . Questo Vettore è manifestamente ortogonale a entrambi i Vettori fattori (quindi al loro piano), ha modulo uguale all'area (assoluta) del parallelogramma costruito su X e Y (cioè del parallelogramma che ha un vertice in O e i vertici contigui in X per cui $\langle O, X \rangle \in X$ e in Y per cui $\langle O, Y \rangle \in Y$), e verso tale che la quaterna ordinata $\langle O, X, Y, Z \rangle$, con $\langle O, Z \rangle \in Z$, sia equiversa alla quaterna ordinata di riferimento di \hat{H}_3 . La sopradefinita operazione binaria \times , lineare rispetto al primo fattore e antisimmetrica, soddisfa l'**identità di Jacobi** (Karl, Potsdam Germ. 1804, Berlino 1851) $(X \times Y) \times U + (Y \times U) \times X + (U \times X) \times Y = 0$ (\equiv zero di \hat{H}_3) per qualunque terna di Vettori $\{X, Y, U\}$ di \hat{H}_3 .²¹ Per la stessa arbitraria terna si prova infine elementarmente che $(X \times Y) \times U = X \cdot UY - Y \cdot UX$. La definizione di prodotto-vettore si può estendere ad una $(n-1)$ -pla ordinata di Vettori di \hat{H}_n , ${}^2X, \dots, {}^nX$; il risultato è

¹⁹ Le considerazioni che portano alle $(1_2, 1_3, \dots, 1_n)$ sono ragionevoli suggerimenti piuttosto che vere dimostrazioni, ma possono rendersi completamente rigorose senza eccessive difficoltà, vedi S.sez. 5.1.2. Qualcuno chiama "n-blocco" l' n -parallelogramma orientato, mentre come già detto per l' $(n+1)$ -edro si usa più frequentemente il termine n -simpleso. Evidentemente, n è la dimensione lineare di un n -blocco e di un n -simpleso.

²⁰ Sebbene la nozione di antisimmetria vi degeneri, quanto precede continua a valere per $n = 1$: il 1-volume con segno ${}_1V \equiv {}_1\mathcal{V}$ (\equiv lunghezza L) dell'1-parallelepipedo (\equiv segmento) orientato di \hat{H}_1 è dato dalla $(1_{n=1})$ secondo $L = {}_1V = x$, dove x è la coordinata del 2° estremo del segmento orientato che ha il suo 1° estremo nell'origine di \hat{H}_1 . Questo risultato in sé banale e atteso conferma la correttezza della (1_n) anche per $n = 1$.

²¹ Lo spazio lineare \hat{H}_3 con in più il "prodotto" (non associativo) antisimmetrico e lineare rispetto al primo fattore \times è un esempio elementare e molto familiare di "algebra di Lie", vedi S.sez. 8.3.2. Una tale algebra si *definisce* aggiungendo ai consueti assiomi di spazio lineare appunto quelli di un "prodotto" (non associativo), antisimmetrico, lineare rispetto al primo fattore e soddisfacente all'identità di Jacobi. Vale a dire, denotando con giustapposizione tale prodotto, per ogni x, y, z dello spazio valgono le $xy + yx = 0$, (antisimmetria), $(\alpha x + \beta y)z = \alpha(xz) + \beta(yz)$ (linearità rispetto al primo fattore), e $(xy)z + (yz)x + (zx)y = 0$ (identità di Jacobi). Beninteso, dalla linearità rispetto al primo fattore segue subito quella rispetto al secondo, perché $z(\alpha x + \beta y) = -(\alpha x + \beta y)z = \alpha(-xz) + \beta(-yz) = \alpha(zx) + \beta(zy)$. Il carattere non associativo di \times si vede ad esempio perché $x \times (y \times z) = (x \cdot z)y - (x \cdot y)z$, mentre $(x \times y) \times z = -[z \times (x \times y)] = -(z \cdot y)x + (z \cdot x)y$, e in generale $(x \cdot y)z \neq (z \cdot y)x$.

ancora un Vettore dello stesso \hat{H}_n , ortogonale a tutti i Vettori fattori, secondo la $Z_i = Y_{i_1 i_2 \dots i_n} x_{i_1} x_{i_2} \dots x_{i_n}$ ($i = 1, \dots, n$, somma da 1 a n sugli indici ripetuti i_1, \dots, i_n).

La nozione di orientamento dello spazio *empirico* (o di un piano *empirico*) si deve considerare primitiva, nel senso che necessita di un “campione di orientamento” (come del resto quella di distanza necessita di un campione di lunghezza) per essere definita operativamente. Questo campione di orientamento può essere la mano destra o la sinistra, il campo magnetico generato dal moto (di data direzione e verso) di cariche elettriche positive o negative, uno zucchero come il destrosio o il levulosio, e comunque *uno di due oggetti enantiomorfi*.²²

Nello spazio empirico orientato, il prodotto-vettore di un segmento orientato uscente da un'origine O^* , diciamo OX^* ²³ per un simile segmento orientato OY^* (in quest'ordine) è definito come segue. Supposti OX^* e OY^* non collineari, il prodotto-vettore del primo per il secondo è il segmento orientato OZ^* di modulo pari all'area (assoluta) del parallelogramma costruito su OX^* e OY^* , perpendicolare al loro piano, e verso tale che la quaterna ordinata di punti empirici $\langle O^*, X^*, Y^*, Z^* \rangle$ sia equiversa a quella di riferimento. Se poi OX^* e OY^* sono collineari, oppure se $O^* = X^*$ e/o $O^* = Y^*$, la definizione si completa ponendo $Z^* = O^*$. È evidente il carattere assoluto di questa definizione, una volta che sia assegnato l'orientamento dello spazio empirico e che l'unità di area sia quella del quadrato di lato unitario. Nella biiezione esistente tra \hat{H}_3 e lo spazio empirico orientato, le due definizioni di prodotto-vettore si corrispondono se i due spazi sono equiversi, cioè se il tetraedro di riferimento di \hat{H}_3 è l'immagine di quello dello spazio empirico. Di norma, per quest'ultimo si prende il tetraedro mimato dal palmo (l'origine) della mano destra, e dalle estremità del suo pollice, indice e medio presi in quest'ordine (**orientamento standard dello spazio empirico**).²⁴

²² Sperimentiamo visivamente uno spazio isometrico ma antiverso rispetto a quello reale in cui siamo immersi quando guardiamo in uno specchio l'immagine di un oggetto la cui vista “diretta” ci è familiare (quindi in particolare *non* del nostro viso): le distanze, la profondità e l'altezza sono invariate, mentre la destra e la sinistra sono scambiate tra loro. Ma in generale, la diversità dell'immagine di un oggetto generico rinviataci dallo specchio, rispetto a quella abituale, non ci colpisce più di un tanto. Un caso tipico in cui il rovesciamento dello spazio reale ci sorprende è invece quello in cui vediamo *noi stessi* doppiamente riflessi da due specchi angolati (come li hanno ad esempio i sarti) e quindi ci vediamo esattamente come ci vede il nostro prossimo; oppure quello in cui cerchiamo (da neofiti) di inquadrare un oggetto con una vecchia macchina fotografica reflex priva di pentaprisma (o di prisma di Porro).

²³ Qui l'asterisco rammenta che stiamo considerando oggetti empirici, cfr S.sez. 1.1.2.

²⁴ È chiaro che questa definizione presuppone che si attribuisca carattere assoluto alla nostra percezione psicofisica di “mano destra”, di “pollice”, ecc.

1.4.3) CONSIDERAZIONI CONCLUSIVE

Possiamo ormai avviarcì alla conclusione di questo primo capitolo. La assiomatizzazione della geometria euclidea n -dim ha messo in luce la notevole ricchezza della struttura di \hat{H}_n e del suo modello $\hat{\mathbf{R}}^n$. Fondamentale, in questi sviluppi, si è rivelato il ruolo del modello $\hat{\mathbf{R}}^n$, spesso vero e proprio “apripista euristico” verso la crescita strutturale di \hat{H}_n , innanzitutto in senso algebrico. Vogliamo in particolare riconsiderare in quest’ottica le trasformazioni affini (non singolari) di coordinate cartesiane generalmente oblique introdotte alla fine della S.sez. 1.2.4 (v. (1.2.4, 1)). L’insieme di queste trasformazioni forma evidentemente un gruppo, che si dice **gruppo affine** e che si denoterà qui con Ω . Sotto una trasformazione di Ω , una retta di $H_n \approx \mathbf{R}^n$ ²⁵ si trasforma in una retta, un piano in un piano, ... un $(n-1)$ -piano in un $(n-1)$ -piano, e $H_n \approx \mathbf{R}^n$ in una sua copia generalmente traslata e distorta linearmente, cioè in una sua “copia affine” di identico supporto. È proprio per questa ragione che le rette, i piani, ... gli $(n-1)$ -piani di $H_n \approx \mathbf{R}^n$ si dicono sue **(sotto)varietà affini**.²⁶ Se poi H_n e \mathbf{R}^n vengono congruentemente orientati (cioè se $\hat{H}_n \approx \hat{\mathbf{R}}^n$), allora la generica trasformazione affine trasforma $H_n \approx \mathbf{R}^n$ in una sua copia affine equiversa o antiversa all’originale a seconda che il determinante della trasformazione sia positivo o negativo. Le trasformazioni affini ortogonali (cfr. 1.3.1, 9) formano un sottogruppo di Ω (**gruppo affine ortogonale**) che denoteremo qui con Ω_{\perp} . Un sistema di coordinate cartesiane ortogonali normalizzate è trasformato in un sistema analogo dalle trasformazioni affini ortogonali, e $\hat{H}_n \approx \hat{\mathbf{R}}^n$ diventa allora una sua copia generalmente traslata e ruotata, equiversa o antiversa all’originale a seconda che il determinante della trasformazione sia uguale a $+1$ o a -1 . Inoltre sotto tali trasformazioni resta invariata la distanza (d in H_n e $*d$ in \mathbf{R}^n , quindi $d \equiv *d$), e più in generale il prodotto interno tra due punti di H_n o tra due n -ple ordinate di \mathbf{R}^n . Questa isometria, assicurata dalle trasformazioni affini ortogonali, non è tuttavia strettamente necessaria per preservare quei caratteri delle figure geometriche che interessano *specificamente* la geometria euclidea: una più generale similitudine basta allo scopo. Il gruppo delle similitudini è un ampliamento banale di quello affine ortogonale, e si ottiene da esso semplicemente moltiplicando per una (unica) costante positiva σ^2

²⁵ Indichiamo qui con \approx l’isomorfismo tra H_n e \mathbf{R}^n , e poco più sotto tra \hat{H}_n e $\hat{\mathbf{R}}^n$, senza menzionare esplicitamente l’origine delle coordinate in H_n .

²⁶ Le “porzioni finite” di dato contorno di queste sottovarietà affini di $H_n \approx \mathbf{R}^n$ sono caratterizzate dall’aver la minima estensione tra tutte le analoghe porzioni di sottovarietà della stessa dimensione e con lo stesso contorno. Così un segmento di estremi dati è la linea di lunghezza minima tra quegli estremi. Teoremi di questo tipo sono naturalmente di grande importanza, ma presuppongono la conoscenza della teoria differenziale di queste sottovarietà immerse, vedi soprattutto la Sez. 5.3. Opportunamente modificata e generalizzata, la proprietà sopramenzionata continua a valere per sottovarietà affini di spazi pseudoeuclidei.

(con $\sigma > 0$) i simboli di Kronecker a 2° membro della (1.2.4, 2). In questo modo la distanza trasformata è pari a σ volte quella originale, e quindi σ uguaglia il fattore di similitudine.

Il gruppo affine ortogonale Ω_{\perp} può considerarsi come “prodotto diretto” (nel senso standard della teoria dei gruppi) del gruppo delle traslazioni, del gruppo delle rotazioni proprie, e di quello della “identità-inversione”.²⁷ In $H_3 \approx \mathbf{R}^3$, il **gruppo delle rototraslazioni proprie**, diciamo Ω_{\perp}^+ (sottogruppo di Ω_{\perp}), risulta essere a sei parametri *continui*: i nove elementi della matrice della trasformazione, legati dai sei vincoli di ortonormalità (restano tre parametri liberi) più i tre parametri della traslazione. Passando a $H_n \approx \mathbf{R}^n$, il numero dei parametri continui di Ω_{\perp}^+ diventa $n^2 - n(n+1)/2 + n = n(n+1)/2$. Il gruppo Ω_{\perp} è sostanzialmente più generale di Ω_{\perp}^+ perché le possibili inversioni, presenti in Ω_{\perp} ma non in Ω_{\perp}^+ , *non* sono continue; quindi Ω_{\perp} non può dirsi a parametri continui (o almeno a parametri tutti continui).

Sotto una rototraslazione (propria o invertente che sia) le componenti cartesiane ortogonali di un Vettore di H_n subiscono la trasformazione *omogenea* associata; le sue nuove componenti si dicono trasformate “per cogredienza” delle vecchie. Questa materia sarà trattata nel Cap 3. Tra l’altro, si vedrà che l’“algebra vettoriale” che si istituisce su $\hat{H}_n \approx \hat{\mathbf{R}}^n$ (somma, moltiplicazione per un reale, prodotto interno, prodotto esterno) si può estendere ad una “algebra tensoriale” analoga, ma sostanzialmente più ricca della prima. La definizione dei tensori di generico “ordine” $\kappa \geq 0$ e della loro algebra si ottiene in termini di forme “ κ -lineari” su \mathbf{R} , o “ κ -forme”; oppure, equivalentemente, per mezzo di convenienti assiomi in base ai quali lo spazio dei tensori e quello delle forme multilineari sono canonicamente isomorfi rispetto alle loro algebre.

Come il gruppo delle similitudini caratterizza la geometria euclidea, il gruppo affine analogamente caratterizza la **geometria affine**, che può essere definita da questa circostanza. Vale a dire, mentre la geometria euclidea si occupa delle proprietà delle figure geometriche che non sono alterate dalle trasformazioni del gruppo delle similitudini, per definizione quella affine si occupa delle loro proprietà che non sono alterate dalle più generali trasformazioni del gruppo affine. Una ovvia conseguenza di ciò è che le proprietà (delle figure) pertinenti alla geometria affine sono *meno* stringenti di quelle pertinenti alla geometria euclidea. Proprietà che non sono alterate dalle trasformazioni affini includono ad esempio il parallelismo (di rette, piani ecc.), ma non gli angoli liberi o i rapporti tra distanze. Proseguendo nella stessa direzione, si potrebbe anche considerare il cosiddetto **gruppo proiettivo completo** di trasformazioni tra coordinate, a sua volta più ampio di quello affine: le proprietà che non sono alterate da trasformazioni del gruppo proiettivo sono per

²⁷ Il gruppo della identità-inversione è isomorfo al più semplice gruppo di ordine finito, essendo costituito da un solo elemento \mathcal{J} oltre all’unità \mathcal{I} , e risulta automaticamente abeliano. La sua tabella moltiplicativa è completamente descritta dalla $\mathcal{J} \circ \mathcal{J} = \mathcal{I}$, perché le altre tre composizioni $\mathcal{I} \circ \mathcal{I} = \mathcal{I}$ e $\mathcal{I} \circ \mathcal{J} = \mathcal{J} \circ \mathcal{I} = \mathcal{J}$ sono conseguenze banali della definizione di unità. Dalla $\mathcal{J} \circ \mathcal{J} = \mathcal{I}$ segue anche che \mathcal{J} è inverso di se stesso. L’elemento \mathcal{J} può ben dirsi “inversione”.

l'appunto quelle pertinenti alla **geometria proiettiva**. Questo modo di vedere le geometrie fu proposto esplicitamente per la prima volta da F. Klein (Felix, Düsseldorf Germ. 1849, Gottinga 1925) nel suo famoso "Programma di Erlangen" (1872),²⁸ e a partire dal tardo Ottocento il suo nucleo più astratto ha progressivamente assunto sempre maggiore importanza nello sviluppo non solo della geometria, ma della stessa fisica. Va da sé che, come avviene per la geometria euclidea (o per la geometria assoluta), queste geometrie più deboli si possono descrivere come (convenienti) sistemi assiomatici.

Ci fermiamo qui con il nostro sintetico sguardo d'assieme allo spazio euclideo, anche se il quadro che ne emerge *non è* quello che sarebbe stato opportuno (o inderogabile) se lo si fosse programmato come obiettivo a sé stante.²⁹ La conclusione fondamentale, lo ripetiamo, è la seguente: si è stabilita una corrispondenza biunivoca tra $\hat{H}_3 \approx \hat{\mathbf{R}}^3$ e lo spazio fisico orientato (convenientemente idealizzato) della esperienza comune, rispetto alla quale essi risultano isomorfi, nel senso che gli enunciati veri del primo corrispondono uno ad uno a verità empiriche del secondo, e viceversa, entro le solite approssimazioni osservative. Rileviamo che ciò è stato ottenuto facendo ricorso da una parte alla sola nozione primitiva di distanza d tra due punti [$*d$ tra due terne ordinate] di \hat{H}_3 [di $\hat{\mathbf{R}}^3$], e dall'altra al solo protocollo operativo per la misura della corrispondente distanza empirica tra due punti dello spazio fisico (oltre alla nozione primitiva di punto di \hat{H}_3 e alla definizione operativa di punto empirico); e in ultima analisi, in accordo con la già menzionata penetrante intuizione di B. Riemann, vedi la S.sez 1.2.2.

Chiudiamo con una considerazione che ci è suggerita dalla nota classificazione delle strutture matematiche proposta da Bourbaki nel quarto/quinto decennio del secolo scorso. Secondo questa classificazione, le strutture matematiche si distinguono in algebriche, topologiche, di ordine, e loro ibridi. Attraverso gli assiomi che traducono le verità empiriche idealizzate dello spazio fisico scelte a fondamento della sua rappresentazione formale \hat{H}_3 , e il provato isomorfismo tra \hat{H}_3 e \mathbf{R}^3 (v. Sez. 1.3), \hat{H}_3 riceve appunto una struttura algebrica e una struttura topologica (quest'ultima essenzialmente attraverso la nozione di prodotto interno). Quanto all'orientamento, esso è evidentemente una struttura d'ordine. Il fatto che strutture dei tre tipi bourbakiani siano già tutte presenti in \hat{H}_3 , e quindi nello spazio fisico idealmente euclideo, rende ben ragionevole – in realtà quasi banale – l'idea che in quest'ultimo spazio l'invenzione matematica abbia trovato una fonte di ispirazione fondamentale.

²⁸ Nelle parole di Klein: «Per una data varietà, ed un dato gruppo di trasformazioni su di essa, si studieranno le forme appartenenti a quella varietà relativamente alle loro proprietà che si mantengono inalterate attraverso le trasformazioni del gruppo dato».

²⁹ Ulteriori informazioni ancora riferibili allo spazio euclideo si troveranno nei capitoli che seguono e nelle appendici generali.

APP. SPEC. CAP. 1

.APP. 1.A LE RELAZIONI D'ORDINE

Ripassiamo qui alcune nozioni standard sulle relazioni d'ordine. Sia ρ una relazione binaria (che nel seguito scriveremo “tra”) sull'insieme con uguaglianza X (cioè definita per ogni coppia ordinata di suoi elementi), e siano x,y,z,\dots elementi arbitrari di X . La relazione ρ si dice:

- (1) **transitiva** se $x\rho y \wedge y\rho z \Rightarrow x\rho z$;
- (2) **riflessiva** se $x=y \Rightarrow x\rho y$;
- (3) **antisimmetrica** se $x\rho y \wedge y\rho x \Rightarrow x=y$;
- (4) **differenziante** se $x\rho y \Rightarrow x \neq y$;
- (5) **asimmetrica** se $x\rho y \Rightarrow \neg y\rho x$;

Alcune di queste proprietà possono eventualmente coesistere, nel qual caso si hanno le seguenti denominazioni per ρ :

- (6) ρ è un **ordine** se valgono le (1,2,3);
- (7) ρ è un **ordine stretto** se valgono le (1,4,5).

Seguono alcune proposizioni.

- (a) «se ρ è un ordine, la relazione (binaria, su X) σ definita dalla $x\sigma y \Leftrightarrow: x\rho y \wedge x \neq y$, è un ordine stretto», che si dice **l'associato all'ordine** ρ ;
- (b) «se σ è un ordine stretto, la relazione (binaria, su X) definita da $x\rho y \Leftrightarrow: x\sigma y \vee x=y$, è un ordine», che si dice **l'associato all'ordine stretto** σ ;
- (c) «l'ordine ρ' associato all'ordine stretto associato all'ordine ρ equivale a ρ , $\rho' \Leftrightarrow \rho$ »;
- (d) «l'ordine stretto σ' associato all'ordine associato all'ordine stretto σ equivale a σ , $\sigma' \Leftrightarrow \sigma$ ».

Nel caso di un ordine ρ , due elementi x,y di X si dicono **confrontabili** se $x\rho y \vee y\rho x$ è comunque vera; nel caso di un ordine stretto σ , essi si dicono **strettamente confrontabili** se $(x=y) \vee x\sigma y \vee y\sigma x$ è comunque vera. Un ordine [un ordine stretto] si dice **totale** o **lineare** se tutte le coppie di elementi di X sono confrontabili [strettamente confrontabili]. Per brevità, scriveremo “t-ordine” per “ordine totale” e “s-ordine” (cioè ordine semplice) per **ordine stretto totale**. Si verifica subito che valgono proposizioni analoghe alle $(a \div c)$, diciamo $(a' \div c')$, per i t-ordini e gli s-ordini.

Nel seguito, denoteremo con \leq un generico t-ordine e con $<$ l's-ordine associato, e viceversa denoteremo con $<$ un generico s-ordine e con \leq il t-ordine associato. Sia ora $(X, <)$ un insieme

s-ordinato: un suo elemento x_0 per il quale $\forall(y)\{(y \in X \setminus \{x_0\}) \Rightarrow (x_0 < y)\}$ [per il quale $\forall(y)\{(y \in X \setminus \{x_0\}) \Rightarrow (y < x_0)\}$] si dice « il **primo** » [l'**ultimo**] elemento di X . La definizione è consistente perché, se esiste, tale x_0 risulta unico. $(X, <)$ si dice **<-denso** se per ogni coppia di suoi elementi x, y con $x < y$ risulta $\{z \in X | x < z < y\} \neq \emptyset$; e un suo sottoinsieme proprio Y si dice **<-denso in X** se per ogni tale coppia x, y risulta $\{z \in X | x < z < y\} \cap Y \neq \emptyset$. (Si noti che quest'ultima definizione è simile a, ma diversa da, quella che si dà di “ Y denso in X ” quando X è uno spazio topologico, vedi App. Gen. C.) Una retta, una semiretta, un segmento di qualunque tipo (tutti s-ordinati con $<$) sono **<-densi**. Sempre con riferimento a $(X, <)$, si consideri ora una coppia $\{X_1, X_2\}$ di sottoinsiemi di X non vuoti e per i quali

(i): $X = X_1 \cup X_2$ e

(ii): $\forall(x_1, x_2)\{x_1 \in X_1 \wedge x_2 \in X_2 \Rightarrow x_1 < x_2\}$.

Tale coppia (di insiemi necessariamente disgiunti) $\{X_1, X_2\}$ si dice una **separazione di X** , e può soddisfare una delle quattro proprietà mutuamente escludentisi

(α) “né X_1 ha l'ultimo elemento né X_2 ha il primo”;

(α') “ X_1 ha l'ultimo elemento e X_2 non ha il primo”;

(α'') “ X_1 non ha l'ultimo elemento e X_2 ha il primo”;

(α''') “ X_1 ha l'ultimo elemento e X_2 ha il primo”.

Ovviamente, la disgiunzione (esclusiva) di ($\alpha, \alpha', \alpha'', \alpha'''$) è comunque vera. Se una separazione di X soddisfa (α), essa si dice una **lacuna di X** . In ciascuno degli altri tre casi si dice una **sezione di X** ; e in particolare se si verifica ($\alpha' \vee \alpha''$) si dice una **sezione di Dedekind** (o **D-sezione**) di X , e l'elemento $z \in X$ per cui $\forall(x_1 \in X_1, x_2 \in X_2)\{x_1 \leq z < x_2 \vee x_1 < z \leq x_2\}$ si dice il **separatore** della D-sezione. Se per $z' \in X$, $z < z'$ [$z' < z$], allora $z' \in X_2$ [$z' \in X_1$], perché l'ipotesi opposta $z' \in X_1$ [$z' \in X_2$] condurrebbe ad una contraddizione. Se $(X, <)$ è **<-denso**, si verifica facilmente che si deve escludere il caso (α'''), e quindi una sezione è sempre una D-sezione; vale a dire, se $(X, <)$ è **<-denso** una separazione di X è una sua lacuna aut una sua D-sezione. $(X, <)$ si dice avere la **proprietà di Dedekind** (o **D-proprietà**) se ogni sua separazione è una sua D-sezione. In luogo di $x < y$ [$x \leq y$] si scrive anche, equivalentemente, $y > x$ [$y \geq x$]. Non vi è affatto uniformità, in letteratura, sulle precedenti definizioni. Ad esempio, spesso quello che qui abbiamo chiamato “ordine” viene detto “ordine parziale” (in inglese è di uso comune il parziale acronimo “poset” [“toset”] per “partially [totally] ordered set”). Altri dicono “ordine parziale” una relazione binaria transitiva e riflessiva, altri ancora dicono quest'ultima “pre-ordine”, ecc.

APP. 1.B SULLE FUNZIONI Cos E Sin

Avendovi fissato O nel solito ruolo di origine, nel piano ρ di H_n si consideri il raggio $p' =: SR(O|O^*)$, il raggio $q' =: SR(O|X)$, e il raggio $r' =: SR(O|Y)$, supponendo X e Y distinti e nello stesso semipiano ρ' di bordo p (la retta includente p'), e sotto i vincoli $d(O,O^*) = d(O,X) = d(O,Y) = 1$. Poniamo $\mathfrak{A} =: [p'|q']$ e $\mathfrak{B} =: [q'|r']$; allora $\mathfrak{A} + \mathfrak{B}$ esiste, e la $\mathfrak{A} + \mathfrak{B} = [s'|p']$ determina unicamente il raggio s' di origine O dello stesso semipiano ρ' , e quindi anche $Z \in s'$ sotto il vincolo $d(O,Z) = 1$. Alla luce della definizione di congruenza tra angoli, risulta:

$$(1_1) \quad d(Z,Y) = d(X,O^*) \text{ se } [r',q',p'], \text{ e}$$

$$(1_2) \quad d(Z,X) = d(Y,O^*) \text{ se } [q',r',p'].$$

Poiché trattiamo in modo simmetrico q' e r' , cioè X e Y , questi due casi *non* sono essenzialmente diversi, e potremo riferirci soltanto a uno di essi, diciamo a quello della (1₁). Si prova allora facilmente la relazione:

$$(2) \quad OX_p = OY_p OZ_p + d(Y,Y_p)d(Z,Z_p),$$

ove OX_p, OY_p, OZ_p sono reali con segno, mentre $d(Y,Y_p)$ e $d(Z,Z_p)$ sono > 0 , e possono eliminarsi in favore di OY_p, OZ_p mediante la

$$(3) \quad d(U,U_p) = (1 - (OU_p)^2)^{1/2},$$

valida per il generico U di ρ . In conformità con quest'ultima, è utile introdurre la funzione $\text{Sin} =: (1 - (\text{Cos})^2)^{1/2}$ (per la funzione Cos , vedi S.sez. 1.3.2) definita in $(0,\pi)$ e ivi con valori in $(0,1]$; quindi, fruendo anche dei valori limite di Cos (essendo banalmente Sin funzione continua di Cos), anche in $[0,\pi]$, e ivi con valori in $[0,1]$. A questo punto la (2) si traduce in termini delle funzioni Cos e Sin , e precisamente nella forma

$$(4) \quad \text{Cosa} = \text{Cos}(a+b)\text{Cos}b + \text{Sin}(a+b)\text{Sin}b,$$

avendo convenuto, in questa appendice, di scrivere a, b, \dots in luogo di $|\mathfrak{A}|, |\mathfrak{B}|, \dots$ per maggior convenienza tipografica. Posto $c =: a + b$ (risulta quindi $\mathfrak{A} < \mathfrak{C}, \mathfrak{B} < \mathfrak{C}$), ed eliminando a , la (4) si trascrive come

$$(5) \quad \text{Cos}(c - b) = \text{Cos}c\text{Cos}b + \text{Sin}c\text{Sin}b \quad (\text{sotto } b < c),$$

che diremo la **formula di sottrazione per Cos**. Se invece si elimina b , la (4) diventa $\text{Cosa} = \text{Cos}(c-a)\text{Cos}c + \text{Sin}(c-a)\text{Sin}c$; se in questa si esprime $\text{Cos}(c-a)$ mediante la (5), essendo $\text{Sin}c > 0$ si ha:

$$(6) \quad \text{Sin}(c-a) = \text{Sin}c\text{Cosa} - \text{Sin}a\text{Cos}c \quad (\text{sotto } a < c),$$

cioè la **formula di sottrazione per Sin**.

È ora la volta delle **formule di addizione** (per Sin e Cos). Ancora con $b = c - a$, dalla (6) si ha:

$$(7) \quad \text{SinaCos}(a+b) = \text{Sin}(a+b)\text{Cosa} - \text{Sin}b:$$

sostituendo questa nella (4) moltiplicata per Sina (> 0), risulta:

$$(8) \quad \text{Sin}(a+b)(\text{CosaCos}b + \text{SinaSin}b) = \text{SinaCosa} + \text{Sin}b\text{Cos}b.$$

Qui $\text{Sin}(a+b)$ è ovviamente > 0 ; il fattore di $\text{Sin}(a+b)$ a 1° membro e il 2° membro si annullano quindi insieme sse $a - b = \pi/2$ ($b < a$) aut $b - a = \pi/2$ ($a < b$). Escludendo questo caso eccezionale, e moltiplicando i due membri della (8) per $\text{SinaCos}b + \text{CosaSin}b$ si trova:

$$(9) \quad \text{Sin}(a+b) = \text{SinaCos}b + \text{CosaSin}b.$$

Sostituendo questa nella (7), e dividendo per Sina (> 0), si ha infine:

$$(10) \quad \text{Cos}(a+b) = \text{CosaCos}b - \text{SinaSin}b.$$

La (9) e la (10) sono tuttavia sub *judice* nel caso eccezionale, e per sciogliere ogni dubbio occorre provarne la validità anche in questo caso. Posto dunque $a = \pi/2 + b$ (quindi $b < \pi/4$), si ha, per quanto riguarda la (9), $\text{Sin}(\pi/2+2b) = \text{Cos}(2b)$ ¹ per il suo 1° membro, e $(\text{Cos}b)^2 - (\text{Sin}b)^2 \equiv \text{Cos}(2b)$ per il 2°: cioè, la (9) è valida anche nel caso eccezionale. Analogamente passando all'esame della (10) nel caso eccezionale, si ha, per il suo 1° membro, $\text{Cos}(\pi/2+2b) \equiv -\text{Sin}(2b)$ (vedi la nota ⁽¹⁾), e $-2\text{Sin}b\text{Cos}b \equiv -\text{Sin}(2b)$: anche la (10) è valida dunque senza eccezioni. La (9) e la (10) sono la **formula di addizione per Sin** e rispettivamente la **formula di addizione per Cos**. Si sottolinea che la dimostrazione della loro validità senza eccezioni ha richiesto il ricorso *diretto* alla teoria H_n (l'asserto "due triangoli rettangoli con l'ipotenusa della stessa lunghezza e un angolo uguale sono uguali" è un teorema di H_n), secondo quanto è stato illustrato nella nota ⁽¹⁾.

Torniamo ancora alla (4), che è la fonte, direttamente connessa a teoremi della teoria H_n , dei precedenti risultati. Come sappiamo Cos (e quindi anche Sin) è una funzione continua, e nulla vieta di *supporla* anche derivabile nell'intervallo aperto $(0,\pi)$. Operando per il momento in modo formale, deriviamo dunque la (4) rispetto a b . Denotando le derivate con un apice, ponendo ancora $c = a + b$ ed eliminando a , con semplici passaggi si trova (Sin b e Sin c sono ovviamente positivi):

$$(11) \quad (\text{Cos}'b/\text{Sin}b - \text{Cos}'c/\text{Sin}c)(\text{Cos}b\text{Sin}c - \text{Cos}c\text{Sin}b) = 0 \quad (\text{sotto } b < c).$$

La (11) è stata qui ricavata per $b < c$, ma è chiaro che vale per b e c qualsiasi in $(0,\pi)$. Il secondo fattore è manifestamente diverso da zero per $b \neq c$, e quindi

$$(12) \quad (\text{Cos}'b/\text{Sin}b - \text{Cos}'c/\text{Sin}c) = 0$$

¹ Che sia $\text{Sin}(\pi/2+d) = \text{Cos}d$ per $d < \pi/2$ si verifica subito mediante la $d(U',U'_p) = OU_p$, ove $U \leftrightarrow d$ e $U' \leftrightarrow \pi/2 + d$ sul cerchio unitario, in base all'uguaglianza dei triangoli di vertici ordinati $\langle U_p, U, O \rangle$ e $\langle U'_p, O, U' \rangle$. In modo simile si prova che $\text{Cos}(\pi/2+d) = -\text{Sin}d$. Naturalmente queste due relazioni sono in accordo con la (9) e rispettivamente con la (10); ponendo infatti $a = \pi/2$ nella (9) [nella (10)] si ha $\text{Sin}(\pi/2+b) = \text{Sin}(\pi/2)\text{Cos}b + \text{Cos}(\pi/2)\text{Sin}b = \text{Cos}b$ [si ha $\text{Cos}(\pi/2+b) = \text{Cos}(\pi/2)\text{Cos}b - \text{Sin}(\pi/2)\text{Sin}b = -\text{Sin}b$].

per $b \neq c$, che è come dire che $\text{Cos}'b/\text{Sin}b$ è una costante universale, diciamo K ; cioè, non indicando esplicitamente l'argomento di Cos e Sin ,

$$(13) \quad \text{Cos}' = K\text{Sin}.$$

Quindi la derivata di Cos *esiste* ed è proporzionale a Sin secondo una costante universale. La (13) può iterarsi, perché $\text{Sin}' = -K\text{Cos}$ scende dalla (13) e dalla $\text{Cos}\text{Cos}' + \text{Sin}\text{Sin}' = 0$. Si ottiene così $\text{Cos}'' = -K^2\text{Cos}$, $\text{Cos}''' = -K^3\text{Sin}$, $\text{Cos}'''' = K^4\text{Cos}$, ... e così via. È dunque possibile sviluppare $\text{Cos}(\pi/2+h)$ in serie di Taylor (B. Taylor, 1685-1731) per $h \in (-\pi/2, \pi/2)$. Il risultato è evidentemente:

$$(14) \quad \text{Cos}(\pi/2+h) = Kh - (Kh)^3/3! + (Kh)^5/5! - \dots \equiv \sin Kh.$$

In modo simile si trova che

$$(15) \quad \text{Sin}(\pi/2+h) = 1 - (Kh)^2/2! + (Kh)^4/4! - \dots \equiv \cos Kh,$$

per h nello stesso intervallo $(-\pi/2, \pi/2)$. Il valore di K si desume facilmente imponendo le condizioni ai limiti. Basta fare $h \rightarrow \pi/2$ nella (14): si ha $-1 = \sin(K\pi/2)$, cioè $K = -1, +3, -5, \dots$ (dalla (15) si ricaverebbe invece la condizione meno forte $|K| = 1, 3, \dots$). D'altra parte è evidente che non può essere $|K| \neq 1$: infatti, è $\text{Sin}(\pi/2+h) > 0$, per h in $(-\pi/2, \pi/2)$, mentre $\cos Kh$ non può essere > 0 nello stesso intervallo per $|K| = 3, 5 \dots$. In definitiva, $K = -1$, e $\text{Cos}(\pi/2+h) = -\sin h$, $\text{Sin}(\pi/2+h) = \cos h$; ossia, tenendo conto delle due formule ricavate nella nota ⁽¹⁾, $\text{Cos} = \cos$ e $\text{Sin} = \sin$ nell'intervallo $[0, \pi]$.

CAP. 2

LA GEOMETRIA PSEUDOEUCLEIDEA ¹

2.1) PREMessa: LO SPAZIO-TEMPO DI EUCLIDE-NEWTON

2.1.1) ELEMENTI DI CINEMATICA EUCLIDEA-NEWTONIANA

Passiamo ora a considerare un aspetto completamente nuovo e diverso – rispetto al quadro che ne abbiamo delineato nel primo capitolo – della geometria fisica. Oltre che in una “dimensione spaziale”, ci rendiamo conto di essere immersi in una “dimensione temporale” percepita come ben distinta dalla prima; ma è chiaro che, posto in questi termini, il problema compete più alla fisiologia e alla psicologia della percezione che alla fisica. Il punto di vista del fisico tende piuttosto ad essere connaturalmente operazionalista: i concetti evolvono dai fatti osservati, e quindi sono connessi ai protocolli operativi – messi a punto per via induttiva – che governano l’osservazione, e non il contrario. Per cominciare, il fisico riconosce quindi che quello che abbiamo fin qui descritto come spazio fisico tridimensionale “atemporale” è piuttosto una successione di sue copie ciascuna delle quali etichettata da una quarta coordinata che chiama “tempo”. Essa si misura mediante uno strumento ideale *non localizzato* detto **orologio**, o più precisamente **orologio normale**, con cui si misura il **tempo normale**.

Un orologio (normale) è un sistema fisico che percorre un “ciclo di stati”, munito di un contatore dei cicli modificabile in senso traslativo, e ragionevolmente sottratto ad ogni influenza esterna che possa alterarne il “regolare ritmo di marcia”. ² Da un punto di vista concettuale, quest’ultima proprietà è analoga alla rigidità dei regoli con cui si misurano le distanze, e come

¹ N.B. Per la grande frequenza con cui tale simbolo si presenta, da questo capitolo in avanti rinunciamo a denotare **R** in grassetto, passando a **R**.

² Un orologio generico si distingue da un orologio normale anche in quanto non necessariamente isolato da azioni esterne: esso è cioè soltanto un misuratore di tempo “non-normale”, ovvero di un tempo funzione generica, ma continua e monotona, del tempo normale. È chiaro, tuttavia, che abbiamo a che fare con nozioni *primitive*, e che la sostanza pratica della distinzione tra orologi normali e non-normali sta nell’espressione “regolare ritmo di marcia” (non a caso virgolettata) che abbiamo usato più sopra. Nel seguito del capitolo ci riferiremo soltanto a orologi normali e al tempo normale, nominandoli semplicemente come “orologi” e “tempo”. Orologi generici verranno presi in considerazione nel Cap. 9, nell’ambito della teoria della relatività generale.

quella deve essere accolta come essenzialmente primitiva. Inoltre incontriamo un problema analogo a quello che ci metteva in difficoltà quando dovevamo raffigurarci un punto (geometrico) ideale come limite di una piccola regione spaziale; vale a dire, dobbiamo anche in questo caso rinunciare all'iter definitorio “intervallo temporale” → “istante” – come a suo tempo abbiamo rinunciato all'analogo iter “regione spaziale” → “punto” – in favore di quello opposto. Va poi da sé che gli **istanti**, controparte dei punti, che possiamo così considerare, sono in realtà intervalli temporali con durata di ordine inferiore a quella del ciclo dell'orologio. Ad esempio, per un orologio atomico questa durata è l'inverso della frequenza di una ben definita riga spettrale, la quale è tipicamente di 10^{10} o 10^{11} s^{-1} .

Ovviamente un orologio reale è localizzato; la possibilità di de-localizzarlo, rendendolo spazialmente pervasivo, è in realtà un pregiudizio acriticamente accettato fino alla fine del XIX secolo, e dovuto alla apparente facilità con cui orologi identici collocati in punti distinti dello spazio possono essere “sincronizzati” (cioè i loro contatori possono essere regolati in modo che segnino lo stesso numero di cicli – ad es. zero – in un comune istante). Seguendo quest'ordine di idee, il tempo diventa un parametro che etichetta tutto lo spazio istante per istante, e la sua coordinatazione naturale è quella dei “tic” di un *unico* orologio delocalizzato, naturalmente orientato dal “passato” al “futuro”.³ Avendo fissato per esso una certa origine e una certa unità di misura multipla del ciclo dell'orologio, il tempo è quindi in corrispondenza biunivoca con l'insieme degli interi letti nel suo contatore – o meglio dei razionali con denominatore uguale al numero di cicli contenuti nell'unità di tempo prescelta. Come in quello della coordinatazione dello spazio, anche in questo caso è vantaggioso – o meglio essenziale – passare dall'insieme dei razionali a quello dei reali \mathbb{R} , pensando ad un orologio con ciclo di durata tendente a zero. In questo modo lo spazio-tempo diventa isomorfo allo spazio-prodotto $\mathbb{R}^3 \times \mathbb{R}$. Dotando \mathbb{R} della sua metrica standard (oltre che del suo ordine standard) lo spazio-tempo $\mathbb{R}^3 \times \mathbb{R}$ può essere a sua volta dotato, in quanto spazio-prodotto di spazi metrici, di una metrica: ad esempio della metrica pitagorica $(d_s^2 + d_t^2)^{1/2}$, dove d_s e d_t sono la distanza spaziale e rispettivamente la distanza temporale tra coppie di suoi elementi, oppure della metrica $\sup(d_s, d_t)$, ecc.⁴ Il generico elemento dello spazio-tempo si dirà **punto-istante**, o (talvolta) **evento**⁵; quindi punto-istante sta per la coppia (punto, istante).

³ Abbiamo cautelativamente virgolettato anche questi sostantivi perché per il momento non possiamo affidarci che ad una interpretazione intuitiva del loro significato.

⁴ Come tuttavia vedremo più oltre, salvo situazioni particolari la possibilità di metrizzare in tal modo lo spazio-tempo non presenta grande interesse.

⁵ Il significato di “evento”, nel linguaggio corrente, è piuttosto quello di “accadimento”, di “qualcosa che avviene” nel dato punto-istante. Così ad es. il “presente” è l'insieme dei punti-istanti presenti, ma anche, nel linguaggio corrente, l'insieme degli accadimenti nei punti-istanti presenti. Da un punto di vista operazionalista, è tuttavia ovvio che un dato punto-istante – pensato come la quaterna ordinata delle sue coordinate – può essere identificato soltanto dallo specifico

In linea di principio, possiamo pensare di collocare stabilmente in ogni punto dello spazio un “registratore” (puntiforme) capace di conservare traccia, al modo di una moderna telecamera nascosta, degli eventi (accadimenti) che occorrono in quel punto in ogni istante successivo (si ricordi che il tempo è orientato) alla sua installazione. In questo modo, sia pur per ricognizione indiretta, potremmo conoscere gli eventi passati, in quel punto e almeno a partire da un certo istante; ma non abbiamo alcuna analoga possibilità di conoscere gli eventi futuri. Questo stabilisce una asimmetria lungo la direzione del tempo, che i fisici chiamano gergalmente **freccia del tempo**, avendola peraltro definita in modo molto diverso. Se poi non avessimo predisposto la rete rigida di registratori locali, anche la conoscenza (di parte) degli eventi passati ci sarebbe preclusa. Tradotto in termini psicologici, questo insieme di fatti ci rende coscienti della inaccessibilità, alla osservazione *diretta*, di tutti gli eventi salvo quelli “presenti”, in contrasto con l’evidente accessibilità di principio di tutti i punti dello spazio. In altre parole, mentre l’osservatore può spostare liberamente i suoi strumenti, per fare esperimenti in questo o quel punto dello spazio, una analoga libertà gli è preclusa quando si tratta di raggiungere punti-istanti con tempi *diversi* da quello presente, e in particolare punti-istanti co-locali ma non-presenti. Soltanto il “presente” (\equiv insieme degli eventi presenti) è direttamente accessibile: possiamo osservare il mondo, in particolare facendovi esperimenti, soltanto “ora”, non domani o ieri, come invece lo possiamo osservare facendo esperimenti “qui” o “lì”. Il “passato” – cioè l’insieme degli eventi passati – possiamo ricostruirlo e descriverlo (se e solo se è stato registrato dalla rete rigida di registratori localizzati, e con l’approssimazione del documento di cui disponiamo); mentre il “futuro” – l’insieme degli eventi futuri – possiamo (tentare di) prevederlo, progettarlo, influenzarlo senza alcuna certezza di successo, e naturalmente, piacendo al nostro destino, attenderlo. Insomma, il tempo scorre per suo conto dal cosiddetto e inaccessibile passato al cosiddetto e inaccessibile futuro, attraversando continuamente il cosiddetto e accessibile, ma fuggevole, presente.⁶ Il fatto infine che il mondo fisico ci proponga innumerevoli tipi di fenomeni tempo-irreversibili, e che la freccia del tempo possa correlarsi nel più naturale dei modi a quella tempo-irreversibilità, esula dal presente approccio alla geometria spazio-temporale.

Nello schema classico che abbiamo delineato, spazio e tempo sono “separati”: come si diceva, il tempo non è altro che un parametro t che etichetta copie identiche dello spazio (e a pensarci un momento questo è proprio il significato di prodotto cartesiano spazio \times tempo). Tuttavia spazio e tempo si “coniugano” nella nozione di **movimento** (o **moto**) di un punto empirico.

accadimento che vi si verifica; esattamente come un punto dello spazio, pensato come la terna ordinata delle sue coordinate, può essere identificato soltanto dallo specifico “cartello indicatore” che vi si colloca.

⁶ Qualcuno ha detto che se da una parte non possiamo (o forse non possiamo ancora, si veda il contributo di Gödel a questo problema) “viaggiare attraverso il tempo”, dall’altra è il tempo che “viaggia attraverso di noi”, con continua e inesorabile regolarità.

Precisamente, diremo che un punto empirico “si muove” nello spazio se, avendolo in qualche modo contrassegnato con una “etichetta” P , al variare di t esso viene a trovarsi in punti diversi delle copie di spazio etichettate mediante t . La traiettoria di P nello spazio-tempo $R^3 \times R$ si dice sua **linea d’universo**. Il presente modello di questo fenomeno, se R^3 è il modello dello spazio empirico e R è il modello del tempo empirico, è dunque quello di una terna ordinata ξ di coordinate spaziali (ad es. cartesiane ortogonali x, y, z) funzioni di t , date per t in un conveniente intervallo-base I (possibilmente, $I = (-\infty, \infty)$) e supposte convenientemente regolari. Siano ora $\xi_1 \equiv (x_1, y_1, z_1)$ la posizione di P al tempo t_1 , e $\xi_2 \equiv (x_2, y_2, z_2)$ la sua posizione al tempo $t_2 \neq t_1$, entrambi nell’intervallo-base I , che converrà pensare come aperto. Passiamo dalla coppia ordinata (ξ_1, ξ_2) al vettore $\Delta\xi =: \xi_2 - \xi_1$, lo **spostamento** di P dal tempo t_1 al tempo $t_2 = t_1 + \Delta t$, e facciamo il prodotto di $\Delta\xi$ per $1/\Delta t$. Supponendo che in R^3 esista il limite (nelle topologie standard di R^3 e di R) di $\Delta\xi/\Delta t$ al tendere di t_2 a t_1 , cioè la derivata $d\xi/dt$ di ξ al tempo $t = t_1$, chiameremo **velocità di P** (a quel tempo t) questo vettore limite, e lo denoteremo $v(t) \equiv d\xi/dt(t)$. In modo analogo potremo procedere considerando, invece delle posizioni, le velocità di P al variare di t , ottenendo un corrispondente vettore limite $a(t) \equiv dv/dt(t)$, o **accelerazione di P** al tempo t . Affinché la velocità [l’accelerazione] di P sia continua in I , come supporremo, la sua posizione dovrà essere espressa come terna ordinata di funzioni ivi di classe C^1 [C^2]. Come ognuno sa, e come giustificheremo fra un momento, non occorre spingersi oltre con analoghe operazioni di differenziazione della traiettoria ai fini della istituzione di una “dinamica” di P . Queste sono le nozioni più elementari della **cinematica classica del punto**.

Alla cinematica classica appartiene anche lo studio dei precedenti concetti a fronte di trasformazioni *continue*, con parametro t , della coordinatazione cartesiana ortogonale. Ciò corrisponde al movimento di un sistema di coordinate rispetto all’altro. Poiché le inversioni non sono continue, dovremo limitarci alle rototraslazioni proprie. Sarà comodo denotare con \mathfrak{S} il riferimento (cartesiano ortogonale) da considerare convenzionalmente come “fisso”, e con \mathfrak{S}' quello da considerare come “mobile” rispetto al primo. La posizione di \mathfrak{S}' rispetto a \mathfrak{S} è unicamente determinata assegnando sei parametri (posizione dell’origine di \mathfrak{S}' – tre parametri – e orientamento di \mathfrak{S}' – altri tre parametri – ad es. i tre angoli di Eulero) come funzioni abbastanza regolari di t nell’intervallo-base I . Il moto di un punto P rispetto a \mathfrak{S}' (che si può dire suo moto “relativo”), quando sia noto quello rispetto a \mathfrak{S} (che si può dire suo moto “assoluto”), è allora a sua volta unicamente determinato, e il suo calcolo esplicito è l’oggetto della **cinematica relativa del punto**. L’App. Spec. 2.A offre un riassunto di tale cinematica, le cui grandi linee si presumono peraltro note al lettore. Ricordiamo qui che la **velocità assoluta** (rispetto a \mathfrak{S}) di P risulta uguale alla somma

vettoriale della sua **velocità relativa** (quella rispetto a \mathcal{S}') e della cosiddetta sua **velocità di trascinamento**, cioè quella che P avrebbe se fosse immobile rispetto a \mathcal{S}' . Le cose sono un po' più complicate per quanto riguarda l'accelerazione: l'**accelerazione assoluta** (rispetto a \mathcal{S}) di P è uguale alla somma della sua **accelerazione relativa** (quella rispetto a \mathcal{S}'), della sua **accelerazione complementare** (o **di Coriolis** (Gustave, 1792-1843)), pari al doppio del prodotto vettoriale della **velocità angolare** (o **spin**: per brevità useremo anche questo termine, improprio in senso stretto) di \mathcal{S}' rispetto a \mathcal{S} per la velocità relativa di P, e infine della sua **accelerazione di trascinamento** – ancora, quella che P avrebbe se fosse immobile rispetto ad \mathcal{S}' . Se tuttavia la rototraslazione del sistema mobile si riduce ad una traslazione con velocità costante (o “traslazione uniforme”), allora sia l'accelerazione di Coriolis che quella di trascinamento risultano nulle, e si è ridotti all'eguaglianza tra l'accelerazione assoluta e quella relativa.

Come esempio fondamentale di traslazione uniforme, si immagini che il riferimento \mathcal{S}' trasli lungo l'asse X di \mathcal{S} (qui e nel seguito si denotano con le stesse lettere senza apice o con apice oggetti omologhi relativi a \mathcal{S} o rispettivamente a \mathcal{S}') con velocità costante (valore e segno) Y , in modo che i piani coordinati (x',y') e (x',z') di \mathcal{S}' “scivolino” lungo i piani omologhi (x,y) e (x,z) di \mathcal{S} , e che le origini dei due riferimenti coincidano al tempo $t = t' = 0$. Allora le leggi di trasformazione dalle coordinate (x,y,z,t) di \mathcal{S} alle coordinate (x',y',z',t') di \mathcal{S}' sono evidentemente le

$$(1_1) \quad x' = x - Yt, \quad y' = y, \quad z' = z,$$

cui si aggiunge la

$$(1_2) \quad t' = t,$$

ovvero l'ipotesi che il tempo sia delocalizzato.

Pur restando la notazione delle (1) quella più flessibile e di generale riferimento (anche nel seguito, in altre simili situazioni), sarà spesso opportuno alleggerirla sostituendo *all'assenza di apice il carattere maiuscolo del simbolo, e alla presenza dell'apice il suo carattere minuscolo*. Confidiamo che il lettore imparerà facilmente ad orientarsi con sicurezza tra le due notazioni, la seconda delle quali sarà di regola la preferita finché possibile. Proseguiremo dunque la sezione in notazione “minuscolo-maiuscolo”, cominciando col riscrivere le (1) come

$$(1_{1bis}) \quad x = X - YT, \quad y = Y, \quad z = Z,$$

$$(1_{2bis}) \quad t = T.$$

Le $(1 \equiv 1_{bis})$ rappresentano una **traslazione uniforme speciale**, caratterizzata da un unico parametro costante Y ; per $Y \neq 0$, essa è la più semplice trasformazione tra coordinate di due riferimenti cartesiani ortogonali in \mathbb{R}^3 , a parte l'identità cui si riduce per $Y = 0$. Evidentemente, queste traslazioni uniformi speciali costituiscono un gruppo commutativo continuo ad un parametro.

Derivando le (1₁) rispetto a $t \equiv T$, e denotando con (v_x, v_y, v_z) le $(dx/dt, dy/dt, dz/dt)$ e con (V_X, V_Y, V_Z) le $(dX/dT, dY/dT, dZ/dT)$ – le velocità di P in s e rispettivamente in \mathfrak{S} –, otteniamo

$$(2) \quad v_x = V_X - \Upsilon, \quad v_y = V_Y, \quad v_z = V_Z.$$

($v \equiv (v_x, v_y, v_z)$ è la velocità relativa di P, e $V \equiv (V_X, V_Y, V_Z)$ la sua velocità di trascinamento.)

Derivando ancora, si ha, per le accelerazioni di P nei due riferimenti,

$$(3) \quad a \equiv (a_x, a_y, a_z) = A \equiv (A_X, A_Y, A_Z),$$

che mostra appunto l'uguaglianza delle due accelerazioni, relativa e assoluta, di P nel caso considerato. Le (1) si dicono **trasformazioni galileiane speciali**, e sono un caso particolare delle **trasformazioni galileiane omogenee**, cioè delle traslazioni uniformi con velocità (costante) generica più una rotazione propria “una tantum”, e con coincidenza delle origini di \mathfrak{s} e \mathfrak{S} al tempo $t = T = 0$.

Anche queste più generali trasformazioni formano un gruppo continuo, adesso a sei parametri costanti: le tre componenti della velocità di traslazione di s e ad es. i tre angoli di Eulero della sua rotazione una tantum. È immediato vedere che la velocità V di P rispetto a \mathfrak{S} è ancora uguale alla somma vettoriale della sua velocità v rispetto a \mathfrak{s} , o velocità relativa, e della sua velocità di trascinamento, pari alla velocità di traslazione di \mathfrak{s} rispetto a \mathfrak{S} ; e che le due accelerazioni di P sono uguali.

Le più generali trasformazioni che fanno passare da un riferimento ad un altro della corrispondente classe di equivalenza formano ancora un gruppo continuo, ogni elemento del quale è caratterizzato, oltre che dalla velocità di traslazione di \mathfrak{s} rispetto a \mathfrak{S} (tre parametri) e dalla rotazione una tantum (altri tre parametri), da una traslazione una tantum delle origini di tutte e quattro le coordinate (altri quattro parametri): gruppo quindi a dieci parametri (costanti), che è detto **gruppo delle traslazioni uniformi**.

La relazione “ \mathfrak{s} è in traslazione uniforme rispetto a \mathfrak{S} ” è manifestamente una relazione di equivalenza tra \mathfrak{s} e \mathfrak{S} . Se si rinuncia sia alla rotazione una tantum che alla traslazione una tantum delle quattro origini (inessenziali dal presente punto di vista), si è ridotti al **gruppo delle traslazioni uniformi parallele-omogenee**, a soli tre parametri, in cui la terna di versori coordinati di \mathfrak{s} si muove con velocità costante rispetto a quella di \mathfrak{S} *restando ad essa parallela*, e le origini delle due terne *coincidono* al tempo $t = T = 0$. Di esse, le traslazioni galileiane speciali sono un caso particolare.

2.1.2) ELEMENTI DI DINAMICA NEWTONIANA

La nostra presentazione della cinematica classica (o newtoniana-euclidea) deve a questo punto sostanzialmente allargarsi in una direzione che a prima vista sembra allontanarci da essa, introducendoci nella categoria dei fenomeni “dinamici”. Tra le classi di equivalenza, chiamiamole \mathcal{E} , dei riferimenti in traslazione uniforme l’uno rispetto all’altro, ve n’è una che diremo **classe inerziale** e denoteremo \mathcal{I} . Evidentemente, l’accelerazione di un punto cinematico è invariante all’interno di una generica classe di equivalenza \mathcal{E} , e in particolare all’interno della classe inerziale \mathcal{I} ; in questo caso l’accelerazione si dirà **strettamente assoluta** (ma per brevità ancora soltanto “assoluta”). D’altra parte, incluso nell’insieme di tutte le traiettorie $\xi = \xi(t)$ di punti cinematici rispetto ad un riferimento generico, esiste un sottoinsieme, che diremo insieme delle **traiettorie dinamiche** (o **traiettorie di punti materiali**), per le quali l’esperienza e l’induzione suggeriscono la disponibilità di un modello \mathcal{M} che fornisce il valore dell’accelerazione assoluta del punto materiale P a meno di un fattore strettamente positivo *dependente soltanto dalla identità* di P . Il reciproco di questo fattore si dice **massa inerziale** (o semplicemente **massa**) di P , e sarà denotato per maggior chiarezza e per il momento con m_P piuttosto che con m .

Più precisamente \mathcal{M} fornisce il vettore accelerazione assoluta di P al tempo t come una funzione f di t , della posizione ξ (al tempo t) e della velocità v (al tempo t) di P , moltiplicata per il fattore $1/m_P$. Nel seguito scriveremo x in luogo di ξ , confidando che il lettore non confonda tale x con la prima componente di ξ , che abbiamo appunto denotato con x . I valori di (t,x,v) sono quelli valutati in un riferimento della classe \mathcal{I} , e quindi il valore di f deve essere anch’esso invariante all’interno della classe \mathcal{I} affinché tutto lo schema funzioni. Poiché (t,x,v) sono a priori generici, \mathcal{M} deve fornire $f = f(t,x,v)$ in un conveniente aperto \mathcal{D} dello spazio delle sue sette indeterminate scalari (t,x,v) . Il vettore $f(t,x,v)$, definito in \mathcal{D} , si dice **forza agente su P** (che si trova in x e nello “stato di moto” v), e viene qui assunto continuo rispetto ai suoi argomenti insieme alle sue derivate parziali rispetto alle x e alle v . Sulla base di \mathcal{M} , in un qualunque riferimento della classe \mathcal{I} vale dunque (in \mathcal{D}) l’equazione differenziale ordinaria del 2° ordine, in forma normale (\equiv risolta rispetto alla derivata di ordine massimo), nella funzione $x = x(t)$ (**traiettoria di P** nel riferimento in oggetto):

$$(1) \quad d^2x/dt^2(t) \equiv d_t^2 x(t) = f(t,x,dx/dt)/m_P.$$

Poiché per definizione m_P non varia con t , la (1) può equivalentemente scriversi, con $v =: d_t x$, come $d_t(m_P v) = f$. Il vettore $m_P v$ si dice, con Cartesio, **quantità di moto** (QdM) di P . È chiaro che la QdM si “conserva” (\equiv non varia con t) se $f \equiv 0$.⁷

⁷ Nella letteratura di lingua inglese, la quantità di moto è detta “momentum”, talvolta tradotto in italiano con “momento”. Ciò può creare confusione perché questo termine ha anche un altro significato nella nostra lingua. È

Una elementare conseguenza scalare della equazione vettoriale (1) si ottiene moltiplicandola scalarmente per v :

$$(1') \quad d_t(m_P v^2/2) = v \cdot f.$$

La quantità $m_P v^2/2$ (dimensionalmente, il prodotto di una forza per una lunghezza, cioè una energia) si dice **energia cinetica** di P. Ovviamente anche l'energia cinetica si conserva se $f \equiv 0$.⁸

Secondo la (1), se due punti materiali di masse diverse sono soggetti alla stessa forza, essi subiscono accelerazioni inversamente proporzionali alle loro masse. Se la costante positiva m_P nella (1) è nota, per la data f la (1) determina unicamente $x(t)$, di classe C^2 , assegnando i valori x_0 e $(d_t x)_0$ di x , e rispettivamente di $(d_t x)$, al tempo t_0 , sotto la condizione $(t_0, x_0, (d_t x)_0) \in \mathcal{D}$, per t in un intorno di t_0 . Vi è tuttavia un caso in cui la (1), e quindi il modello \mathcal{M} assieme alle due condizioni iniziali, fornisce la traiettoria di P *anche non conoscendo* la massa m_P , ed è quello per cui $f \equiv 0$ in \mathcal{D} : allora $x(t) = (d_t x)_0(t-t_0) + x_0$ (per t nel solito intorno di t_0). Questo risultato si dice **legge d'inerzia** (o **prima legge di Newton** (Isaac, Woolsthorpe Ingh. 1642, Londra 1727)), ed afferma che P si muove di moto uniforme con velocità uguale alla sua velocità iniziale, intorno alla sua posizione iniziale, per un certo tempo intorno al tempo iniziale.⁹

opportuna la seguente breve digressione di “terminologia comparata”. In italiano, “momento” è usato i) nel senso di “momento di un cursore F applicato in P, rispetto ad un polo O”, uguale a $(P-O) \times F$, cui corrisponde il “moment” inglese; ii) in dinamica lagrangiana, nel senso di variabile canonica della seconda serie, cioè di “momento cinetico” o anche “momento generalizzato” (ma spesso semplicemente “momento”), che in inglese è ancora “momentum”, o “generalized momentum”. (In effetti, nel caso di un unico punto e di coordinate cartesiane ortogonali, il momento generalizzato e la QdM coincidono). Il “momento della QdM” è reso in inglese come “angular momentum” più spesso che come “moment of momentum”, ed è anche comunemente tradotto in italiano (specie dai fisici) come “momento angolare”. Come si vede, nonostante l'elementarità dei concetti in gioco c'è qualche motivo di disorientamento per il neofita.

⁸ Sottolineiamo che in generale la forza in \mathcal{D} non è data necessariamente da un *unico* modello. Forze fornite da modelli diversi si considerano uguali se, applicate allo stesso generico punto materiale, gli imprimono la stessa accelerazione. Questo permette di verificare la reciproca consistenza tra modelli diversi, e induce talvolta a privilegiare aprioristicamente, sul piano di una esposizione teoretica, un modello sugli altri: ad esempio, in quest'ordine di idee una forza può *definirsi* una volta per tutte come la deformazione relativa che subisce una “molla campione” ad essa soggetta (a meno di un conveniente fattore dipendente dalle unità di misura). La presente definizione è più flessibile e più aderente alla realtà: il punto materiale in questione può essere di fatto soggetto ad una forza elastica (\equiv prodotta da una molla), ma anche ad una forza gravitazionale, ad una forza elettrica, ecc., variando di volta in volta il modello che fornisce quella forza. In una accezione più astratta e generale, il modello addirittura *identifica* il sistema dinamico di interesse; ad esempio, nella dinamica analitica classica la funzione hamiltoniana si può agli effetti pratici *identificare* con l'oggetto di cui si vuole studiare il “moto” o evoluzione (risolvendo le equazioni canoniche sotto le convenienti condizioni accessorie). Segnaliamo infine, per la sua importanza, il modello dell'accelerazione di P secondo il quale $f = -\nabla\varphi$, ove φ è una funzione data della sola posizione x di P. In questo caso, moltiplicando scalarmente (\cdot) la (1) per $m_P d_t x \equiv m_P v$ risulta $0 = m_P v \cdot d_t v + v \cdot \nabla\varphi = d_t(m_P v^2/2 + \varphi)$. La quantità $m_P v^2/2 + \varphi$ è dunque una costante rispetto a t , o un “invariante del moto” che si dice **energia meccanica di P**, somma della sua energia cinetica $m_P v^2/2$ e della sua **energia potenziale** φ .

⁹ Se in particolare la velocità iniziale è nulla, $(d_t x)_0 = 0$, allora $x(t) = x_0$; cioè P permane nella sua posizione iniziale. In questa accezione, la legge d'inerzia è accessibile all'osservazione più grossolana, ed era già nota ad Aristotele. La legge d'inerzia nella sua formulazione completa è invece il frutto di una elaborazione secolare, e fu stabilita con sufficiente chiarezza soltanto da Galileo (ma forse fu già intuita da Leonardo (Vinci It. 1452, Cloux Fr. 1519) e da Copernico (Mikolaj, Toruń Pol. 1473, Frauenberg Pol. 1543)). Nelle parole di Newton (Philosophiæ Naturalis Principia Mathematica, 1687), «Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directum, nisi

Ovviamente questo fatto vale in qualunque riferimento della classe \mathcal{I} e di essa soltanto. Pertanto la classe \mathcal{I} potrebbe definirsi come quella in cui vale la legge d'inerzia. È allora evidente che, ai fini di una fondazione assiomatica della dinamica newtoniana, occorre un assioma del tipo “esiste un riferimento inerziale”. Tuttavia un siffatto asserto, non costruttivo, non sarebbe accettabile da parte di un fisico. Infatti un fisico pretende comunque asserti di esistenza, se non costruttivi, almeno “ostensivi”; in compenso, possibilmente li accetta anche se sono “approssimati”. Una tale “ostensione approssimativa” era sostanzialmente disponibile già qualche tempo dopo la teoria gravitazionale di Newton, e avrebbe potuto recitare: «è inerziale un sistema di riferimento avente la sua origine nel baricentro di un corpo estremamente massivo (ad esempio del sole) ed orientato verso le cosiddette “stelle fisse”.» (Le stelle fisse sono oggetti materiali che ci appaiono come fissi in forza della loro presunta estrema lontananza). Sono evidenti, in questa definizione, gli elementi di approssimazione (il corpo di riferimento deve essere “estremamente” massivo, le stelle fisse devono essere “estremamente” lontane, nonché quelli di circolarità (la lontananza delle stelle fisse è “presunta”, ecc.) Curiosamente, anche molto tempo dopo Newton non fu disponibile una definizione ostensiva sostanzialmente migliore di riferimento inerziale.

La determinazione di m_p pone un problema sperimentale e fondazionale abbastanza delicato se si resta nell'ambito del modello \mathcal{M} . Ovviamente le quantità x_0 , $(d_t x)_0$, $(d_t^2 x)_0$, $(d_t^3 x)_0$, ... sono tutte accessibili all'osservazione diretta, seguendo la traiettoria di P in un intorno di t_0 di ordine 0, 1, 2, 3 ... Se quindi si misurasse l'accelerazione iniziale di P, $(d_t^2 x)_0$, trovandola diversa da 0, la (1) stessa fornirebbe m_p come $f(t_0, x_0, (d_t x)_0) / (d_t^2 x)_0$. Sfortunatamente la misura di $(d_t^2 x)_0$ può essere ardua per un moto generico; essa è molto più semplice per un **moto uniformemente accelerato** (\equiv ad accelerazione costante nel tempo.) Basta dunque considerare la possibilità di usare un modello \mathcal{M}^* , se esiste, che fornisca una forza in pratica indipendente da (t, x, v) . Detta f^* questa forza costante agente su P, nota in base al modello \mathcal{M}^* , secondo la (1) la corrispondente traiettoria di P è $x(t) = f^*(t-t_0)^2 / (2m_p) + (d_t x)_0(t-t_0) + x_0$. Conoscendo la forza costante f^* , la massa m_p si ricava così mediante una (unica) misura della posizione di P (o della sua velocità) ad un tempo prefissato t non troppo discosto da t_0 .^{10, 11}

quatenus a viribus impressis cogitur statum illum mutare». Evidentemente, le linee d'universo di un punto in moto inerziale sono rette dello spazio-tempo classico.

¹⁰ Un siffatto modello \mathcal{M}^* , che fornisce una forza praticamente costante in un intorno non troppo esteso di un qualunque punto x_0 della superficie terrestre, indipendentemente da t e dalla velocità del punto materiale considerato, è quello che ne dà il peso p , verticale e diretto verso il basso. La corrispondente accelerazione, in virtù della (1) anch'essa costante, verticale e diretta verso il basso, è l'**accelerazione di gravità** g (che risulta uguale a circa $9,81 \text{ m/s}^2$); si conclude così che $m_p = p/g$.

¹¹ La formula del moto uniformemente accelerato si semplifica ponendo in essa, senza sostanziali limitazioni di generalità, $x_0 = 0$, $t_0 = 0$ e $(d_t x)_0 = 0$ (l'ultima condizione si soddisfa passando ad un altro riferimento inerziale, in moto traslatorio uniforme con velocità $(d_t x)_0$ rispetto al primo). Si ottiene così la traiettoria rettilinea $x(t) = f^* t^2 / (2m_p)$, con velocità (di direzione costante e lineare in t) $d_t x(t) = f^* t / m_p$. In questo caso la misura di m_p è anche più semplice,

L'eq. (1), con m_p determinata in questo o analogo modo, si dice **equazione fondamentale della dinamica** (o **seconda legge di Newton**)¹²; la legge d'inerzia si può dunque considerare come un suo semplice corollario. È importante sottolineare che le precedenti affermazioni *non* dicono che la forza f agente su P possa essere *direttamente* osservata/misurata in qualche modo; ciò che si osserva/misura direttamente sono i suoi effetti cinematici su P (1° membro della (1)). È piuttosto il modello \mathcal{M} che, “alimentato” con opportune osservazioni/misure (che *in generale non sono direttamente quelle della forza f in \mathcal{D}*), fornisce tale forza in \mathcal{D} , e dunque, con le due condizioni accessorie, consente di determinare predittivamente la traiettoria di P e di confrontarla con la sua ricognizione sperimentale.¹³

Alla seconda legge si associa un **principio di sovrapposizione degli effetti dinamici**. Si verifica sperimentalmente, cioè, che l'azione simultanea di più forze f_1, f_2, \dots, f_n (provvedute da altrettanti modelli) sullo stesso punto materiale P induce su di esso l'accelerazione che indurrebbe la forza somma *vettoriale* delle forze individuali $f_i, \sum_{i=1}^n f_i$; e si eleva poi questo risultato al rango di principio. Quindi, poiché l'accelerazione indotta da f_i è $a_i = f_i/m_p$, l'accelerazione a indotta dalle f_i agenti simultaneamente è la somma vettoriale delle accelerazioni individuali, $a = \sum_{i=1}^n a_i$.

Esiste anche una terza e altrettanto importante legge dinamica newtoniana, indipendente dalla seconda e usualmente nota come “principio di azione e reazione”. Per illustrarla, cominciamo col considerare il seguente esperimento ideale. Siano P_1 e P_2 due punti materiali di masse m_1 e m_2 , soggetti a forze f_1 e f_2 . Immaginiamo allora di “legare” insieme i due punti in modo che essi siano costretti ad occupare in ogni istante la stessa posizione, quindi a percorrere la stessa traiettoria, e in particolare ad avere la stessa accelerazione. Affinché ciò avvenga, i due punti devono essere soggetti, oltre che alle forze f_1 e rispettivamente f_2 , a certe forze addizionali che impongano loro il detto legame di posizione, diciamo $f_{2 \rightarrow 1}$ agente da P_2 su P_1 , e $f_{1 \rightarrow 2}$ agente da P_1 su P_2 . Indicando con

risultando dalla $m_p = f^*/(2x(t))$, oppure dalla $m_p = f^*/(d_x x(t))$, valide entrambe identicamente rispetto a t in un conveniente intervallo attorno a $t = 0$.

¹² «mutationem motus proportionalem esse vi motrici impressæ, et fieri secundum lineam rectam qua vis illa imprimatur» (Newton, loc. cit.). Newton trasse la sua seconda legge dai famosi esperimenti di Galileo, che quella legge in qualche modo intuì. È precisamente alla legge dinamica (1) che risale la sufficienza di un'analisi differenziale del 2° ordine – e non di ordine maggiore – della traiettoria del punto.

¹³ Per fare un esempio importante, \mathcal{M} può dirci che la forza f agente su P è il prodotto di una costante positiva univocamente legata a P , diciamo m_{gp} , per il gradiente di un **potenziale gravitazionale** φ soluzione di un'equazione di Poisson (Siméon, 1781-1840), fornendoci al contempo tutte le informazioni necessarie a determinare unicamente tale gradiente mediante la soluzione di quell'equazione: cioè \mathcal{M} provvede in questo caso il termine libero dell'equazione di Poisson e le relative condizioni al contorno che rendono unico il gradiente della sua soluzione. La costante m_{gp} è la **massa gravitazionale di P** . A posteriori, e senza che emerga alcuna necessità concettuale di questo fatto, la massa gravitazionale m_{gp} e la massa inerziale m_p di P risultano con altissima precisione (secondo le più aggiornate stime, dell'ordine di 10^{-11}) proporzionali tra loro, al punto che vengono di norma numericamente, ed anche dimensionalmente, identificate (“principio di equivalenza debole”, v. S.sez. 9.1.3). Ciò fa sì che in questo caso la traiettoria di P , soggetta alla forza-gradiente, *non* sia influenzata dalla sua massa, perché questa compare omogeneamente nei due membri dell'equazione di Newton: come si impara (o si dovrebbe imparare) già nella prima adolescenza, “nel vuoto – cioè, in assenza d'aria – una piuma e una moneta cadono con la stessa accelerazione”.

a la comune accelerazione dei due punti, per il principio di sovrapposizione le due leggi dinamiche diventano così $m_1 a = f_1 + f_{2 \rightarrow 1}$ e rispettivamente $m_2 a = f_2 + f_{1 \rightarrow 2}$. Sommando queste si conclude che

$$(2) \quad (m_1 + m_2) a = f_1 + f_2 + f_{2 \rightarrow 1} + f_{1 \rightarrow 2}.$$

L'esperienza prova che le forze di legame $f_{2 \rightarrow 1}$ e $f_{1 \rightarrow 2}$, quali esse siano, *non* influenzano il moto del punto P "composto" dai due punti legati P_1 e P_2 ; e questo può aversi quindi, in generale, se e solo se

$$(3) \quad f_{2 \rightarrow 1} + f_{1 \rightarrow 2} \equiv 0.$$

La (2) ci dice quindi che P ha l'accelerazione di un unico punto di massa uguale alla somma delle masse dei punti componenti e che subisce l'azione della somma vettoriale delle forze individuali f_1 e f_2 . Newton estende poi questa idea al caso di punti materiali P_1 e P_2 *non coincidenti*, richiedendo in tal caso che i vettori $f_{2 \rightarrow 1}$ e $f_{1 \rightarrow 2}$ (i) siano uguali ed opposti secondo la (3), e (ii) abbiano la stessa **retta di applicazione**.¹⁴ Il **principio di azione e reazione** consiste precisamente nella congiunzione delle due richieste (i) e (ii) (la (ii) essendo vuota se i due punti coincidono).¹⁵ Per esprimerlo matematicamente occorrono e bastano *due* formule vettoriali, la (3) e la

$$(3bis) \quad f_{2 \rightarrow 1} \times (P_1 - P_2) = 0,$$

(dove \times è il prodotto vettoriale, vedi S.sez. 1.4.2), che afferma appunto il parallelismo fra $f_{2 \rightarrow 1}$ e $P_2 - P_1$. Ferma restando la (3), alla (3bis) può anche sostituirsi la

$$(3ter) \quad f_{2 \rightarrow 1} \times (P_1 - O) + f_{1 \rightarrow 2} \times (P_2 - O) = 0,$$

dove O è un terzo punto qualsiasi. Se f è la forza applicata al punto P, il vettore $f \times (P - O)$ si dice suo **momento rispetto a O**. La (3ter) afferma quindi che la somma dei momenti rispetto a O delle forze applicate a P_1 e P_2 (la somma delle quali forze è nulla, eq. (3)) è nulla. Si vede subito che se la (3ter) vale per un certo O, sotto la (3) essa vale per qualunque O' in luogo di O; ovvero, che sotto la (3), (3bis) \Leftrightarrow (3ter). In una versione più generale, ma *deducibile* da quella a due punti, il principio di azione e reazione si enuncia dicendo che "la somma delle forze (*binarie*) interne ad un insieme qualsiasi di punti materiali (cioè delle forze tra coppie di punti distinti dell'insieme), e quella dei loro momenti rispetto a un arbitrario punto O, sono entrambe nulle".¹⁶ Come è facile intuire, il

¹⁴ La retta di applicazione di un vettore applicato in P è la retta, unicamente determinata, parallela al vettore e passante per P. Se due vettori applicati in punti distinti P_1 e P_2 hanno la stessa retta di applicazione, questa deve essere la retta passante per P_1 e P_2 .

¹⁵ «actioni contrariam semper et æqualem esse reactionem: sive corporum duorum actiones in se mutuo semper esse æquales et in partes contrarias dirigi» (Newton, loc. cit.).

¹⁶ Può essere utile dare qui qualche ragguaglio sugli "insiemi di vettori applicati" e sulla loro possibile equivalenza. Un **vettore applicato** (o **curso**re) è la coppia formata da un vettore v e dal suo punto di applicazione P, $\{v, P\}$. Di un insieme di cursori $\Sigma \equiv \{v_i, P_i\}_{i=1, \dots, N}$, la somma $R = \sum_i v_i$ si dice suo **risultante**, e la somma $M_O = \sum_i [v_i \times (P_i - O)]$ si dice suo **momento totale rispetto a O**. Moltiplicando scalarmene il risultante di Σ per il suo momento totale rispetto a O, si ottiene uno scalare $R \cdot M_O$ che non dipende più da O: infatti $M_{O'} = \sum_i [v_i \times (P_i - O')] = M_O + R \times (O - O')$, da cui $R \cdot M_{O'} = R \cdot M_O$. Questo scalare indipendente da O si dice **automomento** o **momento invariante** di Σ . Due insiemi di cursori $\Sigma \equiv \{v_i, P_i\}_{i=1, \dots, N}$ e $\Sigma' \equiv \{w_j, Q_j\}_{j=1, \dots, M}$ si dicono **equivalenti** (o talvolta, **equipollenti**) se hanno uguali (i) i **risultanti** e (ii) i **momenti totali** rispetto ad un polo O. Si vede subito che, sotto la (i), la (ii) è soddisfatta per *qualunque* O se lo è per un *certo* O. Infatti $M_{O'} = M_O + R \times (O - O')$, e similmente $M'_{O'} = M'_O + R' \times (O - O')$; quindi, sotto $R = R'$, la $M_O = M'_O$

principio di azione e reazione ha avuto un ruolo centrale nello sviluppo della teoria newtoniana della gravitazione universale (vedi S.sez. 6.4.3).

L'equazione vettoriale (1) equivale ad un sistema di tre equazioni differenziali ordinarie (scalari) del secondo ordine in forma normale nelle tre incognite $x_i(t)$ (con $i = 1,2,3$). Se la forza è data nel modello \mathcal{M} esplicitamente come funzione del solo t in un certo intervallo temporale base, questo sistema si risolve mediante quadrature sotto i due soliti dati accessori (vettoriali, pari a sei dati numeri reali). Ma come sappiamo, in generale la forza è data in \mathcal{M} come funzione non solo di t , ma anche della posizione e/o della velocità di P al tempo t . Allora il sistema non può più risolversi per quadrature; ma come insegna l'Analisi, esiste comunque una ed una sola sua soluzione sotto ben note condizioni sufficienti (di cui abbiamo dato l'esempio canonico di continuità di f e delle derivate parziali f_x e f_v) e gli stessi (o convenienti altri ¹⁷) dati accessori.

La teoria dei sistemi differenziali ordinari ci insegna che il sistema differenziale ordinario in oggetto (tre equazioni del secondo ordine in forma normale nelle tre incognite $x(t)$) può sempre scriversi nella forma di un equivalente sistema di sei equazioni del primo ordine (sempre in forma normale), nelle sei incognite $x(t)$ e $v(t)$. Il problema si dice **autonomo** se l'espressione della forza non contiene esplicitamente t . ^{18, 19} Questo schema si generalizza facilmente al caso di un insieme finito di punti materiali P_1, \dots, P_N , che dà luogo ad un sistema di $6N$ equazioni differenziali del primo ordine (in generale non disaccoppiabili in N sottosistemi differenziali indipendenti di 6 equazioni l'uno se l'espressione della forza su di un punto dipende *essenzialmente*, oltre che da t ,

equivale alla $M_O = M'_O$. La relazione: “ Σ e Σ' sono equivalenti” è evidentemente una relazione di equivalenza. Una **coppia** è un insieme di due cursori del tipo $\{v, P\}$ e $\{-v, Q\}$ con $P \neq Q$; quindi ha risultante nullo e momento totale rispetto a O uguale a $v \times (P-O) - v \times (Q-O) = v \times (P-Q)$, indipendente da O . Un insieme qualsiasi di cursori $\Sigma = \{v_i, P_i\}$ è equivalente all'insieme di cursori Σ' formato dal risultante R di Σ applicato in un punto arbitrario O e da una coppia di momento pari al momento totale M_O rispetto a O di Σ (e per il resto arbitraria). Infatti il momento totale rispetto a O di Σ' si riduce a quello della sua coppia, che è M_O per definizione. Un sistema Σ di cursori avente risultante e momento totale rispetto ad un certo polo entrambi nulli si dice **nullo**; oltre che risultante nullo, esso ha quindi nullo il momento totale rispetto ad un polo qualsiasi. Consideriamo un sistema di punti materiali $\{P_i\}_{i=1, \dots, N}$ *isolato* (\equiv sul quale non agiscono forze esterne) ma soggetto a forze interne *binarie*, quindi del tipo $f_{i \rightarrow j}$ su P_j da parte di P_i , per $i \neq j$. Per il principio di azione e reazione, $f_{i \rightarrow j}$ e $f_{j \rightarrow i}$ sono uguali ed opposte e hanno la stessa retta di applicazione, quella che passa per P_i e P_j . Segue che $R = \sum_{i,j} f_{i \rightarrow j} = 0$ (dove l'apice sulla doppia somma Σ significa che in essa si salta $i = j$). Similmente $M_O = \sum_{i,j} f_{i \rightarrow j} \times (P_j - O) = (1/2) \sum_{i,j} f_{i \rightarrow j} \times (P_j - P_i) = 0$. Il principio di azione e reazione si enuncia dunque come è fatto nel testo tra le “ ”, e le forze sui punti costituiscono nel loro insieme un sistema di cursori nullo.

¹⁷ Ad esempio, assegnando la posizione di P in due punti della sua traiettoria.

¹⁸ Per l'omogeneità rispetto a t , la descrizione di un moto autonomo (\equiv soluzione di un sistema dinamico autonomo) è dunque indipendente dalla scelta dell'origine dei tempi.

¹⁹ Si consideri l'equazione differenziale (*) $d^2x/dt^2 = h(t, x, dx/dt)$ in $t \in (0,1)$, con le condizioni ai limiti $x(t=0) = a$, $x(t=1) = b$, e si supponga che h soddisfi alla simmetria (**) $h(z, x, y) = h(1-z, x, -y) \forall (x, y) \text{ e } \forall (z) \in (0,1)$. Posto allora $t+\tau = 1$, il moto “inverso” $x = x(\tau)$ tra $x(\tau=0) = b$ e $x(\tau=1) = a$, che percorre all'indietro il moto “diretto” $x = x(t)$, è a sua volta soluzione della equazione che si ottiene dalla (*) sostituendovi formalmente t con τ . Questa proprietà si dice **reversibilità del moto** tra $t = 0$ e $t = 1$, il caso di maggior interesse pratico essendo quello in cui la simmetria (**) è soddisfatta perché h non dipende né da y né da z . Essa si estende facilmente ai sistemi, relativi a N punti mobili (vedi il testo subito appresso), di $6N$ equazioni accoppiate di tipo (*), con simile simmetria (**) e simili condizioni ai limiti, e (banalmente) per $t \in (t_1, t_2)$, $t_1 < t_2$. Un moto retto dalla (*) supposta *non reversibile* è una particolare manifestazione empirica della freccia del tempo, che ne definisce il verso.

anche dalle posizioni e/o dalle velocità degli altri punti), nelle $6N$ incognite corrispondenti (posizioni e velocità dei punti), sotto il vincolo del principio di azione e reazione (“problema degli N corpi”). L’intera cosiddetta **Dinamica Analitica** classica (vedi Sez 6.3 e 6.4), sostanzialmente iniziata da Lagrange (Joseph, Torino 1736, Parigi 1813) e Hamilton (William, Dublino 1805-1865), si sviluppa, sotto certe condizioni²⁰ e attraverso importanti generalizzazioni ed astrazioni, da questo semplice nucleo di idee.²¹

Riferendoci di nuovo al moto di un solo punto materiale P di massa m_P , supponiamo di passare dal riferimento assoluto (della classe \mathcal{I}) \mathfrak{S} ad un riferimento \mathfrak{s} in moto arbitrariamente dato rispetto ad esso, cioè di tipo rototraslatorio con i sei parametri funzioni continue di t (di classe C^2) arbitrariamente date. Secondo l’esperienza, si suppone che il *vettore* “forza agente su P ” *non vari* passando al nuovo riferimento²², ovvero che $F = f$. Inoltre m_P dipende esclusivamente da P , e quindi anch’essa *non varia* passando da un riferimento all’altro. Invece l’accelerazione A di P (secondo \mathfrak{S}) diventa *un altro* vettore nel riferimento mobile \mathfrak{s} , cioè la sua accelerazione relativa a , generalmente diversa dall’accelerazione assoluta A . Conseguenza da ciò che la legge dinamica *non è invariante* per generiche rototraslazioni (proprie) del riferimento, date in funzione di t . Sappiamo tuttavia che esiste una (e una sola) eccezione, quella delle traslazioni uniformi, per le quali accelerazione assoluta e relativa coincidono. Come abbiamo detto, la classe di equivalenza dei

²⁰ La precisa identificazione di queste condizioni costituisce il cosiddetto **problema inverso** della dinamica analitica.

²¹ Sistemi fisici la cui evoluzione temporale è retta da equazioni differenziali si dicono spesso, sotto certe restrizioni, **sistemi dinamici**. In generale, non è detto che la soluzione delle equazioni dinamiche sia “praticamente determinabile” dopo che sia trascorso un tempo abbastanza lungo dal tempo iniziale, anche se la sua esistenza/unicità è assicurata dalle usuali condizioni e dalla prescrizione dei dati iniziali. Infatti essa può risultare *instabile* – cioè discontinua – rispetto a tali dati (o addirittura rispetto al modello dinamico di cui essi fanno parte, cfr. S.sez. 0.1.2). Questa situazione di “impredicibilità virtuale” della evoluzione del sistema mina alla base la vecchia idea di Laplace (Pierre, 1749-1827) che il suo stato sia predicibile a qualunque tempo $t > 0$ (per un dato e non perturbabile modello deterministico della evoluzione) *con la stessa precisione* con cui è conosciuto al tempo iniziale $t = 0$. Nelle parole di Poincaré: «Se la conoscenza delle leggi naturali ci permettesse di predire la situazione successiva (e abbastanza lontana, ndr) di un dato universo con la stessa approssimazione (con cui conosciamo quella iniziale, ndr), questo è tutto ciò che chiediamo, e diremmo che il fenomeno è stato predetto. Ma non sempre è così: può infatti succedere che piccole differenze nelle condizioni iniziali producano un grandissimo errore in quelle successive. La predizione diventa allora impossibile.» (“Science et Méthode”, 1908). La impredicibilità può essere già presente in sistemi dinamici assai semplici, ed è specificamente legata ad una non-linearità delle equazioni di evoluzione. Circa mezzo secolo fa, problemi di questo tipo cominciarono ad essere collegati a quella che poco più tardi si sarebbe detta **teoria del caos deterministico**. Questa teoria studia leggi di evoluzione deterministica in cui *coesistano l’iterazione indefinita di un processo ricorsivo e la non-linearità*, e per il resto del tutto generali. Non è difficile intuire che in tali condizioni l’evoluzione può assumere carattere caotico, e quindi diventare virtualmente impredicibile, da un certo tempo in poi. Un esempio semplicissimo e assai noto di impredicibilità virtuale è quello originato dalla legge deterministica, ricorsiva e non lineare (geralmente detta “logistica”) data da $x_{n+1} = \alpha x_n(1-x_n)$, $n = 0, 1, \dots$, per α (\equiv un parametro positivo fisso) abbastanza grande e naturalmente per $x_0 > 0$. La teoria del caos deterministico – in particolare lo studio della possibile transizione al caos del sistema dinamico considerato – e le sue applicazioni alle scienze fenomenologiche hanno riscosso largo interesse, e intorno ad esse si è ormai sviluppato un importante capitolo della fisica matematica contemporanea (vedi App. Gen. D.5 per qualche dettaglio).

²² Ciò implica che la forza su P , supposta generalmente funzione della sua posizione X e velocità V (oltre che di t), dipenda dai suoi argomenti in modo in generale *diverso* per S ed s , diciamo come $F(t, X, V)$ e rispettivamente $f(t, x, v)$; ma diverso proprio così che i valori di $F(t, X, V)$ e di $f(t, x, v)$ siano uguali, $F(t, X(t), V(t)) \equiv f(t, x(t), v(t))$. Considerazioni analoghe valgono nel caso di un sistema N di punti materiali.

riferimenti in traslazione uniforme l'uno rispetto all'altro *che ha tra i suoi elementi il riferimento assoluto* è la classe inerziale \mathcal{I} ; il gruppo continuo a dieci parametri indipendenti dal tempo delle corrispondenti traslazioni uniformi si dice **gruppo galileiano**. Quindi la legge dinamica (1) è invariante rispetto alle trasformazioni del gruppo galileiano, e soltanto rispetto ad esse. Ciò implica che gli stessi esperimenti dinamici (su un punto materiale) eseguiti in riferimenti diversi della classe inerziale \mathcal{I} devono dare risultati indistinguibili. Sotto i principi di sovrapposizione degli effetti e di azione e reazione, questo vale anche per insiemi arbitrari, e al limite continui, di punti materiali. Si conclude che un osservatore della classe \mathcal{I} non può rilevare la sua velocità rispetto al riferimento assoluto mediante esperimenti dinamici qualsiasi (**principio di relatività galileiana**).²³,²⁴ Se poi questa conclusione viene generalizzata dicendo “mediante esperimenti di ogni tipo” anziché “mediante esperimenti dinamici”, ed elevata al rango di principio, allora si parla di **principio di relatività tout court**.

Naturalmente non si deve pensare che la traiettoria di P non si possa calcolare in un riferimento \mathfrak{s} *non-inerziale*, in moto dato rispetto a quello inerziale-assoluto \mathfrak{S} : vi sarà soltanto da tener conto del fatto che, in quel riferimento, la legge dinamica consiste nell'uguagliare il prodotto della massa di P per la somma dell'accelerazione relativa a \mathfrak{s} , dell'accelerazione di Coriolis (nota) e di quella di trascinamento (nota anch'essa) alla forza (vettore intrinseco) su P. L'equazione differenziale vettoriale risultante è dello stesso tipo della (1), cioè del secondo ordine e in forma normale, generalmente non autonoma, e può dirsi **legge dinamica generalizzata**. Si intuisce facilmente, tuttavia, che passando ad un altro riferimento \mathfrak{s}^* in moto traslatorio uniforme rispetto ad \mathfrak{s} , tale equazione *non* si conserva formalmente invariata (salvo il caso che \mathfrak{s} sia in moto inerziale, e quindi anche \mathfrak{s}^* lo sia). Un approfondimento di questo punto è riportato nell'App. Spec. 2.A.

²³ Il lettore certamente conosce la famosa ed efficacissima descrizione che Galileo dà di questa situazione, immaginandola verificarsi nella stiva di un grosso vascello in navigazione “rettilenea uniforme”; cioè in un mare idealmente calmo e sotto una spinta costante che faccia esattamente equilibrio alla resistenza idrodinamica.

²⁴ Un analogo principio di relatività (invarianza dei risultati di esperimenti dinamici per osservatori inerziali) vale anche in relatività speciale, seppure in forza di meccanismi più complicati che nel caso classico (vedi S.sez. 2.2.1). Si parla allora di **principio di relatività einsteiniana**. Emerge dall'insieme abbastanza disordinato di occasioni in cui il termine “relatività” compare nel discorso con significati diversi, come la denominazione standard della teoria einsteiniana (storicamente *non* attribuibile a Einstein) presenti qualche inconveniente. La confusione aumenta, presso il grande pubblico, perché praticamente in tutte le lingue l'attributo “relativistico” significa sia “attinente al relativismo” che “attinente alla teoria della relatività”.

2.2) INTRODUZIONE ALLA CINEMATICA E ALLA DINAMICA RELATIVISTICHE SPECIALI

2.2.1) CINEMATICA RELATIVISTICA SPECIALE E TRASFORMAZIONI DI LORENTZ

Come si diceva nella precedente sezione, la percezione di un tempo delocalizzato che etichetta istante per istante tutto lo spazio è perdurato per millenni a causa della possibilità di comunicare da punto a punto dello spazio, e in tempi che più o meno inconsciamente si stimavano trascurabili rispetto a quelli di reale interesse. Seppure ci si poneva un problema di genere così esotico, si pensava cioè che dopotutto un segnale luminoso, o anche acustico-meccanico, viaggia a velocità abbastanza grande da potersi considerare infinita a tutti gli effetti pratici.

La rivoluzione relativista della meccanica newtoniana inizia con una critica di pretto stampo operazionalista sulla possibilità di principio di definire un tale tempo delocalizzato, svolta da A. Einstein nella sua celebre memoria del 1905.¹ Il presupposto di poter trasmettere segnali da punto a punto con velocità infinitamente grande è infatti falso: ogni segnale viaggia con velocità finita, e di questa finitezza bisogna tenere opportunamente conto. Inoltre nella realtà fisica esistono soltanto orologi localizzabili (anzi, orologi da pensare abbastanza piccoli nella scala macroscopica cui ci si riferisce); supposti identici tra loro, questi orologi si dovranno immaginare posti “stabilmente”² in questo o quel punto, per cui ogni punto avrà il “suo” tempo, indicato dall’orologio locale. Ma è chiaro che si cade in un circolo vizioso se questi orologi si vogliono sincronizzare tenendo conto della velocità dei segnali, perché è impossibile valutare la velocità di qualcosa che si muove tra due punti prefissati se non si dispone di orologi *già tra loro sincronizzati* in quei punti.

A prima vista, si potrebbe pensare di trasportare i vari orologi in una opportuna “stazione di sincronizzazione” situata in un punto prefissato dello spazio, sincronizzarli, e poi riportarli nella loro posizione originale; ma nulla assicura che questo trasporto “avanti e indietro” non alteri la regolarità del loro ritmo di marcia.³ Nella consapevolezza di questa insidia, è opportuno rinunciare

¹ A. Einstein, “Zur Elektrodynamik bewegter Körper” *Annalen der Physik*, 17, 891, 1905. In questo fondamentale lavoro il problema “tempo delocalizzato vs. tempo localizzato” è per la prima volta affrontato lucidamente, e pienamente risolto. Beninteso, e a buona ragione, il segnale considerato da Einstein nei suoi esperimenti mentali è un segnale luminoso, o comunque elettromagnetico.

² In questa possibilità si cela evidentemente la nozione di “sistema rigido di punti”, che fa capo a quella anche più generale e intuitiva di “corpo rigido” (e in particolare di “asta rigida” con cui misurare le distanze). Si tratta in ogni caso di nozioni primitive. Al presente, alla misura di una distanza si preferisce ormai la più precisa misura del tempo necessario a percorrerla “avanti e indietro” da parte di un segnale luminoso. La necessaria precisazione appena riportata tra “ ” è discussa in dettaglio più avanti in questa sezione.

³ Esperimenti relativamente recenti provano che questo è proprio quanto si verifica nella realtà, seppure in misura trascurabile nelle condizioni della esperienza comune.

cautelativamente all'idea della stazione di sincronizzazione; ed Einstein la scarta infatti senza esitazioni (non nominandola nemmeno). Resta un'altra possibile procedura. Vale a dire – argomenta Einstein –, avendo inviato un segnale (abbastanza localizzato nello spazio e nel tempo) da un orologio posto in A (“orologio trasmittente”) verso un orologio posto in B (“orologio ricevente”), non è necessario conoscere il suo tempo di arrivo in B per avere una stima *plausibile* della sua velocità di propagazione; basta rinviarlo ad A immediatamente dopo la sua ricezione (idealmente, nello stesso istante, come farebbe uno specchio appunto ideale), e *presupporre*⁴ che i due viaggi abbiano uguale durata. In questo modo la durata del viaggio di andata da A a B (o di ritorno da B ad A) viene uguagliata alla metà di quella dell'intero percorso, da A ad A dopo la riflessione in B, diciamo $\delta t > 0$. L'orologio in B viene poi congruamente regolato in modo che, nel momento in cui si verifica la riflessione del segnale, segni un tempo t_B uguale a quello della sua partenza da A, diciamo t_A , aumentato della presunta durata del viaggio da A a B, cioè di $\delta t/2$; ovvero, equivalentemente, uguale alla media aritmetica tra il tempo t_A e quello del suo ritorno in A dopo la riflessione in B, diciamo t_A^* (due eventi co-locali, per cui $\delta t = \tau_A^* - \tau_A$), ovvero $\tau_B = (\tau_A + \tau_A^*)/2$.

Quella che abbiamo descritto è precisamente la procedura di sincronizzazione (di un orologio “incollato ad A” rispetto ad un orologio identico “incollato a B”) proposta da Einstein – il quale si riferiva esplicitamente ad un segnale *luminoso* – nel suo succitato lavoro. Resta tuttavia il fatto che una simile procedura non può essere interamente fondata su una convenzione: occorre che essa «sia libera da contraddizioni» (vedi Einstein, loc. cit.). Einstein vede dunque questo problema della coerenza, e correttamente lo specifica nella richiesta (necessaria e sufficiente) che la relazione “l'orologio in B è sincronizzato all'orologio in A” sia una relazione di equivalenza, cioè sia simmetrica e transitiva (la riflessività è banalmente evidente).⁵ Tuttavia egli non *dimostra* la coerenza della procedura partendo da un fatto sperimentale primitivo, ma si limita a postularla dopo aver introdotto la sua definizione di sincronismo. In altri termini, Einstein *prima* fissa (induttivamente) un criterio convenzionale di sincronismo, e *poi* confida che la fisica sia benevola in ordine alla coerenza della sua definizione.⁶

⁴ Einstein dice «stabilire *per definizione*» (il corsivo è originale); ma è ovvio che una tale scelta resta una convenzione, per quanto ragionevole. Vale a dire, si può al più costruire una teoria che accoglie tale convenzione, studiarne la consistenza interna, e confrontare le sue implicazioni con l'esperienza.

⁵ Si dovrebbe qui aggiungere la richiesta che la relazione in oggetto sia indifferente rispetto ad uno spostamento rigido della coppia di orologi e rispetto all'istante di partenza del segnale. Questa richiesta di *omogeneità*, sia dello spazio che del tempo, è data tacitamente per soddisfatta.

⁶ È possibile che Einstein abbia preferito procedere in questo modo per non introdurre la nozione di costanza della celerità (\equiv modulo della velocità) della luce prima di quando a suo parere fosse strettamente indispensabile, poco più di mezza pagina di testo.

Va anche ricordato che, a complicare la situazione, prima che Maxwell pubblicasse la sua teoria elettromagnetica (1873) non si aveva nozione di segnali che potessero propagarsi *nel vuoto*; ma anche più tardi, per qualche tempo ancora questa idea si dimostrò così radicata da obbligare i fisici dell'epoca a postulare l'esistenza di un "mezzo" di non ben definita natura che, permeando tutto lo spazio, fungesse da supporto alla propagazione dei campi elettromagnetici, il cosiddetto "etere". Ma a quel punto si poneva naturale la domanda: "come si muove il laboratorio terrestre (ad esempio) rispetto all'etere?" Vedremo tra un momento che la risposta doveva avere ricadute sostanziali sul problema della sincronizzazione degli orologi.

Se verificare che la durata del percorso di un segnale da A a B \neq A è uguale a quella del percorso da B ad A è *logicamente* impossibile, si può peraltro confrontare la durata del suo viaggio di andata e ritorno $A \rightarrow B \rightarrow A$ con quella di un analogo viaggio $A \rightarrow B' \rightarrow A$, dove $B' \neq B$ ma $|AB'| = |AB|$. Il caso più significativo, al quale ci riferiremo nel seguito, è evidentemente quello in cui i segmenti di uguale lunghezza AB e AB' sono perpendicolari. In presenza di un (presunto) "vento d'etere", detta $|v|$ la celerità (\equiv modulo della velocità) di questo rispetto al laboratorio terrestre, e $c > 0$ la celerità del segnale luminoso rispetto al mezzo in cui si propaga (cioè all'etere stesso), si può facilmente dimostrare che la differenza tra il tempo di percorrenza del percorso $A \rightarrow B \rightarrow A$ e quello del percorso $A \rightarrow B' \rightarrow A$ è dell'ordine del quadrato del rapporto $|v|/c$, se questo è supposto abbastanza piccolo da trascurarne le potenze superiori alla prima. Precisamente, se $0 \leq \theta \leq \pi/2$ è l'angolo tra la direzione del vettore velocità v del vento d'etere e quella del segmento AB, detti τ_B e $\tau_{B'}$ i due tempi di percorrenza si trova che

$$(1) \quad \tau_B - \tau_{B'} = (|v|^2/c^3)|AB|\cos 2\theta.$$

Questa differenza è quindi massima in valore assoluto per $\theta = 0$ e $\theta = \pi/2$ (positiva nel primo caso e negativa nel secondo), mentre si annulla, cambiando segno, per $\theta = \pi/4$; come è abbastanza naturale se ci si riflette un momento.

Questo è nella sua essenza il significato del famoso esperimento di Michelson (Albert, 1852-1931) e Morley (Edward, 1838-1923), iniziato a partire dal 1881 e poi ripetuto in successive versioni sempre più precise e raffinate. Astutamente, il confronto in questione fu effettuato mediante tecniche interferometriche. Supponendo (ragionevolmente) l'etere solidale con il sole, o addirittura con la Galassia (cioè con l'intero universo quale lo si pensava allora), il rapporto $|v|/c$ si stimava dell'ordine di 10^{-4} (rapporto tra la velocità orbitale terrestre, ca. 30 km/s, e la velocità della luce, ca. 300.000 km/s); mentre la sensibilità dell'esperimento si valutava essere circa 40 volte

maggiore di quella minima necessaria a rivelare il vento d'etere.⁷ Tuttavia l'esperimento di Michelson e Morley (MM) non diede affatto il risultato che ci si sarebbe aspettati ragionando in modo classico: comunque si orientasse il triangolo isoscele ABB' rettangolo in A , non si riuscì mai ad evidenziare l'esistenza di un vento d'etere. Naturalmente si fecero una quantità di ipotesi di ripiego atte a spiegare l'esito dell'esperimento (di) MM (poi accanitamente ripetuto fino al 1887 in una serie di versioni sempre più raffinate), ma senza successo. Il rimedio radicale a tale impasse fu alla fine indicato da Einstein nel suo articolo del 1905,⁸ meno di una trentina di pagine che hanno cambiato per sempre il corso della fisica. Come abbiamo anticipato nella sezione precedente, esso consiste *innanzitutto* nel rimuovere l'idea, fino a quel punto quasi unanimemente accettata, di un tempo "universale" $t \equiv T$ altrettanto buono (possibilmente a meno di una costante additiva) per \mathfrak{S} e per \mathfrak{s} , distinguendo *sostanzialmente* T (tempo di \mathfrak{S}) da t (tempo di \mathfrak{s}).

L'esito degli esperimenti MM implicava che la terra fosse solidale con l'etere, e quindi suggeriva una visione assurdamente geocentrica dell'universo all'epoca noto. Non solo, ma il risultato non cambiava rifacendo il test in diversi momenti dell'anno, allorché il vento d'etere, se davvero presente, avrebbe dovuto consistentemente cambiare in entità e direzione: un vero e proprio problema nel problema.

Non disponendo di un reticolo (tridimensionale) rigido di orologi identici già sincronizzati tra loro non è possibile definire una velocità se non riferendola ad un percorso *chiuso*; ad esempio quello costituito dallo stesso segmento percorso avanti e indietro (Einstein), oppure da un triangolo,

⁷ Concettualmente, e sempre in un'ottica classica, il risultato dell'esperimento MM (per semplicità con $\theta = 0$) è del tutto equivalente a quello che ci si attende nella ben nota metafora didattica "dei due nuotatori". Due nuotatori N' e N , capaci della stessa velocità c (rispetto all'acqua in cui nuotano) si cimentano nella gara seguente: N' attraversa avanti e indietro un fiume di larghezza l nuotando perpendicolarmente alle rive, mentre N percorre avanti e indietro la stessa distanza l nuotando lungo il filo della corrente. Se v è la velocità della corrente, si vede subito che il tempo impiegato da N' è $\tau_{N'} = (2/c)(1-v^2/c^2)^{-1/2}$, mentre quello impiegato da N è $\tau_N = (2/c)(1-v^2/c^2)^{-1}$. Si conclude quindi che $\tau_N - \tau_{N'} = (2/c)[1-(1-v^2/c^2)^{1/2}](1-v^2/c^2)^{-1} \approx v^2/c^3$ nell'approssimazione in cui si trascurano le potenze di $|v|/c$ superiori alla seconda. Ciò è in accordo con quanto si ricava dalla (1) del testo per $\theta = 0$, $|AB| = 1$. La gara è dunque vinta da N' .

⁸ Il già citato "Zur Elektrodynamik bewegter Körper". Sebbene sia poco credibile, qualcuno sostiene che quando Einstein elaborò la sua teoria ignorava il risultato degli esperimenti MM. Ciò contrasta con una frase della sua memoria, ove si accenna agli «infruttuosi tentativi di mettere in evidenza un moto relativo della terra rispetto al mezzo luminifero». Sia come sia, oggetto di discussione è semmai la misura in cui le idee di Einstein furono direttamente influenzate dall'evidenza sperimentale. Il punto è che Einstein teneva indubbiamente conto anche di un fatto teoretico, ossia della invarianza (o covarianza, secondo un altro modo di esprimersi) delle equazioni di Maxwell (James, Edimburgo 1831, Cambridge 1879) rispetto alle trasformazioni di Lorentz (Hendrik, Arnhem Ol. 1853, Haarlem Ol. 1928) (1904), sotto le convenienti formule di trasformazione del campo elettromagnetico e delle sue eventuali sorgenti. Di fatto, la memoria einsteiniana del 1905 è priva di riferimenti bibliografici, (salvo un cenno fuori testo al lavoro di Lorentz del 1904, ma per affermare che esso «non era noto all'autore» all'epoca della sua ricerca), e tanto meno menziona l'esperimento MM. È da presumere che con questa scelta Einstein abbia inteso sottolineare l'indipendenza dei suoi risultati cinematici da ogni precedente ricerca sullo stesso tema (vedremo oltre da quali *principi* Einstein inferì le sue conclusioni). Gli storici della Scienza hanno dedicato molta attenzione a questo punto, e in particolare al rapporto Einstein-Michelson/Morley; si veda, tra gli altri contributi, il documentato saggio di G. Holton, "Constructing a Theory: Einstein's Model", The American Scholar, vol. 48 (1979), trad. ital. in Einaudi, raccolta di saggi dello stesso autore sotto il titolo "L'immaginazione scientifica".

oppure da una spezzata (generalmente non piana) di quattro o più lati.⁹ In un'ottica operazionalistica, il risultato negativo dell'esperimento MM diventava dunque un forte argomento a sostegno della plausibilità della procedura di sincronizzazione proposta da Einstein. D'altra parte, accettando la legge di trasformazione delle velocità (2.1.1, 2) per gli osservatori \mathfrak{S} e \mathfrak{s} , e riferendola ad un segnale luminoso in moto con velocità $c > 0$ lungo l'asse X del riferimento \mathfrak{S} , la velocità del segnale nel riferimento \mathfrak{s} , in moto con velocità Y rispetto a \mathfrak{S} , sarebbe stata $c - Y$. Postulando l'invarianza della celerità della luce misurata da un osservatore inerziale dotato del suo reticolo rigido di orologi sincronizzati "alla Einstein" rispetto ad un altro osservatore inerziale (perciò traslante uniformemente rispetto al primo), similmente dotato del suo reticolo rigido di orologi sincronizzati, si entrava quindi in flagrante conflitto – anche se quantitativamente trascurabile nelle condizioni dell'esperienza comune – con la legge galileiana (2.1.1, 2).

Si può dare al sopradetto postulato la veste di una "verità sperimentale" che reciti: (°) "il rapporto tra la lunghezza di una spezzata chiusa con $n \geq 2$ lati immersa nello spazio *vuoto* e immobile rispetto al reticolo rigido degli orologi (di un dato osservatore inerziale), lungo la quale un segnale luminoso si propaga partendo da un suo punto (ad esempio da un suo vertice) sino al suo ritorno nello stesso punto, ed il tempo da esso impiegato per questo viaggio (misurato dall'orologio incollato a quel punto), è una costante c indipendente da tutte le modalità prescelte per l'operazione, pari a ca. $3 \cdot 10^8$ m/s." Per quanto ci occorre per dimostrare su questa base la natura di equivalenza della relazione "l'orologio in A è sincronizzato con l'orologio in B" è sufficiente limitarsi al caso del triangolo, vedi l'App. Spec. 2.B. Riferito ad un tale triangolo, l'asserto (°) basta a dimostrare la coerenza del seguente criterio di sincronizzazione: (°°) "l'orologio posto in B è sincronizzato rispetto a quello posto in $A \neq B$ se $t_B = t_A + |AB|/c$ ", (ove $|AB|$ è al solito la distanza tra A e B, e $c = 2|AB|/(t'_A - t_A)$ secondo (°)). Questo criterio *implica* quello einsteiniano: utilizzando come circuito il segmento AB percorso dal segnale da A a B e poi riflesso da B in A, da (°) si ottiene $t_B - t_A = t'_A - t_B$ (infatti $t_B - t_A = |AB|/c$ in forza della definizione (°°), e $t'_A - t_B = t'_A - t_A - (t_B - t_A) = 2|AB|/c - |AB|/c = t_B - t_A$), ossia $t_B = (t_A + t'_A)/2$. Sottolineiamo tuttavia che *non* sarebbe possibile dimostrare la transitività della relazione di sincronismo "in generale" limitando l'asserto (°) a circuiti-segmenti: occorre riferirlo anche a circuiti-triangoli non-degeneri (vedi ancora l'App. Spec. 2.B). L'insieme degli orologi sincronizzati mediante il criterio (°°) forma dunque una classe di equivalenza in quello di tutti gli orologi (identici) del reticolo rigido *comunque* regolati; e la scelta di una tale classe è sufficiente ai nostri scopi, attesa l'assunta omogeneità dello spazio e del tempo.

⁹ Questa idea sembra risalire a Fizeau (Armand, 1819-1896), e in modo esplicito, più tardi, a Poincaré.

In definitiva, tutto avviene “come se” l’ipotetico mezzo attraverso il quale il segnale luminoso si propaga, omogeneo rispetto allo spazio e al tempo, sia immobile rispetto al reticolo rigido degli orologi di *qualunque* osservatore inerziale. Il conflitto tra questo fatto e le (2.1.1, 2) non può essere rimosso se non optando per una sola delle due possibilità; e dovendo scartarne una, sembra ovvio tenere per buona la prima, convalidata da una diagnostica molto più raffinata che non la seconda. Si deve dunque ricercare un’alternativa alle (2.1.1, 1) che ne sia una buona approssimazione nel limite in cui $|Y|/c \rightarrow 0$ (coincidendo con essa, e con la trasformazione identica, per $Y = 0$), e che elimini la contraddizione altrimenti emergente.¹⁰

Preferiamo dare subito la soluzione di Einstein-Lorentz del problema riservandoci di giustificarla (induttivamente) e commentarla in un secondo tempo (vedi anche App. Spec. 2.C). Torniamo dunque ai due riferimenti inerziali \mathfrak{S} e \mathfrak{s} della S.sez. 2.1.1 (notazione minuscolo-maiuscolo!) il secondo in moto traslatorio uniforme rispetto al primo lungo l’asse X con velocità Y (valore e segno) per ipotesi di modulo $< c$. La trasformazione galileiana speciale (2.1.1, 1bis) tra le coordinate (X, Y, Z, T) di \mathfrak{S} e (x, y, z, t) di \mathfrak{s} viene allora modificata secondo le

$$(2) \quad x = g(X - YT), \quad y = Y, \quad z = Z, \quad t = g(T - YX/c^2),$$

dove g è un’abbreviazione per $(1 - Y^2/c^2)^{-1/2}$ (**trasformazioni di Lorentz speciali**).^{11, 12} Il fatto che le (2) e le (2.1.1, 1) coincidano con la trasformazione identica per $Y = 0$, e che le prime siano una approssimazione delle seconde tanto migliore quanto più $|Y|/c$ è piccolo rispetto a 1 non merita commenti. Come poi le trasformazioni (2.1.1, 1bis) costituiscono un gruppo continuo ad un parametro Y , altrettanto si può dire delle (2). Per quanto riguarda la relazione tra il parametro e le

¹⁰ Sia \mathfrak{S} la rete rigida inerziale di orologi che fa l’esperimento ($^\circ$), trovando, per ipotesi indipendentemente dalle modalità del suo modo di operare, il valore c ; e sia \mathfrak{s} un altro reticolo rigido di orologi in moto traslatorio uniforme rispetto a \mathfrak{S} . Se mancasse la *universalità* del risultato ($^\circ$), il solo modo sensato che \mathfrak{s} avrebbe di sincronizzare tra loro i suoi orologi sarebbe quella di regolarli sugli orologi di \mathfrak{S} (già tra loro sincronizzati) utilizzando la coincidenza spaziale, per quell’istante che dura, degli uni con gli altri. Questo criterio di sincronizzazione degli orologi di \mathfrak{s} sarebbe logicamente accettabile; ma con i suoi orologi così sincronizzati, \mathfrak{s} non troverebbe lo stesso valore c che trova \mathfrak{S} facendo a sua volta l’esperimento ($^\circ$), salvo il caso che il circuito a questo scopo utilizzato fosse un *segmento parallelo al suo moto relativo rispetto ad \mathfrak{S}* . Utilizzando segmenti diversamente orientati (o un qualsiasi triangolo non-degenere), \mathfrak{s} troverebbe valori diversi, e generalmente diversi tra loro. Ciò è precisamente quanto succederebbe se \mathfrak{S} fosse immobile rispetto al mezzo di propagazione del segnale. Queste considerazioni portano quasi fatalmente ad un esperimento del tipo MM, e il risultato negativo di questo, la riscontrata uguaglianza della velocità della luce lungo i due segmenti perpendicolari percorsi avanti e indietro, condusse alla fine ad una interpretazione dei fatti diversa, quella che ne diede Einstein con la sua teoria della relatività speciale.

¹¹ Le formule di trasformazione (2) erano già conosciute, e non a caso sono universalmente dette “di Lorentz” e non “di Einstein”, secondo l’autorevole suggerimento di Poincaré. Ciò che è importante nel contributo di Einstein è la precisa analisi epistemologica, che porta ad una genuina *interpretazione fisica* delle trasformazioni stesse. Di fatto, Einstein muove essenzialmente da quelli che da allora si diranno **principio di relatività** e rispettivamente **principio della costanza della velocità della luce**, e ne *deduce* le (2), vedi l’App. Spec. 9.A per dettagli.

¹² Può essere vantaggioso scrivere le (2) in modo più simmetrico, ponendo ovunque in evidenza le coordinate temporali ct , cT . Si ottiene così (2*) $x = g(X - cTY/c)$, $ct = g(cT - XY/c)$, nelle quali compaiono soltanto lunghezze e numeri adimensionali. Come sappiamo, il prodotto di c per il tempo (una lunghezza) si dice **coordinata di Römer**. La costante universale c può evidentemente eliminarsi (come fattore) convenendo di usarla come unità di misura di velocità, cioè usando orologi rispetto ai quali la sua misura sia 1.

nozioni standard di trasformazione identica, trasformazione composta e trasformazione inversa, è facile verificare che alla trasformazione inversa corrisponde il cambiamento di segno del parametro Y , e che alla trasformazione composta dal passaggio da \mathfrak{S} a \mathfrak{s} con parametro Y e da \mathfrak{s} a \mathfrak{s}' con parametro Y' , il valore del parametro che le corrisponde *non è* $Y + Y'$ (come nel caso galileiano), ma

$$(3) \quad Y_+ =: (Y+Y')/(1+YY'/c^2).$$

La (3) si giustifica subito scrivendo, con evidente significato dei simboli introdotti, $x' = g'(x - Y't)$ e $t' = g'(t - Y'x/c^2)$, e sostituendo in queste a x ed a t le loro espressioni in funzione di X e T . Ad esempio riferendoci alla prima relazione troviamo $x' = gg'[(1+YY'/c^2)X - (Y+Y')T]$. Scrivendo quest'ultima nella forma $x' = g_+(X - Y_+T)$, otteniamo $g_+ = gg'(1+YY'/c^2)$ e $g_+Y_+ = gg'(Y+Y')$, da cui segue la (3). Alla stessa conclusione si giunge, partendo dalla seconda relazione, attraverso la $t' = gg'[T(1+YY'/c^2) - X(Y+Y')/c^2]$. La Y_+ risulta dunque una funzione simmetrica di Y e Y' , per cui il gruppo delle trasformazioni di Lorentz speciali è commutativo come quello delle trasformazioni galileiane speciali. Supposte Y e Y' dello stesso segno, $|Y_+| < |Y+Y'|$; e al limite per Y e Y' simultaneamente tendenti a $\pm c$, anche Y_+ tende a $\pm c$. È infine ovvio che la (3) si riduce alla $Y_+ =: Y + Y'$ quando $|Y|/c$ e $|Y'|/c$ tendono a zero, ovvero nel limite in cui $c \rightarrow \infty$.

Dalle (2) si trae facilmente il nuovo legame tra le velocità V in \mathfrak{S} e v in \mathfrak{s} , che è:

$$(4) \quad v_x = \alpha(V_X - Y), \quad v_y = \alpha g^{-1}V_Y, \quad v_z = \alpha g^{-1}V_Z,$$

dove α è un'abbreviazione per $1/(1 - YV_X/c^2)$. (Basta calcolare dX/dT [dx/dt] e identificarla con V_X [con v_x], e simili.) Naturalmente anche il denominatore in α va assunto > 0 per qualunque V_X , il che è assicurato supponendo $|V_X|/c < 1$: una limitazione del tutto ragionevole alla luce della precedente $|Y|/c < 1$. Inoltre, se $V_X = \pm c$, $V_Y = V_Z = 0$, le (4) danno $v_x = \pm c$, $v_y = v_z = 0$; non si può dunque ottenere una velocità di modulo maggiore di c mediante composizione. Similmente, se $V_Y = \pm c$, $V_X = V_Z = 0$, si trova $v_x = -Y$, $v_y = \pm c(1 - Y^2/c^2)^{1/2}$, quindi $v_x^2 + v_y^2 = c^2$; e lo stesso scambiando tra loro V_Y e v_y con V_Z e rispettivamente v_z . Più in generale, se $V_X^2 + V_Y^2 + V_Z^2 = c^2$, sempre usando le (4) si trova $v_x^2 + v_y^2 + v_z^2 = \alpha^2[(V_X - Y)^2 + g^{-2}(c^2 - V_X^2)]$; e sostituendo in questa le appropriate espressioni di α e di g , con alcuni passaggi si ha $v_x^2 + v_y^2 + v_z^2 = c^2$. In definitiva, le (4) *rispettano l'invarianza* (rispetto ad osservatori inerziali) *della celerità della luce nel vuoto*. Si osservi ancora che la (3) è in accordo con la prima delle (4) in cui si faccia, *come si deve*, $V_X = Y_+$ e $v_x = Y'$.

Fondamentale è il fatto che in forza delle (2) risulta

$$(5) \quad x^2 - c^2t^2 = X^2 - c^2T^2,$$

ovvero che attraverso la trasformazione (2) delle coordinate è *invariante*, nel passaggio da \mathfrak{S} a \mathfrak{s} , la *differenza* tra il quadrato della coordinata spaziale lungo la quale si effettua il moto relativo tra i riferimenti e c^2 volte il quadrato della coordinata temporale (che è anch'esso il quadrato di una

lunghezza); e ovviamente, qualunque funzione universale di questa differenza, ad es. la differenza opposta.

Le (2) si generalizzano facilmente sostituendo alle traslazioni uniformi speciali traslazioni uniformi generiche in cui la velocità di \mathfrak{s} rispetto ad \mathfrak{S} è un vettore arbitrario, dando a \mathfrak{s} una rotazione (propria) una tantum, nonché spostando (una tantum) l'origine spaziale in x_0 e quella dei tempi in t_0 . Numereremo come (2') le nuove trasformazioni affini, che qualcuno dice ancora "trasformazioni di Lorentz", ma che sarebbe più corretto nominare come **trasformazioni di Poincaré**, a dieci parametri; e come (4') le corrispondenti trasformazioni tra le velocità (vedi (2.D, 2)). Alle trasformazioni tra vettori maiuscoli e vettori minuscoli rispetto a trasformazioni di Lorentz omogenee e senza rotazione spaziale una tantum (**trasformazioni di Lorentz parallele**, a tre parametri Υ), si può dare veste "intrinseca", nel senso che in esse il generico vettore riceve una identità indipendente dalla sua rappresentazione per componenti. Le trasformazioni di questo tipo sono riassunte nell'App. Spec. 2.D. Esse diventano normali trasformazioni tra componenti proiettandole su una terna coordinata ortogonale *comune* a \mathfrak{s} e a \mathfrak{S} , e dipendono appunto dalle tre componenti, rispetto a quella terna, del vettore Υ . In particolare la velocità della trasformazione composta da quella da \mathfrak{S} a \mathfrak{s} (velocità Υ) seguita da quella da \mathfrak{s} a \mathfrak{s}' (velocità Υ') risulta dalla $\Upsilon_+(1+\Upsilon\cdot\Upsilon'/c^2) = \Upsilon[1+(1-g^{-1})\Upsilon\cdot\Upsilon'/\Upsilon^2] + \Upsilon'g^{-1}$, ove g si intende riferita a Υ . Questa espressione di Υ_+ è manifestamente non-simmetrica in Υ, Υ' ; dunque il gruppo delle trasformazioni di Lorentz parallele *non è commutativo* in generale, ma lo diventa nel caso delle trasformazioni speciali, per le quali l'espressione di Υ_+ si riduce alla (3).

La relazione "le coordinate di \mathfrak{s} sono legate a quelle di \mathfrak{S} da equazioni del tipo (2)" è una relazione di equivalenza, e le trasformazioni (2') stesse costituiscono un gruppo continuo (a dieci parametri), o **gruppo di Poincaré**.¹³ Come in base alle (2) e (4), anche in base alle (2') e alle relative (4') si ritrova l'invarianza della celerità della luce nei due riferimenti inerziali; e inoltre, nel caso *omogeneo* (Lorentz), l'invarianza di una quantità scalare che generalizza la differenza tra quadrati di cui alla (5), cioè

$$(5bis) \quad x^2 + y^2 + z^2 - c^2t^2 = X^2 + Y^2 + Z^2 - c^2T^2.$$

¹³ Ci concediamo qui un commento per il lettore abbastanza esperto. Le proprietà del gruppo di Lorentz (associato omogeneo del gruppo di Poincaré) sono simili a, ma essenzialmente diverse da, quelle del gruppo ortogonale (euclideo), le differenze essendo connesse con l'esistenza di due tipi di inversioni, una riferita allo spazio e una al tempo, e al fatto che il gruppo di Lorentz non è compatto, perché la pseudosfera unitaria di uno spazio pseudoeuclideo di quella segnatura (vedi oltre), cioè il luogo dei punti per i quali il modulo della pseudodistanza dall'origine è 1, *non è* compatta. Le applicazioni fisiche delle trasformazioni di Lorentz sono connesse con il principio di relatività einsteiniano, secondo il quale tutte le leggi fisiche, con l'eccezione di quella gravitazionale, sono invarianti rispetto a quelle trasformazioni. In un numero di casi, ad esempio nella teoria assiomatica dei campi quantizzati, principi come quello appena nominato conducono a conclusioni di grande portata sulla reciproca dipendenza di certe quantità fisiche. Infine esistono importanti applicazioni delle trasformazioni di Lorentz nello spazio tangente ad uno spazio pseudoriemanniano, legate alle cosiddette "simmetrie spazio-temporali locali".

Sarà conveniente scrivere queste differenze come quadrati formali, diciamo d^2 e rispettivamente D^2 , anche se essi possono essere negativi.

Se pensiamo ad un insieme di segnali luminosi localizzati nello spazio e nel tempo – o come si può anche dire, di “fotoni” – emessi dall'origine di \mathfrak{S} al tempo $T = 0$, dopo il tempo $T > 0$ essi si trovano sulla sfera di raggio c^2T^2 , di equazione $D^2 = 0$; l'invarianza di D^2 rispetto alle trasformazioni (omogenee) tra coordinate ci assicura allora che secondo l'osservatore \mathfrak{s} , se la sua origine passa per l'origine di \mathfrak{S} al tempo $t = 0$, al corrispondente tempo t quei fotoni si trovano sull'analoga sfera di raggio c^2t^2 di equazione $d^2 = 0$. Detto diversamente, \mathfrak{S} e \mathfrak{s} fanno a questo proposito osservazioni indistinguibili.¹⁴ Si può anzi dimostrare (vedi App. Spec. 2.C) che le (2) [le (2') omogenee] sono *le sole* trasformazioni lineari ad un parametro [a sei parametri] tra coordinate che assicurano l'invarianza di cui alla (5) [di cui alla (5bis)], e che si riducono alla trasformazione identica [ad una rotazione una tantum] per $Y = 0$. Un altro fatto importante da segnalare è che l'operatore iperbolico di d'Alembert (Jean, Parigi 1717-1783) relativo alle coordinate (X, Y, Z, cT) di \mathfrak{S} , diciamo qui “Dal” in \mathfrak{S} e “dal” in \mathfrak{s} , è invariante sotto le trasformazioni (2'), ossia che Dal = dal. La dimostrazione consiste in una verifica materiale.¹⁵ Questa ulteriore implicazione delle (2') è suggestiva, perché l'equazione Dal $\Phi = 0$ per una funzione $\Phi = \Phi(X, Y, Z, cT)$ governa la propagazione ondosa, con celerità costante c , di Φ stessa. In forza delle equazioni di Maxwell nel vuoto, se Φ è una qualsiasi componente del campo elettromagnetico, le Dal $\Phi = 0$ e dal $\phi = 0$ valgono entrambe; confermando così la precedente interpretazione della costante c come *celerità assoluta* di un fotone.

Formalmente, ci si può sempre limitare a trasformazioni di Lorentz in senso stretto (uguagliando a zero le quattro costanti (x_0, t_0)) alla semplice condizione di sostituire le quattro coordinate con le differenze Δ tra coordinate omologhe di un dato punto-istante da quelle di un altro. Ciò corrisponde al passaggio dallo spazio affine dei punti-istanti (o eventi) cui ci siamo fin qui riferiti, al relativo spazio lineare. Diremo (2'_0) queste trasformazioni a sei parametri (le tre componenti di Y e i tre angoli della rotazione una tantum), e (2_0) le corrispondenti trasformazioni speciali ad un solo parametro. Così le (2_0) si otterranno dalle (2) semplicemente premettendo a X , T , x , t l'usuale segno di differenza Δ . Queste relazioni lineari *omogenee* (2_0) nelle differenze delle

¹⁴ In alcuni manuali si parte dalla richiesta dell'*equivalenza* delle due equazioni $D^2 = 0$ e $d^2 = 0$ (per riferimenti con origini coincidenti ai tempi $T = 0$, $t = 0$), dettata dal sopraddetto risultato sperimentale, e da questa equivalenza si “inferisce” l'invarianza di D^2 rispetto a trasformazioni di Lorentz. Si tratta evidentemente di una conclusione inaccettabile se non la si sostiene con altre ragionevoli argomentazioni.

¹⁵ Cioè Dal =: $(\partial^2/\partial X^2 + \partial^2/\partial Y^2 + \partial^2/\partial Z^2) - \partial^2/\partial (cT)^2$ con c costante, e similmente per dal con le variabili minuscole. Considerando le sole coordinate spaziali (X, x) parallele al moto, risulta $\partial_x = g[\partial_X + (Y/c^2)\partial_T]$ e $\partial_t = g(\partial_T + Y\partial_X)$; quindi $\partial_x^2 = g^2[\partial_X^2 + 2(Y/c^2)\partial_X\partial_T + (Y^2/c^4)\partial_T^2]$, $\partial_t^2 = g^2(\partial_T^2 + 2Y\partial_T\partial_X + Y^2\partial_X^2)$. Con Dal = $\partial_X^2 - (1/c)^2\partial_T^2$, e similmente per dal, la tesi segue immediatamente. Non contando la distinzione tra variabili maiuscole e minuscole, la notazione usuale per Dal \equiv dal è \square .

coordinate attestano alcuni fatti caratteristici e fisicamente importanti, e cioè: (i) la cosiddetta “natura relativa della simultaneità”; (ii) la cosiddetta “dilatazione alla Lorentz delle durate”; (iii) la cosiddetta “contrazione alla Lorentz (e FitzGerald (George, 1851-1901)) delle lunghezze parallele al moto”. La loro discussione, elementare dal punto di vista matematico, presenta qualche aspetto delicato da quello interpretativo, e segue qui appresso sotto (1), (2), e (3).

Sotto l’ipotesi $\Upsilon \neq 0$, siano \mathfrak{S} e \mathfrak{s} i due soliti osservatori inerziali, il primo considerato convenzionalmente come “osservatore di quiete” e il secondo come “osservatore di moto”. Abbiamo allora:

§1. Natura relativa della simultaneità. Si considerino due eventi E_1 ed E_2 simultanei rispetto a \mathfrak{S} , quindi $T_1 = T_2$. Rispetto a \mathfrak{s} , $t_1 = g(T_1 - X_1\beta/c)$, e similmente $t_2 = g(T_2 - X_2\beta/c)$, ove β è un’utile abbreviazione per Υ/c (per cui $g^{-1} = (1-\beta^2)^{1/2}$). Si conclude che può essere $t_1 = t_2$ se e solo se (sse) $X_1 = X_2$, ossia sse i due eventi si verificano nello stesso punto per \mathfrak{S} , e quindi sono a tutti gli effetti lo stesso evento. In quest’ultimo caso è poi $x_1 - x_2 = \Upsilon g (T_2 - T_1) = 0$; ossia gli eventi E_1 e E_2 sono lo stesso evento anche per \mathfrak{s} . Con l’eccezione di questo caso banale, la simultaneità è dunque “relativa”: se vale per \mathfrak{S} , non vale per \mathfrak{s} supposto in moto rispetto a \mathfrak{S} ($\Upsilon \neq 0$), e viceversa. §

A questo effetto di relatività della simultaneità si possono aggiungere due altre conseguenze delle trasformazioni speciali di Lorentz (2) (con $\Upsilon \neq 0$). Siano (X_1, T_1) e (X_2, T_2) le coordinate degli eventi E_1 e rispettivamente E_2 secondo \mathfrak{S} , e similmente (x_1, t_1) e (x_2, t_2) quelle degli stessi eventi secondo \mathfrak{s} , e poniamo come anzidetto $\Delta X =: X_2 - X_1$, $\Delta T =: T_2 - T_1$, $\Delta x =: x_2 - x_1$, $\Delta t =: t_2 - t_1$. Le quattro variabili ΔX , ΔT , Δx , Δt sono dunque legate dal sistema omogeneo (2₀) (oppure dal suo inverso che si ottiene dal precedente scambiando le maiuscole con le minuscole e invertendo il segno di Υ). Questo legame è non singolare, cioè permette di ricavare due qualsiasi delle sopraddette quattro variabili avendo assegnato ad arbitrio le altre due. Ad esempio, assegnando ΔX e Δt si ha $\Delta T = \Delta X\beta/c + \Delta t/g$ e $\Delta x = \Delta X/g - \Upsilon\Delta t$. In particolare una (ma non entrambe, pena banalità delle conclusioni) delle due variabili assegnate può supporre nulla. Se si assegna $\Delta X = 0$ [$\Delta x = 0$] i due eventi sono colocali in \mathfrak{S} [in \mathfrak{s}]; se si assegna $\Delta T = 0$ [$\Delta t = 0$] i due eventi sono simultanei in \mathfrak{S} [in \mathfrak{s}]. Vi sono in tutto dodici possibilità di questo tipo: $\Delta X = 0$ e Δx , aut ΔT , aut Δt dato ad arbitrio, e così via per i tre analoghi casi $\Delta T = 0$, ecc. Poiché tuttavia vi è simmetria tra \mathfrak{S} e \mathfrak{s} (cambia soltanto il segno di Υ), si è ridotti a soli sei casi a priori diversi. Questi sono: (I₁) $\Delta X = 0$ e ΔT dato; (I₂) $\Delta X = 0$ e Δt dato; (II₁) $\Delta t = 0$ e Δx dato; (II₂) $\Delta t = 0$ e ΔX dato; (III₁) $\Delta X = 0$ e Δx dato; (III₂) $\Delta t = 0$ e ΔT dato. Nel caso (I₁) [nel caso (I₂)] abbiamo $\Delta x = -\Upsilon g\Delta T$ [$\Delta x = -\Upsilon\Delta t$]; e in entrambi i casi (I₁) e (I₂), $\Delta t = g\Delta T$. Si vede quindi che (I₁) e (I₂) sono equivalenti. Similmente, nel caso (II₁) [nel caso (II₂)] abbiamo $\Delta T = \Delta x g\beta/c$ [$\Delta T = \Delta X\beta/c$]; e in entrambi i casi (II₁) e (II₂),

$\Delta X = g\Delta x$. Si vede quindi che anche (II₁) e (II₂) sono equivalenti. Nel caso (III₁) abbiamo $\Delta t = -\Delta x/\Upsilon$ e $\Delta T = \Delta t/g$, quindi ricadiamo in (I₂). Infine nel caso (III₂) abbiamo $\Delta X = \Delta Tc/\beta$ e $\Delta x = \Delta X/g$, quindi ricadiamo in (II₂). In definitiva dei dodici casi inizialmente considerati solo due sono essenzialmente diversi, diciamo (I₁) e (II₂).

§2. Dilatazione alla Lorentz delle durate (caso (I₁)). Fisicamente, prescrivere ΔT sotto $\Delta X = 0$ significa, per \mathfrak{S} , misurare la durata (in questo caso $\Delta T > 0$) di un fenomeno in un punto *fisso* determinato, dove \mathfrak{S} ha il suo orologio O: ad esempio, misurare la vita media di una particella elementare immobile. Vista da \mathfrak{s} , la particella-orologio si muove con velocità costante $-\Upsilon$, mentre la sua vita media Δt risulta maggiore per un fattore $g (> 1)$ di quella misurata da \mathfrak{S} , secondo la $\Delta t = g\Delta T$. (Si noti che Δt e ΔT hanno come naturale lo stesso segno.) In definitiva, \mathfrak{s} vede lo stesso fenomeno durare di più di quanto lo vede durare \mathfrak{S} . Il tempo ΔT misurato da O è la “durata di Lorentz” del fenomeno in oggetto. §

§3. Contrazione alla Lorentz delle lunghezze (caso (II₂)). Fisicamente, prescrivere ΔX sotto $\Delta t = 0$ significa che \mathfrak{S} misura le posizioni X_1 e X_2 di due eventi E_1 e E_2 simultanei secondo \mathfrak{s} e ne fa la differenza $\Delta X = X_2 - X_1$. Siano X_1 e X_2 le posizioni degli estremi (1) e (2) di una barretta rigida fissa in \mathfrak{S} e parallela all’asse X. Poiché la barretta è fissa, X_1 e X_2 *non* dipendono dai tempi T_1 e T_2 ai quali \mathfrak{S} fa le sue misure di posizione; ma secondo la richiesta di simultaneità in \mathfrak{s} , questi tempi T_1 e T_2 sono quelli che corrispondono ad un comune ma arbitrario tempo $t_1 = t_2$ di \mathfrak{s} , e quindi sono $T_1 = t_1\sqrt{1-\beta^2} + X_1\beta/c$ e $T_2 = t_2\sqrt{1-\beta^2} + X_2\beta/c$, per cui $T_2 - T_1 \equiv \Delta T = (X_2 - X_1)\beta/c \equiv \Delta X\beta/c$; vale cioè, come deve, la prima relazione del caso (II₂). Inoltre $\Delta t \equiv t_2 - t_1 = 0$, e quindi $x_2 - x_1 \equiv \Delta x = \Delta X/g \equiv (X_2 - X_1)/g$ secondo la seconda relazione dello stesso caso. Si noti che Δx e ΔX hanno come ci si aspetta lo stesso segno, ma la barretta appare a \mathfrak{s} più corta che a \mathfrak{S} per un fattore g^{-1} . § ¹⁶

In modo analogo a come si è passati dalle (2,2') alle (2₀,2'₀), si passerà dall’invariante nella (5) a quello che gli corrisponde nelle differenze delle coordinate; e con riferimento alla più generale (5bis), dall’invariante D^2 a quello che gli corrisponde nello stesso senso, e che scriveremo convenzionalmente $(\Delta D)^2 =: (\Delta L)^2 - c^2(\Delta T)^2$ (dove $(\Delta L)^2$ è un’abbreviazione per la somma dei quadrati delle differenze fra coordinate spaziali omologhe, invariante per rototraslazioni). È fondamentale osservare che, mentre in R^n è invariante (rispetto a rototraslazioni a n dimensioni (\equiv combinazioni di una traslazione e di una trasformazione ortogonale, non di necessità congruente)) la *somma* dei quadrati delle differenze tra coordinate omologhe, nel caso presente

¹⁶ Nei precedenti §1, §2, §3, le differenze Δ considerate sono state supposte finite; ma nulla impedirebbe di supporle infinitesime. In questo caso sarebbero possibili interpretazioni più permissive delle varie condizioni descritte; ad esempio, la particella “immobile” rispetto a \mathfrak{S} del caso §2 potrebbe essere soltanto “momentaneamente immobile” rispetto a \mathfrak{S} , ecc.

dello spazio-tempo $R^3 \times R$ lo è (rispetto a trasformazioni di Lorentz) *una delle due differenze* fra quei quadrati, ad es. quella in cui il quadrato $c^2(\Delta T)^2$ è *sottratto* e gli altri sono sommati (è ovvio il carattere convenzionale di quest'ultima scelta, che sarà anche la nostra). Un'altra diversità fondamentale sta poi nel fatto che, mentre l'invariante euclideo è intrinsecamente > 0 (salvo quando tutte le Δ si annullano), quello lorentziano può avere segno qualsiasi; o anche annullarsi *senza* che tutte le differenze omologhe si annullino, e precisamente quando i due punti-istanti in oggetto si riferiscono alle posizioni successive *di uno stesso* fotone. Più specificamente, così come è stato qui definito, $(\Delta D)^2$ è *negativo* quando i due punti-istanti sono sulla linea d'universo di un punto materiale, *nullo* quando i due punti-istanti sono sulla linea d'universo di un fotone, e altrimenti *positivo* (cioè quando i due punti-istanti non sono né sulla linea d'universo di un punto materiale né su quella di un fotone). Insomma, mentre l'invariante euclideo è definito positivo, quello lorentziano è indefinito. Spazi lineari a **metrica non degenerare indefinita**, come appunto lo spazio-tempo della **relatività speciale** (o **ristretta**), sono quelli che in algebra si dicono **pseudoeuclidei**.

Tornando ancora alle trasformazioni di Lorentz speciali (equazioni (2)), esse si chiamano anche **rotazioni immaginarie**, e per la semplice ragione che segue. Ponendo $W =: icT \equiv iU$, con $i =$ unità immaginaria, l'invariante canonico (5) si scrive $X^2 + W^2$, e quindi coincide formalmente con il quadrato della distanza dall'origine del piano (X,W) pensato come reale. Le più generali trasformazioni lineari che lasciano invariata questa distanza sono le rotazioni del piano stesso, che si esprimono per mezzo di ben note formule in termini delle funzioni circolari $\sin\varphi$, $\cos\varphi$, ove φ è l'angolo della rotazione. Tornando alle vecchie coordinate (X,U) , e richiedendo il carattere reale delle trasformazioni, si vede subito che questo è possibile sse l'angolo di rotazione φ è immaginario. Posto dunque $\varphi = i\theta$, ciò fa sì che le funzioni iperboliche Sh , Ch prendano il posto delle corrispondenti funzioni circolari, finendo con le trasformazioni (2) scritte nella forma (con $u =: ct$)

$$(6) \quad x = XCh\theta - USh\theta, \quad u = -XSh\theta + UCh\theta,$$

o equivalentemente, le stesse cambiando il segno di θ

$$(6bis) \quad x = XCh\theta + USh\theta, \quad u = XSh\theta + UCh\theta.$$

Si vede subito che il determinante delle (6) o delle (6bis) è uguale ad 1, proprio come quello della rotazione reale dell'angolo φ . Sempre con riferimento alle (6) [(6bis)], la grandezza adimensionale θ è determinata dalla ovvia richiesta $\Upsilon = cTh\theta$ [$\Upsilon = -cTh\theta$]; fatte le sostituzioni, si ritrovano immediatamente le (2). L'espedito formale illustrato (che può avere qualche utilità euristica nella discussione di problemi meno elementari) è suggerito dalla considerazione comparata delle due

identità $\sin^2 + \cos^2 = 1$ e $\text{Ch}^2 - \text{Sh}^2 = 1$. Nell'App. Spec. 2.C le (6,6bis) sono discusse più a fondo, e la loro correttezza è dimostrata esaurientemente.

Le trasformazioni (2) si possono anche rappresentare graficamente sovrapponendo al piano cartesiano ortogonale (X,U) un piano (x,u) con una opportuna coordinatazione cartesiana obliqua (con la stessa origine), in modo che lo stesso punto sulla carta abbia alternativamente le coordinate maiuscole nel primo piano coordinato e quelle minuscole nel secondo. È facile riconoscere che la coordinatazione obliqua in questione ha assi ruotati in senso opposto, rispetto agli assi originali, dell'angolo $\arctg Y/c$, in senso antiorario per $X \rightarrow x$ se l'angolo retto dall'asse X all'asse U è orientato in senso antiorario (come è usuale), e unità date dalle intercette dei nuovi assi con le due iperboli unitarie coniugate $X^2 - U^2 = \pm 1$.

Diamo per finire le relazioni tra le accelerazioni nei due riferimenti inerziali \mathfrak{S} ed \mathfrak{s} relative alle solite trasformazioni speciali di Lorentz (analogamente a quanto abbiamo fatto con le (4) per le velocità). Il calcolo non presenta difficoltà, e si trova:

$$(7) \quad a_x = \alpha^3 g^{-3} A_x, \quad a_y = \alpha^2 g^{-2} (A_y + \alpha A_x V_y Y/c^2), \quad a_z = \alpha^2 g^{-2} (A_z + \alpha A_x V_z Y/c^2),$$

dove α è ancora $(1 - V_x Y/c^2)^{-1}$. Se in particolare il punto in oggetto è fermo rispetto ad \mathfrak{S} (pur subendo l'accelerazione A), e quindi $V = 0$, $\alpha = 1$, le (7) si riducono alle più semplici

$$(7') \quad a_x = g^{-3} A_x, \quad a_y = g^{-2} A_y, \quad a_z = g^{-2} A_z.$$

2.2.2) RUDIMENTI DI DINAMICA RELATIVISTICA SPECIALE

Va da sé, da tutto quanto precede, che se le trasformazioni di Lorentz hanno posto rimedio al dilemma cinematico della costanza della celerità di un fotone in due riferimenti inerziali (quindi in moto rettilineo uniforme l'uno rispetto all'altro), esse introducono *una nuova difficoltà* per quanto riguarda la legge dinamica, la quale dovrà essere necessariamente modificata in modo opportuno. Infatti, considerando la massa del punto, nonché la forza agente su di esso, come invarianti nel passaggio da un riferimento inerziale ad un altro, e supposta la legge dinamica valere in \mathfrak{S} , essa non vale più in \mathfrak{s} ; e viceversa. Altrimenti detto, il modello dinamico di Newton per il moto di un punto materiale, invariante (o covariante) rispetto a trasformazioni galileiane, non lo è più, come era ben da attendersi, rispetto a trasformazioni di Lorentz. Occorrerà insomma *elaborare un nuovo modello dinamico* che soddisfi al già accennato **principio di relatività speciale einsteiniana**, e secondo il quale *tutti* i modelli non coinvolgenti fenomeni gravitazionali, e non soltanto il modello cinematico del moto di un punto (punto geometrico, cioè non necessariamente materiale), devono essere invarianti rispetto a trasformazioni di Lorentz.

Come nella elaborazione di ogni modello fisico, anche in questo caso dovranno farsi equilibrio il punto di vista osservativo e quello euristico. Supponiamo dunque che, secondo l'osservatore inerziale \mathfrak{S} , il punto materiale P di massa M e soggetto ad una forza F si muova con una certa legge $X = X(T)$ (ove X sta ora per il *vettore* di posizione), quindi con velocità $V = dX/dT = V(T)$. Senza ledere la generalità, potremo sempre scegliere \mathfrak{S} in modo che sia $V(T=0) = 0$ e $X(T=0) = 0$. Sotto il primo vincolo, diremo \mathfrak{S} osservatore (o riferimento) **di quiete istantanea per P**. *Supporremo* che il modello fisico per \mathfrak{S} , a $T = 0$, sia ancora quello di Newton (legge fondamentale della dinamica), cioè che

$$(1) \quad MA(T=0) - F(T=0) = 0,$$

dove M, A, F sono la massa di P, l'accelerazione di P e la forza su P secondo \mathfrak{S} . Le coordinate (x,t) dell'osservatore \mathfrak{s} siano legate a quelle (X,T) di \mathfrak{S} mediante le (2.2.1, 2), quindi $x(t=0) = 0$ e $v(t=0) = (-Y, 0, 0)$. *Converremo* inoltre che \mathfrak{s} attribuisca a P la stessa massa che gli attribuisce \mathfrak{S} , $m = M$. Poiché $(X=0, T=0) \Leftrightarrow (x=0, t=0)$, i due eventi "presenza di P in $X = T = 0$ " e "presenza di P in $x = t = 0$ " sono per definizione lo stesso evento. Sia poi f la forza agente su P secondo \mathfrak{s} .

Ci si chiede quale debba essere la controparte, diciamo

$$(2) \quad \mathfrak{N}(m, a(t=0), f(t=0)) = 0$$

secondo l'osservatore inerziale \mathfrak{s} , della (1), in modo che le due leggi siano equivalenti, ossia che $(1) \Leftrightarrow (2)$. Si tratta dunque di determinare (se esiste) la funzione \mathfrak{N} ; ed è ovvio che il problema *non può avere una sola risposta* a questo livello di generalità. Nel seguito sottintenderemo per brevità che $T = 0$ in A e F, e similmente che $t = 0$ in a e f. Nella (2), conosciamo almeno l'espressione di a, che è $a = (g^{-3}A_X, g^{-2}A_Y, g^{-2}A_Z)$, cfr. (2.2.1, 7'). Ragionevolmente, *supporremo* allora che la funzione \mathfrak{N} sia del tipo $m\Lambda(a) - f$, ove Λ è una funzione *lineare* di a. Un'idea naturale sembrerebbe quella di porre

$$(3) \quad \Lambda(a) = (g^3 a_x, g^2 a_y, g^2 a_z) (= m^{-1}f),$$

che condurrebbe semplicemente a $f = F$. Benché logicamente accettabile, questa posizione *non è in accordo con l'esperienza*, e in particolare con quanto ci insegna l'elettromagnetismo. Si vedrà infatti che la forza elettromagnetica su una carica, diciamo Ω e rispettivamente ω , cambia da \mathfrak{S} a \mathfrak{s} secondo la legge

$$(4) \quad \omega = (\Omega_X, g^{-1}\Omega_Y, g^{-1}\Omega_Z),$$

cfr. (2.D, 5bis). Una relazione analoga deve dunque esistere tra f e F, perché si può sempre immaginare una situazione in cui forze generiche siano equilibrate da forze elettromagnetiche. In conclusione, *ammetteremo* che

$$(4bis) \quad f = (F_X, g^{-1}F_Y, g^{-1}F_Z);$$

e questo ci porta a correggere, nella (3), i due g^2 in due g . Otteniamo insomma per \aleph la funzione $m(g^3 a_x, g a_y, g a_z) - f$, che in \mathfrak{S} diventa $mA - F$. Evidentemente, essa è tale che $\aleph = 0$ equivale alla (1). In definitiva, il moto di P secondo \mathfrak{s} soddisferà alla

$$(5) \quad m(g^3 a_x, g a_y, g a_z) = f,$$

dove il vettore f è legato a F come ω è legato a Ω , cioè secondo la (4bis). Un aspetto rassicurante di questa scelta è nel fatto che, come si dimostra con un calcolo senza difficoltà, il primo membro della (5) è uguale a $m d_t(\gamma v)$, dove $\gamma(t) =: (1 - v^2(t)/c^2)^{-1/2}$ e $v = (v_x, 0, 0)$ è la velocità di P in \mathfrak{s} . Quindi possiamo riscrivere la (5) nella semplice forma

$$(6) \quad m d_t(\gamma v) = f.$$

Questa (6) è da riferirsi ad un generico osservatore inerziale rispetto al quale il punto mobile di massa m ha velocità $v = v(t) = d_t x(t)$. Qui γ è la precedente funzione di t attraverso $v^2(t)$, e f (**forza relativistica**) è assunta come prescritta funzione abbastanza regolare di (t, x, v) . L'equazione differenziale (6) è dello stesso tipo dell'equazione di Newton (cui si riduce per $c^2 \rightarrow \infty$), fatta eccezione per il carattere normale, che è andato momentaneamente perduto. Tale carattere si restaura tuttavia senza difficoltà, ottenendo un sistema di due equazioni differenziali normali del primo ordine in (x, v) , e precisamente $d_t x = v \equiv \Psi_1(t, x, v)$, e $d_t v = m^{-1} \gamma^{-1} [\mathbf{I} + \gamma^2 v v / c^2]^{-1} \cdot f(t, x, v) \equiv \Psi_2(t, x, v)$, dove \mathbf{I} è la diade unitaria, $[\dots]^{-1}$ è la matrice inversa di quella scritta entro le $[\]$ (simmetrica come quest'ultima) e (\cdot) denota il prodotto matriciale standard di $[\dots]^{-1}$ per f (pensata come colonna). Il determinante della matrice da invertire, nonché la matrice inversa, si calcolano elementarmente, il primo risultando uguale a $\gamma^2 > 1$. Il sistema differenziale $d_t x = \Psi_1(t, x, v)$, $d_t v = \Psi_2(t, x, v)$, ha quindi una e una sola soluzione sotto le usuali condizioni accessorie e le usuali richieste di regolarità sulla funzione $f(t, x, v)$. Si noti anche che, come funzione di $|v|$, $\gamma(|v|)$ è uguale alla $g(|Y|=|v|)$.

Si vede immediatamente, infine, che v non dipende dal tempo se $f \equiv 0$, ossia che vale la legge d'inerzia. Infatti, dalla $\gamma v = \sigma$, dove σ è un vettore indipendente da t , con manipolazioni banali si trae $v^2/c^2 = (1 + c^2/\sigma^2)^{-1}$; quindi anche v^2 , e perciò γ , è indipendente da t , e in definitiva lo è $v = \sigma/\gamma$. Detto diversamente, e da un punto di vista più generale, se c^2 è una costante positiva qualsiasi, e se $v^2 \equiv \sum_i v_i^2 < c^2$ (con $i = x, y, z$), i sistemi differenziali $d_t v_i = 0$ e $d_t [v_i / (1 - \sum_i v_i^2 / c^2)^{1/2}] = 0$ sono equivalenti.

Alla legge dinamica relativistica speciale (6) si può dare una veste formalmente (ma anche, in parte, concettualmente) diversa, e che forse aderisce meglio alla natura del fatto fisico. Torniamo innanzitutto all'invariante lorentziano, che scriveremo (per l'osservatore inerziale \mathfrak{s}) $(\Delta d)^2 = (\Delta l)^2 - c^2 (\Delta t)^2$, e riferiamolo allo spostamento di un punto *materiale* P (cioè per $(\Delta l)^2 =$ somma

dei quadrati degli spostamenti coordinati di P (secondo \mathfrak{s}) e $(\Delta t)^2 =$ quadrato del tempo trascorso, secondo gli orologi di \mathfrak{s}). Sapendo che l'invariante $(\Delta d)^2$ è ora certamente negativo, lo scriveremo come $-c^2(\Delta\tau)^2$, dove $\Delta\tau$ è una quantità con le dimensioni di un tempo, per il momento definita a meno del segno, per cui $c^2(\Delta\tau/\Delta t)^2 = -(\Delta l/\Delta t)^2 + c^2$. Al tendere di Δt a zero, il secondo membro diventa la quantità positiva $c^2 - v^2(t)$, mentre il primo membro si deve interpretare come c^2 volte il quadrato della derivata rispetto a t di una funzione $\tau(t)$ definita, a meno di una costante additiva che per i nostri scopi può essere presa uguale a 0, come l'integrale da 0 a t della funzione $\gamma^{-1}(t) < 1$. Ciò corrisponde ad aver scelto la determinazione *positiva* per la derivata $d\tau/dt$ nella $(d\tau/dt)^2 = \gamma^{-2}(t)$. $\tau(t)$ è dunque un tempo funzione *dispari* di t determinata in termini di $v^2(t)$, che è 0 per $t = 0$, cresce monotonamente con t (ma più lentamente di t) e si conserva in valore assoluto minore di t stesso per $t \neq 0$. Essendo così τ in relazione biunivoca con t , possiamo descrivere il moto di P in funzione di τ invece che di t . Tale τ viene detto **tempo proprio di P**. La velocità di P in termini di tempo proprio è, se x ne denota la posizione, $u = d_\tau x = u(\tau) = \gamma(t)v(t)$, e viene detta **velocità propria di P** (al tempo τ). Sottolineiamo che l'incremento infinitesimo $d\tau$ è quello misurato da un orologio in moto rettilineo uniforme, ma istantaneamente in quiete rispetto a P, o come anche si dice, **in moto tangente a quello di P**. $d\tau$ coincide dunque istante per istante con la durata infinitesima di Lorentz corrispondente al valore di $|v|$, secondo la $d\tau = \gamma^{-1}(|v|)dt$. Se supponiamo che esista un tipo di orologio il cui ritmo di marcia non è turbato dalle accelerazioni che esso eventualmente subisce (potremo chiamare questi orologi "a prova di accelerazione"), allora τ potrà pensarsi come misurato da un tale orologio che viaggi "incollato" a P, e che segni 0 quando $t = 0$ ¹⁷. Questa interpretazione di τ è quella normalmente accettata, perché è semplice e intuitiva, ma non è indispensabile seguirla nel presente contesto. Analogamente alla velocità propria $u(\tau)$, in modo simile si potrà definire una **accelerazione propria di P**, $w(\tau)$, secondo $w = w(\tau) =: d_\tau u = d_t(\gamma v)d_\tau t = \gamma d_t(\gamma v)$. Alla luce della (6), potremo dunque scrivere la nuova versione della legge dinamica nella forma

$$(7) \quad mw = md_\tau^2 x \equiv k =: \gamma f.$$

La forza k si dice **forza di Minkowski** (Hermann, Alexotas Lituania 1864, Gottinga 1909) ed è proporzionale a f secondo il fattore (dipendente da v^2) γ . Si osservi che nel solito limite $v^2/c^2 \rightarrow 0$, il tempo proprio si riduce a t e γ diventa uguale a 1 (quindi la forza di Minkowski coincide con la forza f). Infine la lunghezza $c\tau$ (misurata dall'orologio in P) si potrà dire **lunghezza propria di P**.

¹⁷ Gli orologi "a prova di accelerazione" sono strumenti ideali, nel senso in cui lo sono gli orologi puntiformi o i regoli rigidi, e si possono pensare approssimati da corrispondenti strumenti reali con la precisione conveniente alla situazione di interesse. Una nota (e abusata) facezia didattica menziona il caso dell'orologio reale (\equiv non a prova di accelerazione) che cadendo a terra si rompe.

Come dall'equazione vettoriale di Newton (legge dinamica) si deduce, per moltiplicazione scalare per v , l'equazione scalare (della potenza) $md_t(v^2/2) = v \cdot f$, analogamente dalla (6) si ottiene una corrispondente **equazione** (scalare) **della potenza relativistica**, secondo la

$$(8) \quad md_t(\gamma c^2) = v \cdot f.$$

La (8) si giustifica mediante alcuni facili passaggi partendo dalla $d\gamma = \gamma^3 v \cdot dv/c^2$. Definendo un'energia \mathfrak{R} , attraverso la (8), come quello scalare la cui t-derivata è $v \cdot f$ e che si annulla per $v = 0$, si trova $\mathfrak{R} = mc^2(\gamma - 1)$. È tuttavia più conveniente introdurre l'energia $e =: \mathfrak{R} + mc^2 = m\gamma c^2$ (**energia totale**), per mezzo della quale la (8) assume la semplice forma $d_t e = v \cdot f$. In prima approssimazione, se v^2 è molto più piccola di c^2 , $\gamma - 1 \approx v^2/2c^2$, per cui $\mathfrak{R} \approx mv^2/2$; vale a dire, in questa approssimazione \mathfrak{R} è semplicemente l'energia cinetica classica del punto, $mv^2/2$.

In analogia con il caso classico, definiremo la **quantità di moto (QdM) relativistica** del punto come $p =: m\gamma v$. Nel solito riferimento inerziale \mathfrak{S} , rispetto al quale \mathfrak{s} trasla uniformemente con velocità Y (valore e segno) lungo l'asse X , mediante manipolazioni elementari si trova che la QdM relativistica di P e l'energia totale E (rispetto ad \mathfrak{S}), cioè $m\Gamma V$ e rispettivamente $m\Gamma c^2$, ove $\Gamma =: (1 - V^2/c^2)^{-1/2}$ è l'“omologo maiuscolo” di γ , sono legate a (p, e) dalle relazioni

$$(9) \quad p_x = g(P_X - Y E/c^2), \quad p_y = P_Y, \quad p_z = P_Z, \quad e = g(E - Y P_X).$$

Quindi le coppie $(p, e/c)$ e $(P, E/c)$ sono legate in modo formalmente identico a come lo sono le coppie (x, ct) e (X, cT) nelle trasformazioni di Lorentz speciali.¹⁸ Conseguenza immediata di ciò è che le differenze $p^2 - e^2/c^2$ e $P^2 - E^2/c^2$ sono uguali, ovvero che $p^2 - e^2/c^2$ (riferendoci ad \mathfrak{s}) è un invariante rispetto a trasformazioni di Lorentz.

Poiché la legge dinamica relativistica (6) può scriversi nella forma equivalente

$$(10) \quad d_t p = f,$$

e poiché conosciamo come si trasforma p passando da \mathfrak{s} a \mathfrak{S} , potremo scrivere la legge dinamica in \mathfrak{S} come $d_T P = d_T d_t P = F$ (dove F è la forza vista da \mathfrak{S}), e con essa la sua conseguenza di potenza come $d_T E = V \cdot F$. Tenendo conto della

$$(11) \quad d_t T = g(1 + Y v_x/c^2),^{19}$$

ricaviamo la legge di trasformazione delle forze, che risulta essere:

$$(12_1) \quad f_x = F_X - \alpha Y (F_Y V_Y + F_Z V_Z),$$

$$(12_2) \quad f_{y,z} = \alpha g^{-1} F_{Y,Z}.$$

Se in particolare il punto P è in *quiete istantanea* rispetto al riferimento \mathfrak{S} ($V = 0$), allora $\alpha = 1$, e le

(12) si semplificano nelle

¹⁸ Si noti che il rapporto $p : (e/c)$ è adimensionale, come il rapporto $x : ct$.

¹⁹ La (11) si ottiene immediatamente scrivendo la trasformazione inversa della seconda (2.2.1, 2), $T = g(t + Yx/c^2)$, e derivandola rispetto a t .

$$(12\text{bis}) \quad f_x = F_X, \quad f_y = g^{-1}F_Y, \quad f_z = g^{-1}F_Z,$$

o meglio nelle stesse in cui le componenti maiuscole siano affette da un superscritto ($^{\circ}$) con il quale ricordiamo il loro essere relative al riferimento di quiete (a questa norma notazionale ci atterremo spesso nel seguito). Ritroviamo quindi la legge di trasformazione (4bis), che governa il passaggio da F a f .

Il prodotto $m\gamma(|v|)$, per un punto materiale di massa m in moto con velocità v , si dice suggestivamente **massa relativistica** o **massa di moto** di quel punto, mentre m si dice anche, per contrapposizione, sua **massa classica** o **di quiete**. È vantaggioso, e molto più comune, scrivere m come m° e γm° come m . *Adottando questa nuova notazione*, la QdM relativistica è mv , e la legge dinamica relativistica diventa semplicemente

$$(6\text{bis}) \quad d_t(mv) = f,$$

mentre l'equazione della potenza relativistica diventa

$$(8\text{bis}) \quad d_t(mc^2) = v \cdot f. \quad {}^{20}, {}^{21}$$

Sviluppando la derivazione nella (6bis) si trova $v dm/d|v| + m d_t v = f$; e quindi moltiplicando scalarmene questa per il versore tangente, e rispettivamente per il versore normale principale, alla traiettoria, e denotando con $_t$ e rispettivamente $_n$ le corrispondenti componenti, otteniamo

$$(13_t) \quad (|v|dm/d|v| + m)a_t = f_t$$

$$(13_n) \quad ma_n = f_n,$$

²⁰ Nella nuova notazione l'energia totale è scritta come $e = mc^2$. Anche grazie alla sua estrema semplicità formale, la $e = mc^2$ è certamente la formula più popolare, una vera e propria icona, della fisica moderna (se mai ne esiste una). L'importanza, che ad essa a ragione si attribuisce, appare tuttavia *non del tutto giustificata all'interno del presente contesto*. Dopotutto, quello che qui conta sono le *variazioni* dell'energia totale mc^2 ("energia di quiete" $m^{\circ}c^2$ + "energia di moto" $(m - m^{\circ})c^2$) rispetto alla $m^{\circ}c^2$, e non il valore di quest'ultima costante: l'esistenza stessa di una tale energia di quiete, eventualmente disponibile per altre trasformazioni, potrebbe essere illusoria. L'equivalenza (cioè l'uguaglianza a meno dell'unità di misura) tra massa ed energia, suggerita dagli esperimenti sulla radioattività naturale, nonché artificiale a partire dal 1934, rimase una ragionevole ma audace congettura per molto tempo dopo la creazione della relatività speciale; ancora nei primi anni '30, lo stesso Einstein non ne era del tutto persuaso. In realtà la somma delle masse-energie dei costituenti di una molecola, di un atomo o di un nucleo atomico (presi separatamente) è diversa dalla massa-energia del sistema composto come un tutto, perché una parte di quella somma è trasformata in **massa-energia di legame**. Le variazioni di massa-energia di composti chimici (molecole o atomi) sono difficilmente rivelabili per la loro piccolezza relativa; ma lo sono quelle di composti nucleari, quando sono in gioco masse-energie di legame in proporzione assai più grandi. Accertati questi fatti, ed accettato l'ordine di idee che ne consegue, la sola ragione per distinguere tra massa ed energia si riduce allora alla scelta dell'unità di misura. Va anche ricordato che mentre la *proposta/ipotesi*, da parte di Einstein, della completa equivalenza tra massa e energia risale allo stesso 1905 (Ann. d. Phys. **18**, 639; vedi anche in Jahrbuch der Radioaktivität u. Elektronik **4**, 1907), le prime chiare conferme sperimentali arriveranno quasi venticinque anni dopo. I lavori einsteiniani appena citati hanno dunque un'importanza confrontabile con quella dell'assai più noto e contemporaneo "Zur Elektrodynamik bewegter Körper".

²¹ Denotando qui con f° la forza agente sul punto materiale nel limite galileiano, in questo limite alla (8bis) corrisponde la (*) $m^{\circ}d_t v^2/2 = v \cdot f^{\circ}$ (che si ottiene moltiplicando scalarmente per v la legge dinamica newtoniana $m^{\circ}d_t v = f^{\circ}$). Benché f sia generalmente diversa da f° per $v \neq 0$, come sappiamo $v \cdot f = v \cdot f^{\circ}$, e quindi la (8bis) potrebbe anche scriversi come ($^{\circ}$) $m^{\circ}c^2 d_t \gamma = v \cdot f^{\circ}$. Ci si guardi tuttavia dal pensare che $c^2 d_t \gamma = d_t v^2/2$. Di fatto $c^2 d_t \gamma = \gamma^3 d_t v^2/2$; vale a dire, la ($^{\circ}$) e la (*) sono leggi *diverse* (fermo restando che la ($^{\circ}$) confluisce nella (*) per $c \rightarrow \infty$).

dove a sta al solito per $d_t v$. Infatti $v_t = |v|$, $a_t = d_t |v|$ e $v_n = 0$. Calcolando $dm/d|v|$, si vede che la (13_t) può risciversi sostituendo al contenuto della parentesi a 1° membro la massa $m\gamma^2 \equiv m_l > m$, che si dice massa “longitudinale” del punto. Segue da ciò che i vettori a e f *non* sono paralleli come nella dinamica classica (allorché $\gamma = 1$ e $m_l = m = m^0$). Per analogia, la massa di moto m è anche detta “massa trasversale”.²²

Vogliamo ancora sottolineare come, non contando *l'induzione* della legge dinamica relativistica (6) – un punto di importanza centrale nella elaborazione del modello relativistico – a tutti gli altri risultati riportati si perviene, a partire dalle trasformazioni di Lorentz, mediante manipolazioni (algebriche o differenziali) elementari. In altre parole, le procedure *deduttive* connesse allo sviluppo della relatività ristretta sono, contrariamente a quelle *induttive*, sostanzialmente banali.

Ci siamo fin qui limitati alla dinamica relativistica speciale di un *unico* punto materiale, anche perché il passaggio a sistemi di *più* punti solleva alcune tipiche e serie difficoltà. Dalla seconda legge di Newton (2.1.2, 1) unita ai principi di sovrapposizione degli effetti dinamici e di azione e reazione, la dinamica degli insiemi discreti di punti materiali, e al limite dei sistemi materiali continui, si sviluppa in modo assai naturale. In particolare, una fondamentale direttrice di approfondimento e generalizzazione della dinamica classica dei sistemi materiali è quella della cosiddetta “dinamica lagrangiana/hamiltoniana”. Come vedremo nel Cap. 6, questa si occupa di sistemi materiali con un numero finito di “gradi di libertà” soggetti a sollecitazione “conservativa e posizionale (o posizionale-cinetica)”. Orbene, una soddisfacente versione relativistica – ispirata alla (6) invece che alla (2.1.2, 1) – di tale dinamica lagrangiana/hamiltoniana non esiste al presente, almeno in senso generale e sistematico; ed è facile intuirne il perché. Ad esempio, l’affermare che la forza agente ad un certo istante su un certo punto del sistema dipende dalle posizioni (e possibilmente dalle velocità) degli altri punti allo *stesso istante*, presuppone l’esistenza di una “azione a distanza” (\equiv propagazione con velocità infinita di effetti fisici), evidentemente incompatibile con un principio fondamentale della relatività. Invece una versione relativistica della dinamica dei continui – sempre ispirata alla (6) – è possibile, almeno finché l’interazione di un piccolo volume materiale con gli altri è supposta limitarsi ai piccoli volumi *a diretto contatto* con esso. A certi aspetti della meccanica relativistica degli insiemi di punti si accennerà nella App. Spec. 2.E, e a quella dei continui nella App. Spec. 2.H.

Concludiamo osservando che tutto quanto precede è permesso fino a che le masse dei punti materiali, o la densità di massa nel caso di un continuo materiale, sono abbastanza piccole; al di là

²² A vero dire, le espressioni “massa longitudinale” e “massa trasversale”, nonché le relative nozioni – originariamente proposte da Lorentz (1904) – sono oggi un po’ obsolete.

di questa condizione, che è resa quantitativamente precisa dalle “equazioni di campo” di Einstein-Hilbert (v. S.sez. 9.1.3), si entrerebbe a pieno titolo nel dominio della relatività generale. Nella relatività generale, come vedremo, la massa (o piuttosto la densità di massa) influisce sulla metrica dello spazio-tempo (determinandola sotto convenienti condizioni accessorie); quindi lo spazio-tempo non è più pseudoeuclideo, ma una vera e propria “varietà pseudoriemanniana”, e la legge di moto viene ulteriormente generalizzata.

2.3) INTRODUZIONE ALLA GEOMETRIA PSEUDOEUCLEIDEA

2.3.1) GLI SPAZI PSEUDOEUCLEIDEI (NOZIONI DI BASE)

Con questa sottosezione forniamo un insieme di nozioni di prima necessità sull'algebra degli spazi R-lineari (\equiv lineari su R) finito-dimensionali pseudoeuclidei, propedeutiche alla comprensione dell'algebra (dello spazio) di Minkowski o dell'algebra affine ad essa associata. Il lettore interessato ad un quadro organico della materia è rinviato ai trattati di algebra lineare/multilineare.

Nella S.sez. 1.3.1 abbiamo introdotto la nozione di spazio R-lineare E fondandola sugli otto assiomi (a÷h) ¹. Questi assiomi sono stati suddivisi in due insiemi di quattro + quattro, il primo (a÷d) relativo alla proprietà di E di essere un gruppo additivo, e il secondo (e÷h) alle proprietà del prodotto di un numero reale per un elemento di E. Un insieme *equivalente* di sette assiomi, che si ottiene da (a÷h) abolendovi (c) e (d) e aggiungendo l'assioma di "ruolo dello zero di R", $\forall x(\in E)\{0x = 0\}$, ² e che denoteremo come (a1÷a7), è usato da H. Weyl, J. Von Neumann, ecc.

Su un tale spazio R-lineare possiamo immediatamente introdurre le seguenti nozioni. Sia $\{x_i\}_{i=1,\dots,k}$ una famiglia di $k \geq 1$ elementi di E, e $\{\alpha_i\}_{i=1,\dots,k}$ una famiglia di altrettanti numeri reali: allora l'elemento $\sum_{i=1}^k \alpha_i x_i$ di E si dice la **combinazione lineare degli $\{x_i\}$ di coefficienti $\{\alpha_i\}$** . Se per una data famiglia $\{x_i\}$ l'annullarsi di una sua *qualsiasi* combinazione lineare implica che tutti i relativi coefficienti sono nulli, gli elementi di $\{x_i\}$ si dicono **linearmente indipendenti**, e **linearmente dipendenti** in caso contrario. ³ Dire che "E è n-dimensionale" o " $\dim E = n$ " significa che in E esiste una famiglia di n elementi linearmente indipendenti, ma non ne esiste una di n+1 elementi. (Il caso $n = 0$ è ammissibile, ma privo di interesse perché allora E si riduce a $\{0\}$; per cui in pratica si assume sempre $n \geq 1$.) Se E è n-dim per un certo $n \geq 1$, ciò si esprime aggiungendo agli assiomi (a1÷a7) l'**assioma della dimensione finita**

(b) "E è finito-dim", o anche "esiste $n \geq 1$ per cui $\dim E = n$ ";

¹ Gli stessi assiomi (a÷h) definiscono uno spazio lineare su C (corpo dei complessi), o ancora più in generale su un generico corpo commutativo, o campo, K.

² Qui lo zero a primo membro è quello di R, mentre lo zero a secondo membro è quello di E. Anche nel seguito, useremo lo stesso simbolo per questi due diversi oggetti, confidando che ciò non generi equivoci.

³ È ovvio che nelle precedenti definizioni le x_i di $\{x_i\}$ possono considerarsi tutte diverse tra loro e inoltre diverse da zero senza sostanziali limitazioni di generalità (perché $\alpha 0 = 0 =$ zero di E, e ogni spazio R-lineare contiene almeno un elemento, il suo zero). Se $k = 1$ e $x_1 \neq 0$, x_1 è linearmente indipendente, se $x_1 = 0$, x_1 è linearmente dipendente.

vale a dire, gli assiomi (a1÷a7,b) (o gli stessi con **C** in luogo di **R**), definiscono completamente uno spazio **R**-lineare (o rispettivamente **C**-lineare) finito-dim. Nel seguito ci interesseremo esclusivamente a spazi **R**-lineari finito-dim.

Per un dato sottoinsieme Σ di uno spazio **R**-lineare E ($n \geq 1$)-dim, l'insieme degli elementi di E esprimibili come combinazioni lineari di elementi di Σ si dice **varietà lineare** (di E) **abbracciata da** (o **generata da**) Σ , e si indica con $\text{vect}_E \Sigma$ (o più brevemente con $\text{vect} \Sigma$ se si sottintende il riferimento a E).⁴ $\text{vect} \Sigma$ è dunque uno spazio **R**-lineare incluso in E , ovvero un **sottospazio** (SS) di E . Se $\Sigma = \text{vect} \Sigma$, Σ è un sottospazio di E ; in particolare, $\{0\}$ è un SS di E , e E è un SS di E perché $E = \text{vect} E$. Se $H, H_1, H_2 \dots$ sono SS di E , si ha

(i) $0 \in H$;

(ii) $\dim H \leq \dim E = n$,

con l'uguaglianza valida sse (se e solo se) $H = E$;

(iii) $H_1 \cap H_2$ è un SS di E .

((iii) *non* vale in generale per $H_1 \cup H_2$ al posto di $H_1 \cap H_2$!)⁵ Salvo avviso contrario, quando si parla di un SS di E si assume che la sua dim sia ≥ 1 .

Se Σ è costituito da un numero $m \leq n$ di elementi linearmente indipendenti, m uguaglia la dimensione di $\text{vect} \Sigma$. In particolare, $\text{vect} \Sigma$ si dice **retta** se ha dimensione 1, **piano** se ha dimensione 2 (se $n \geq 2$), e **m-piano** se ha dimensione $2 < m < n$. Sia C un elemento di E ma non di $\text{vect} \Sigma$ (quindi $C \neq 0$): l'insieme degli elementi di E del tipo " C + un elemento di $\text{vect} \Sigma$ " si dice la **varietà affine a** $\text{vect} \Sigma$ **e** (da essa) **dislocata di** C . Evidentemente, una tale varietà affine a $\text{vect} \Sigma$ non è un SS di E , ma per definizione ha la stessa dimensione di $\text{vect} \Sigma$. In particolare, la varietà affine (1-dim) associata ad una retta si dice **retta affine**, ecc.⁶

⁴ Con ogni evidenza, lo spazio lineare E specifica le operazioni lineari necessarie a definire la sua varietà lineare abbracciata da Σ , e ciò rende conto della presenza di E nel simbolo $\text{vect}_E \Sigma$.

⁵ Pensando ad E come ad uno spazio 3-dim, un generico SS 2-dim è un piano passante per l'origine, e un generico SS 1-dim è una retta passante per l'origine. È allora intuitivamente ovvio che l'intersezione di due piani distinti passanti per l'origine è una retta passante per l'origine, secondo la (iii).

⁶ Queste rette, piani, ecc., affini possono essere orientati in modo naturale come segue. Sulla retta affine $x = a\mu + C$, dove μ è un parametro reale, e $a \in E$ è (l'unico) elemento di cui consta Σ in questo caso, la coppia ordinata di punti $\langle x(\mu=0), x(\mu=1) \rangle = \langle C, a+C \rangle$, e con essa la retta affine in oggetto, è automaticamente orientata. Analogamente, nel piano affine $x = a\mu + bv + C$, dove μ, v sono parametri reali e a, b sono i due elementi linearmente indipendenti di E ($n \geq 2$) di cui consta Σ in questo caso, le due coppie ordinate di punti $\langle x(\mu=0, v=0), x(\mu=1, v=0) \rangle = \langle C, a+C \rangle$ e $\langle x(\mu=0, v=0), x(\mu=0, v=1) \rangle = \langle C, b+C \rangle$ individuano un ordine sui tre punti $C, a+C, b+C$ se è dato un ordine sulla coppia $\{a, b\}$, ad es. l'ordine $C \rightarrow a+C \rightarrow b+C$, e quindi determinano un orientamento sul piano affine in oggetto. Si prosegue analogamente per i 3-piani ($n \geq 3$), ecc. Si tratta di nozioni che abbiamo già introdotto e discusso a proposito di uno spazio euclideo (identificato con \mathbb{R}^n), e che qui vengono estesi ad un generico spazio lineare su \mathbb{R} .

La struttura di E , \mathbb{R} -lineare e finito-dim, può a questo punto arricchirsi introducendo un'applicazione $B: E \times E \rightarrow \mathbb{R}$ sotto i seguenti quattro assiomi (per ogni x, y, \dots di E e ogni α, β, \dots di \mathbb{R})

- (c1) B è simmetrica: $B(x,y) = B(y,x)$;
- (c2) B è distributiva rispetto al primo argomento: $B(x+y,z) = B(x,z) + B(y,z)$;
- (c3) B è associativa rispetto al prodotto del primo argomento per un reale: $B(\alpha x,y) = \alpha B(x,y)$;
- (c4) B è definita positiva: $B(x,x) \geq 0$ et $(B(x,x) = 0 \Rightarrow x = 0)$.

Si constata immediatamente che questi assiomi (c1÷c4) coincidono con gli assiomi di prodotto interno (i÷n) della S.sez. 1.3.1. Dunque il prodotto interno è un caso particolare della generica applicazione B di $E \times E$ in \mathbb{R} testé considerata.

Se oltre agli assiomi (a1÷a7,b) valgono gli assiomi (c1÷c4) la coppia ⁷ $\{E,B\}$ si dice uno **spazio hilbertiano elementare**. Uno spazio hilbertiano propriamente detto è lineare su \mathbb{C} (e non soltanto su \mathbb{R}), l'applicazione B di $E \times E$ è in \mathbb{C} (e non soltanto in \mathbb{R}), la simmetria (c1) è sostituita dalla **simmetria hermitiana** (c1bis) $B(x,y) = (B(y,x))^*$ (dove $*$ significa "complesso coniugato"), e l'assioma di dimensione finita (b) è in generale sostituito da un **assioma di dimensione numerabilmente infinita** (b_∞); e inoltre, si aggiungono un **assioma di completezza** (d) e un **assioma di separabilità** (e). Questi due ultimi assiomi (d,e) conseguono peraltro dai rimanenti se in luogo di (b_∞) vi è (b), e in tal caso si ignorano. Si noti infine che in forza di (c4) $B(x,x)$ può considerarsi il quadrato di una norma (cfr. S.sez. 1.3.1). Uno spazio hilbertiano è dunque normato, e perciò automaticamente metrico con distanza $d(x,y) = d(y,x) = [B(x-y,x-y)]^{1/2}$.

Tornando allo spazio hilbertiano elementare (a1÷a7,b,c1÷c4), considereremo adesso la possibilità di generalizzarlo in senso diverso da quello appena illustrato (cioè ad uno spazio hilbertiano propriamente detto), e che riguarda l'assioma (c4). Precisamente, fermi restando i rimanenti assiomi (a1÷a7,b,c1÷c3), all'assioma (c4) si sostituisca il più debole assioma

$$(c4bis) \quad \forall x \{ \forall y \{ B(x,y) = 0 \} \Rightarrow x = 0 \}.$$

Denoteremo ancora con $\{E,B\}$ un tale spazio \mathbb{R} -lineare finito-dim definito da (a1÷a7,b,c1÷c3,c4bis). La nozione di retta (piano, ecc) affini orientati si trasporta senza modifiche in tale $\{E,B\}$, con la precisazione che si potranno ora distinguere tre tipi di rette affini $x = a\mu + C$ (ove $a, C \in E$ e $\mu \in \mathbb{R}$) a seconda che il valore di B in (a,a) sia > 0 , < 0 , $= 0$; e analogamente, sei tipi di piani affini $x = a\mu + bv + C$ (ove $a, b, C \in E$, a e b essendo linearmente indipendenti, e $\mu, v \in \mathbb{R}$) a seconda che il valore della forma in (a,a) , e/o in (b,b) , sia > 0 , < 0 , $= 0$, .. e così via.

⁷ A rigore si dovrebbe dire una "coppia ordinata"; ma in casi come il presente, in cui gli elementi della coppia hanno natura a priori diversa, diventa superfluo dare un ordine su di essi. Questo vale del tutto in generale.

Una ulteriore generalizzazione consiste nel sostituire all'assioma (c4bis) l'ancor più debole assioma

$$(c4ter) \forall x \{ \forall y \{ B(x,y) = 0 \} \Rightarrow x \in N \},$$

dove N è il **nucleo** di B , un SS di E di dimensione $0 < \delta \leq n$ per il quale vale l'implicazione opposta a quella in (c4ter). (Se, in deroga alla definizione standard di SS, si permettesse $\delta = 0$ si tornerebbe all'assioma (c4bis)). Denoteremo ancora come $\{E, B\}$ tale spazio R -lineare finito-dim definito da (a1÷a7, b, c1÷c3, c4ter). Valgono ancora, infine, le precedenti osservazioni circa le rette (piani, ecc.) affini.

Richiamiamo ora alcune nozioni generali dell'algebra delle forme bilineari, e delle associate forme quadratiche, (reali) su uno spazio R -lineare finito-dim E . (Nel seguito ci limiteremo a considerare spazi E di questo tipo.) Una **forma bilineare su E** è un'applicazione $B: E^2 \rightarrow R$ con le seguenti proprietà ($\forall (x, y, z)$ di E e $\forall \lambda$ reale):

$$(1) \quad B(x+y, z) = B(x, z) + B(y, z) \quad \text{e} \quad B(x, y+z) = B(x, y) + B(x, z);$$

$$(2) \quad B(\lambda x, y) = B(x, \lambda y) = \lambda B(x, y).$$

Conseguenza immediata della (2) è che $B(-x, -y) = B(x, y)$. Ci interesseranno in particolare le **forme bilineari simmetriche**, per le quali cioè $B(x, y) = B(y, x) \quad \forall (x, y)$ di E (si verifica subito che tale simmetria è compatibile sia con la (1) che con la (2)). Nel seguito supporremo sempre che B sia simmetrica e soddisfi gli assiomi (c1, c2, c3).

Una **forma quadratica** su E è un'applicazione $Q: E \rightarrow R$, con le proprietà ($\forall (x, y)$ di E e $\forall \lambda$ reale):

$$(3) \quad \text{"}Q(x+y) - Q(x) - Q(y) \text{ è una forma bilineare in } x, y\text{" (evidentemente simmetrica);}$$

$$(4) \quad Q(\lambda x) = \lambda^2 Q(x).$$

Una conseguenza della (4) è che $Q(-x) = Q(x)$. Se scriviamo la forma bilineare simmetrica di cui alla (3) come $2B(x, y)$, cioè poniamo

$$(3bis) \quad 2B(x, y) =: Q(x+y) - Q(x) - Q(y),$$

troviamo immediatamente:

$$(5) \quad B(x, x) = Q(x).$$

Sotto le (3,4) la $B(x, y)$ definita dalla (3bis), oltre che essere simmetrica soddisfa *per definizione* le (1,2), e questo si traduce in proprietà cui deve soddisfare la Q per suo conto; ad esempio, con riferimento alla (1), $Q(x+y+z) = Q(x+y) + Q(y+z) + Q(z+x) - Q(x) - Q(y) - Q(z)$.

Le forme, bilineare (simmetrica) B e quadratica Q così legate si dicono **associate** tra loro, e sono univocamente definite l'una in termini dell'altra secondo le (3bis,5): ad ogni B corrisponde una e una sola Q , e viceversa. Se dunque una proprietà o un asserto vale "con riferimento a B ",

potremo anche dire che vale “con riferimento a Q”, e viceversa. In ogni caso, lo spazio E corredato della forma bilineare simmetrica B oltre che con $\{E,B\}$ si potrà denotare equivalentemente $\{E,Q\}$ se B e Q sono associate.

Se (per i dati E, B) $B(x,y) = 0$, x e y si dicono **ortogonali rispetto a B**, o **B-ortogonali** (o anche soltanto **ortogonali** sottintendendo il riferimento a B) tra loro. È chiaro che se $x = 0$ e/o $y = 0$, x e y sono automaticamente ortogonali; come pure che, se x e y sono ortogonali, allora lo sono anche λx e y, $\forall \lambda$ reale.

Nel seguito, il riferimento a E (spazio R-lineare finito-dim) inteso come “universo del discorso”, sarà spesso sottinteso. Se B è una forma bilineare definita positiva [negativa] su E, la coppia $\{E,B\}$ si dice uno **spazio euclideo** [**antieucclideo**]. Se la forma bilineare B soddisfa l’assioma (c4bis), la coppia $\{E,B\}$ si dice uno **spazio pseudoeuclideo**, e **propriamente pseudoeuclideo** se non è né euclideo né antieucclideo; se infine soddisfa l’assioma (c4ter), la coppia $\{E,B\}$ si dice uno **spazio semieucclideo**, e **propriamente semieucclideo** se $\delta > 0$. In quest’ultimo caso, il SS N di E (vedi (c4ter)) è il nucleo di B, e la sua dimensione $0 < \delta \leq n$ ($\equiv \dim E$) si dice (**indice di deficienza di** $\{E,B\}$). La differenza $\rho =: n - \delta$, che è ≥ 0 e $< n$, si dice **rango di B**, e si denota $\text{rng}(B)$. Se con la solita deroga si ammette $\delta = 0$, $\{E,B\}$ è pseudoeuclideo, $\rho = n$, e B si dice **non singolare**. Se fosse $\rho = 0$, allora $E = N$ e $B(x,y) = 0 \forall (x,y \in E)$. Qui e nel seguito escluderemo sempre questo caso degenero di B nulla su tutto E.

Queste definizioni si possono anche svincolare dagli assiomi (c4bis) o (c4ter) avendo introdotto una conveniente “matrice di B”. Sia $\{e_1, \dots, e_n\}$ una base ⁸ di E, e si ponga $B_{ij} =: B(e_i, e_j)$, $1 \leq i, j \leq n$; a B corrisponde così, in quella base, un’unica $(n \times n)$ -matrice simmetrica, la **matrice di B nella base** $\{e_1, \dots, e_n\}$ (e viceversa ad una $(n \times n)$ -matrice simmetrica corrisponde un’unica B che la ammette come sua matrice nella stessa base). Come vedremo (v. T8 appresso), tale matrice ha lo stesso rango (nel senso dell’algebra matriciale ⁹) $\text{rng}(B)$ *comunque si scelga la base di riferimento*.

Ricordiamo il seguente teorema sul rango di una $(n \times n)$ -matrice:

T0. «se C e A sono due matrici quadrate dello stesso ordine, e C è non singolare, le matrici-prodotto CA e AC hanno rango uguale a quello di A.»

⁸ Ricordiamo (cfr. S.sez. 1.3.1) che una base di E (n-dim) è un insieme di suoi elementi linearmente indipendenti tale che ogni elemento di E si possa rappresentare come combinazione lineare degli elementi della base (tale rappresentazione essendo unica). Ovviamente il numero degli elementi della base coincide con n.

⁹ Vi sono più definizioni equivalenti del rango $r \leq n$ di una matrice quadrata di ordine n, delle quali la più comunemente nota è quella di “massimo ordine dei suoi minori non nulli”. Se $r = n$, la matrice in oggetto è non singolare; se $r < n$, esiste un minore non nullo di ordine r, ma non uno di ordine $r+1$. Il caso $r = 0$ si può considerare privo di interesse, perché corrisponde ad una matrice nulla. La stessa definizione di rango si applica anche a $(n \times m)$ -matrici (non quadrate), nel qual caso deve essere $r \leq \min(n,m)$. Se una matrice ha rango $r > 0$ non tutti i suoi minori di ordine $< r$ sono nulli (teorema di Laplace). In italiano il rango di una matrice si nomina anche come sua “caratteristica”.

A, A_1, A_2, \dots saranno generici SI, e H, H_1, H_2, \dots generici SS ($\neq \{0\}$ salvo avviso contrario) di E , che si supporrà (salvo avviso contrario) generalmente semieuclideo e n -dim. $A^{\perp, B}$ (o brevemente A^{\perp} sottintendendo il riferimento a B) si definisce come il SI costituito da tutti (e soli) gli elementi (di E) che sono (B -)ortogonali ad ogni elemento di A , cioè $A^{\perp} =: \{x | \forall y (\in A) \{B(x, y) = 0\}\}$, e si dice **ortogonale** (sost.) **di** A . Si constata subito che A^{\perp} è un SS, ovvero che

$$(6) \quad A^{\perp} = \text{vect}A^{\perp} = (\text{vect}A)^{\perp}.$$

In base alla definizione si prova anche che

$$(7) \quad A^{\perp\perp} (\equiv (A^{\perp})^{\perp}) \supset A,$$

$$(8) \quad (A_1 \cup A_2)^{\perp} = A_1^{\perp} \cap A_2^{\perp},$$

$$(9) \quad \{0\}^{\perp} = E, \text{ }^{10} \text{ e}$$

$$(10) \quad \dim H + \dim H^{\perp} \geq \dim E = n$$

Per dati SS H_1, H_2, H_1+H_2 si definisce come il SI (di E) degli elementi esprimibili come somma di un elemento di H_1 e di un elemento di H_2 , cioè

$$(11) \quad H_1+H_2 =: \{z | \exists \langle x, y \rangle (\in H_1 \times H_2) \{z = x + y\}\}.$$

H_1+H_2 è banalmente simmetrica rispetto a H_1 e H_2 , per cui si dirà **somma di H_1 e di H_2** , Inoltre l'operazione $+$ è associativa. In forza di questa seconda proprietà si passa alla somma di più (di due) SS (ma quest'ultima si può anche direttamente definire attraverso una ovvia generalizzazione della (11)). Si verifica che H_1+H_2 è un SS, e che H_1 e H_2 sono SS di H_1+H_2 . Si prova anche facilmente che $A_1^{\perp} + A_2^{\perp} \subset (A_1 \cap A_2)^{\perp}$.

In modo analogo, $H_1 \oplus H_2$ si definisce come il SI degli elementi esprimibili come somma di un *unico* elemento di H_1 e di un *unico* elemento di H_2 , cioè

$$(12) \quad H_1 \oplus H_2 =: \{z | \exists! \langle x, y \rangle (\in H_1 \times H_2) \{z = x + y\}\},$$

e si dice **somma diretta di H_1 e H_2** . La somma diretta di SS è un particolare tipo di somma ($+$) degli stessi SS (la (12) differisce dalla (11) soltanto per la sostituzione di \exists con la più severa $\exists!$); quindi gode di tutte le proprietà di quest'ultima, e possibilmente di proprietà più specifiche. Ad esempio, la somma \oplus è anch'essa simmetrica ed associativa. Inoltre, ovviamente,

$$(13) \quad H_1 \oplus H_2 \subset H_1 + H_2.$$

Seguono i teoremi:

$$T1. \quad \langle H_1 \oplus H_2 = H_1 + H_2 \text{ sse } H_1 \cap H_2 = \{0\} \rangle;$$

cioè «la decomposizione di $z \in H_1+H_2$ in $x(\in H_1) + y(\in H_2)$ è unica sse $H_1 \cap H_2 = \{0\}$ ».

Dim. della parte «se»: per assurdo, supponiamo esistere due decomposizioni $x+y$ e $x'+y'$ di z , con $x, x' \in H_1$ e $y, y' \in H_2$, quindi $x + y = x' + y'$. Ma $x - x' = y - y'$ appartiene a $H_1 \cap H_2$, e in forza

¹⁰ La duale di questa, $E^{\perp} = \{0\}$, vale sse E è pseudoeuclideo, perché $E^{\perp} = N$ e $\delta = 0$ sse E è pseudoeuclideo.

dell'ipotesi $H_1 \cap H_2 = \{0\}$ deve essere uguale a 0, cioè $x = x', y = y'$, qed. Dim. della parte “solo se”: ancora per assurdo, supponiamo $H_1 \cap H_2 \neq \{0\}$, cioè che a $H_1 \cap H_2$ appartenga un $z \neq 0$. Ma $z = z(\in H_1) + 0(\in H_2) = 0(\in H_1) + z(\in H_2)$; quindi la decomposizione non è unica, contro l'ipotesi che lo sia, qed. #

T2. « $\dim(H_1 \oplus H_2) = \dim H_1 + \dim H_2$ »

Dim: Sia $\{e_1, \dots, e_p\}$ una base di H_1 e $\{f_1, \dots, f_q\}$ una base di H_2 . Se $z \in H_1 \oplus H_2$, esiste un'unica coppia ordinata $\langle x, y \rangle \in (H_1 \times H_2)$ per cui $z = x + y$, quindi $(\circ) z = \sum_{i=1}^p x^i e_i + \sum_{j=1}^q y^j f_j$, ove i coefficienti x^i e y^j sono unicamente determinati. Segue da una parte che i vettori $\{e_1, \dots, e_p, f_1, \dots, f_q\}$ sono linearmente indipendenti, e dall'altra che $H_1 \oplus H_2 = \text{vect}\{e_1, \dots, e_p, f_1, \dots, f_q\}$. In definitiva $\{e_1, \dots, e_p, f_1, \dots, f_q\}$ è una base di $H_1 \oplus H_2$, da cui la tesi. #

T3. « $H \cap H^\perp = \{0\}$ » \Rightarrow « $\dim H + \dim H^\perp = n$ »

Dim: si ha $n \leq \dim H + \dim H^\perp$ (secondo la (10)) = $\dim(H \oplus H^\perp)$ (secondo (T2)) $\leq n$ (perché $H \oplus H^\perp$ è un SS), e dunque i due \leq precedenti devono essere degli $=$. Da qui la tesi. # Inoltre, dalla $\dim(H \oplus H^\perp) = n$ scende $H \oplus H^\perp = E$ (v. punto (ii) tra le proprietà dei generici SS).

T4. In termini di somma diretta, è facile provare che

$$(14) \quad \langle A_1^\perp \oplus A_2^\perp \subset (A_1 \cap A_2)^\perp$$

per generici SI di E .» Dim: esercizio. #

T5. «Se il rango di B è massimale (cioè $\rho = n$), vale la stessa tesi $\dim H + \dim H^\perp = n$ di T3.»

Dim: La dimostrazione si fonda sull'isomorfismo che esiste allora tra E e il suo duale E^* (vedi S.sez. 3.1.4). L'immagine di H^\perp attraverso questo isomorfismo è infatti l'insieme delle forme lineari $B(y, \cdot)$ per cui $y \in H^\perp$, cioè che si annullano in H ; o come anche si dice, è il SS di E^* “B-ortogonale* di H ”. Se scriviamo H^* per questo B-ortogonale*, risulta $\dim H^* = \dim H^\perp$. La tesi scende quindi dal fatto, facilmente provabile, che $\dim H + \dim H^* = n$. #

T6. «Ancora per B di rango massimale, vale l'uguaglianza

$$(15) \quad H^{\perp\perp} = H.$$

Dim: Se B è non singolare, la tesi di T3 vale anche con H^\perp in luogo di H :

$$(16) \quad \dim H^\perp + \dim H^{\perp\perp} = n,$$

e dunque, per differenza con $\dim H + \dim H^\perp = n$ (v. T3)

$$(17) \quad \dim H^{\perp\perp} = \dim H.$$

In generale, da “ $H_1 \supset H_2$ et $\dim H_1 = \dim H_2$ ” segue $H_1 = H_2$. (infatti, se si suppone stretta l'inclusione $H_1 \supset H_2$, una base di H_1 si forma aggiungendo qualche elemento a una base di H_2 , il

che implica $\dim H_1 > \dim H_2$, contro la seconda ipotesi.) Ora alla (17) va associata la (7), cioè proprio la $H^{\perp\perp} \supset H$. Si conclude che deve valere la (15), qed. #

Sia $\{B_{ij}\}_{i,j=1,\dots,n}$ la matrice di B (come anzidetto, escluderemo sempre il caso di B nulla, ovvero $\rho = 0$) nella base $\{e_1, \dots, e_n\}$, e siano $x = \sum_i x_i e_i$ e $y = \sum_j y_j e_j$ le rappresentazioni (uniche) di x e y in quella base ($x_i = B(x, e_i)$, $y_j = B(y, e_j)$). Allora $B(x, y) = \sum_{i,j} B_{ij} x_i y_j$; e se Q è l'associata di B , $Q(x) = \sum_{i,j} B_{ij} x_i x_j$ ($\equiv \sum_i B_{ii} x_i^2 + 2 \sum_{i < j} B_{ij} x_i x_j$). Se $Q(x)$ viene scritta come $\sum_{i,j} Q_{ij} x_i x_j$ (ove Q_{ij} si può supporre simmetrica senza limitazioni di generalità), Q_{ij} è similmente detta la **matrice di Q** (nella stessa base), ed è evidente che le due matrici, quella di B e quella di Q nella stessa base, sono identiche secondo la definizione data. Per data B su E , una base $\{e_1, \dots, e_n\}$ per cui risulta $B_{ij} = 0$ per ogni coppia (i, j) con $i \neq j$ si dice **B-ortogonale**, o semplicemente **ortogonale** sottintendendo il riferimento a B ; se più in particolare è $B_{ij} = \delta_{ij}$ (simbolo di Kronecker), e quindi $Q(x) = \sum_{i=1}^n x_i^2$, la base si dice **B-ortonormale** (o **ortonormale**). (Per $n = 1$, qualunque base è ortogonale perché l'antecedente $i \neq j$ è falso.)¹¹ Seguono (gli enunciati di) altri teoremi validi per spazi generalmente semieulidei $\{E, B\}$, dei quali preferiamo posticipare la dimostrazione.

T7. (della base ortogonale): «In un generico spazio semieulideo $\{E, B\}$ (finito-dim) esiste una base B-ortogonale.» Dim: vedi S. Sez. 9.2.1 #

T8. (del rango): «In un generico spazio semieulideo $\{E, B\}$, il numero degli elementi diagonali diversi da zero della matrice di B in una base B-ortogonale (ne esiste *almeno* una per T7) non dipende dalla base prescelta.» Dim: vedi S.sez. 9.2.1 # Questo numero, dunque intrinseco a B , è il rango ρ (per ipotesi ≥ 1) di B precedentemente introdotto, che riceve così una definizione alternativa e (si dimostra) equivalente.

T9. (teorema di Sylvester, o legge di inerzia): «Ancora riferendoci allo spazio semieulideo $\{E, B\}$, il numero degli addendi positivi (quindi anche di quelli negativi) dello sviluppo $\sum_{i=1}^{\rho} Q_{ii} x_i^2$ di $Q(x)$ in una base B-ortogonale non dipende dalla base prescelta.» Dim: vedi S.sez. 9.2.1 #

Sempre riferendoci allo spazio semieulideo $\{E, B\}$, sia ancora $\{e_1, \dots, e_n\}$ una base ortogonale di E , e sia π il numero (base-indipendente) degli elementi diagonali positivi della matrice di Q . Ordiniamo la base in modo che i suoi primi $\pi \geq 0$ elementi siano quelli per cui $Q(e_i) > 0$. Allora $Q(x) = \sum_{i=1}^{\pi} Q(e_i) x_i^2 + \sum_{j=\pi+1}^{\rho} Q(e_j) x_j^2$, dove $Q(e_i) > 0$ per $i = 1, \dots, \pi$ e $Q(e_j) < 0$ per $\pi+1 \leq j \leq \rho$. (Si intende che se $\pi = 0$ [$\pi = \rho$] manca la prima [la seconda] somma.) Passando alla

¹¹ Ricordiamo che, per data B definita positiva, una base ortonormale si ricava mediante la classica procedura (o algoritmo) di Gram-Schmidt (vedi S.sez. 3.2.4) partendo da una base qualsiasi. Se si ordina la base di partenza, la base ortonormale prodotta dalla procedura è unica a meno del segno (comune) dei suoi elementi unitari; e con la scelta usuale di questo segno risulta equiversa alla base originale.

base, ortogonale come la $\{e_1, \dots, e_n\}$, $\{e'_1, \dots, e'_n\} =: \{e_1/|Q(e_1)|^{1/2}, \dots, e_\rho/|Q(e_\rho)|^{1/2}, e_{\rho+1}, \dots, e_n\}$, quindi con $Q(e'_i) = 1$ per $i = 1, \dots, \pi$, $Q(e'_j) = -1$ per $j = \pi+1, \dots, \rho$, otteniamo:

$$(18) \quad Q(x) = \sum_{i=1}^{\pi} x'_i{}^2 - \sum_{j=\pi+1}^{\rho} x'_j{}^2 \equiv \sum_{h=1}^{\rho} \varepsilon(h) x'_h{}^2$$

dove $\varepsilon(1 \leq h \leq \pi) =: 1$ (se $\pi \geq 1$), e $\varepsilon(\pi+1 \leq h \leq \rho) =: -1$ (se $\rho \geq \pi+1$), e le x' sono le componenti di x nella base $\{e'\}$. Questa rappresentazione di $Q(x)$, in cui gli interi $0 \leq \pi \leq \rho$ e $1 \leq \rho \leq n$ sono *intrinseci* alla forma B , si dice **canonica**. Una base di E in cui per ogni x $Q(x)$ assume la forma canonica (18) si dice **B-semiortonormale**; e se in particolare B è non singolare ($\rho = n$), **B-pseudortonormale** (o **B-pseudopitagorica**). (Correntemente, i “B” che precedono questi attributi possono essere sottintesi.) Se più in particolare $\pi = \rho = n$, torniamo alla base B -ortonormale (o B -pitagorica) di uno spazio euclideo o antieucideo. Il numero π , intrinseco a B e compreso tra 0 e ρ si dice **indice positivo** (o brevemente **indice**) di B , e la coppia ordinata di numeri (anch'essi compresi tra 0 e ρ) $\langle \pi, v =: \rho - \pi \rangle$ **segnatura di B** (all'occorrenza, v potrà nominarsi come **indice negativo di B**). Essendo base-indipendenti, il rango e l'indice delle forme bilineari simmetriche [delle forme quadratiche associate] sullo spazio finito-dim R -lineare $E \times E$ [E] possono essere usati per classificare tali forme.

Due forme bilineari (simmetriche, non nulle) B_1 e B_2 su E (finito-dim), si dicono **equivalenti** se esiste un automorfismo φ di E per il quale $B_2(\varphi x, \varphi y) = B_1(x, y)$ per ogni x, y di E . La definizione può anche darsi dicendo che esistono due basi di E , diciamo $\{e_1, \dots, e_n\}$ e $\{f_1, \dots, f_n\}$, per le quali $B_1(e_i, e_j) = B_2(f_i, f_j) \quad \forall (i, j)$ (la verifica è lasciata al lettore). Chiaramente la relazione “ B_1 è equivalente a B_2 ” è una relazione di equivalenza. La condizione di equivalenza tra B_1 e B_2 può anche esprimersi, in modo ovvio, in termini delle forme quadratiche associate Q_1 e Q_2 .

Posto in termini generali, il problema di individuare condizioni necessarie e sufficienti alla equivalenza di due date forme B_1 e B_2 può non essere semplice. Tuttavia ci si può limitare a considerare basi $\{e_1, \dots, e_n\}$ e $\{f_1, \dots, f_n\}$ che siano B_1 -ortogonale la prima e B_2 -ortogonale la seconda. Vale allora il teorema, diretta conseguenza della legge d'inerzia e del suo corollario (18):
T10. «Affinché due forme B_1 e B_2 siano equivalenti, occorre e basta che abbiano la stessa segnatura». ¹² Dim: esercizio. #

Se Q è non singolare ($\rho = n$), essa è **definita positiva** se $\pi = n$, **definita negativa** (cioè, opposta ad una forma definita positiva) se $v = n$, e **indefinita** in tutti gli altri casi. Come sappiamo, nel primo caso $\{E, B\}$ è uno spazio euclideo, nel secondo uno spazio antieucideo e nel terzo uno spazio (propriamente) pseudoeucideo. Uno spazio n -dim pseudoeucideo di indice π si denota

¹² Se tuttavia E è C -lineare invece che R -lineare, allora (T10) si modifica nel più forte (T10bis) «Affinché due forme B_1 e B_2 [Q_1 e Q_2] su E siano equivalenti, occorre e basta che abbiano lo stesso rango.»

genericamente $E_{n,\pi}$; uno spazio euclideo, che è un $E_{n,n}$, semplicemente come E_n , e uno spazio antieuclideo, che è un $E_{n,0}$, semplicemente come E_{-n} . Se invece Q è singolare ($\rho < n$) con indice di deficienza $\delta = n - \rho$, lo spazio semieuclideo $\{E,B\}$ si denota come $E_{n,\pi,\delta}$ (quindi $E_{n,\pi,0}$ è uno spazio pseudoeuclideo $E_{n,\pi}$).

Va da sé che una $Q(x)$ non singolare è $> 0 \forall x \neq 0$ sse è definita positiva e < 0 sse è definita negativa. Se la forma è singolare ($\rho < n$), essa si dice **semidefinita positiva** se $\pi = \rho$ (o $v = 0$) e **semidefinita negativa** se $v = \rho$ (o $\pi = 0$). Una tale forma singolare è $\geq 0 \forall x \neq 0$ sse è semidefinita positiva, e ≤ 0 sse è semidefinita negativa.

Lo **pseudomodulo** di un vettore x di uno spazio pseudoeuclideo $E_{n,\pi}$ si definisce come la radice quadrata del *valore assoluto* di $Q(x)$, $|Q(x)|^{1/2}$; se Q è definita positiva ($\pi = n$) lo pseudomodulo si riduce a $(Q(x))^{1/2}$, cioè all'usuale modulo pitagorico $[\sum_{i=1}^n x_i^2]^{1/2}$ in una base ortonormale.

Ancora in uno spazio pseudoeuclideo $E_{n,\pi}$, i vettori x per cui $|Q(x)| = 1$ si dicono **pseudounitari** (positivi aut negativi) e quelli per cui $Q(x) = 0$ si dicono **isotropi**. Il vettore $x = 0$ è banalmente isotropo. A parte il vettore nullo, in uno spazio pseudoeuclideo $E_{n,\pi}$ vi sono tre tipi di vettori: i **vettori pseudoreali** (o **euclidei**) per cui $Q(x) > 0$, i **vettori pseudoimmaginari** (o **pseudoeuclidei**) per cui $Q(x) < 0$, e i nominati vettori isotropi $\neq 0$, per cui $Q(x) = 0$. Un vettore non isotropo può essere sempre reso pseudounitario moltiplicandolo per un numero positivo (il reciproco della $\sqrt{|Q(x)|}$ della sua Q se pseudoreale o della sua $-Q$ se pseudoimmaginario). L'insieme delle rette isotrope di $E_{n,\pi}$ passanti per l'origine si dice **cono isotropo**. Un vettore isotropo è ortogonale a se stesso, o ad un qualunque vettore (isotropo) ad esso proporzionale secondo un fattore reale arbitrario.

Nel caso euclideo ($\pi = n$, con Q definita positiva) [antieuclideo, $v = n$, con Q definita negativa], $(Q(x))^{1/2}$ [$(-Q(x))^{1/2}$] può essere adottata come **norma** (norma "standard"), diciamo $N(x)$, di E , perché soddisfa a tutti i relativi assiomi.¹³ Poiché una norma $N(\cdot)$ induce sempre una distanza $d(\cdot, \cdot)$ (basta porre $d(x,y) = N(x-y) = N(y-x)$), cioè una metrica, e quindi una topologia, un E euclideo o antieuclideo può sempre pensarsi come metrico e topologico con la norma standard. Ciò non è più possibile se E è *propriamente* pseudoeuclideo ($1 \leq \pi \leq n-1$): infatti $|Q(x)|^{1/2}$ non può più considerarsi una norma perché il suo annullarsi non implica (in generale) $x = 0$.

¹³ Come già visto (Cap 1), questi assiomi sono: 1. $N(x) \geq 0$ e $Nx = 0$ solo se $x = 0$; 2. $N(x+y) \leq N(x) + N(y)$ (disuguaglianza triangolare); 3. $N(\alpha x) = |\alpha|N(x)$ (per ogni x,y di E e ogni α di \mathbb{R}). Identificando $N(x)$ con la radice quadrata di $|Q(x)|$ di uno spazio euclideo o antieuclideo, 1. e 3. sono evidentemente soddisfatti, mentre 2. si dimostra equivalente alla disuguaglianza di Cauchy-Schwarz.

In uno spazio propriamente pseudoeuclideo vi sono diversi tipi di piani: i **piani euclidei** E_2 ($\equiv E_{2,2}$) abbracciati da due vettori euclidei, i **piani pseudoeuclidei** $E_{2,1}$ e $E_{2,0}$ abbracciati da due vettori uno dei quali (caso di $E_{2,1}$) non euclideo, o entrambi (caso di $E_{2,0}$) non euclidei; i piani contenenti vettori isotropi, che sono **piani semieuclidei** con deficienza 1, del tipo $E_{2,1,1}$ e $E_{2,0,1}$, e i **piani isotropi**, semieuclidei con deficienza 2 (di tipo $E_{2,0,2}$), tutti i vettori dei quali sono isotropi. Sottolineiamo che uno spazio propriamente pseudoeuclideo *non* può essere metrico (a meno che non gli si assegni una metrica prescindendo da B). In particolare, in esso *non* vale la disuguaglianza triangolare $\|x + y\| \leq \|x\| + \|y\|$ (avendo indicato con $\|\cdot\|$ lo pseudomodulo), almeno per coppie di vettori x, y generici. Se ad esempio due vettori determinano un piano pseudoeuclideo del tipo $E_{2,0}$, allora (si dimostra) vale per essi la **disuguaglianza triangolare inversa**, cioè con \leq sostituito da \geq . Infine in uno spazio propriamente pseudoeuclideo ci sono tre tipi di “sfere” centrate nell’origine, cioè di superfici $(n-1)$ -dim per cui $Q(x) = \text{cost}$: quelle con $Q(x) = \text{cost} = R^2 > 0$, o **sfere tout court**, quelle con $Q(x) = \text{cost.} = -R^2 < 0$ o **pseudosfere**, e quella con $Q(x) = 0$ o **sfera isotropa**. La costante R che figura nella equazione di una sfera o di una pseudosfera si dice (il suo) **raggio**. In una base B-pseudortonormale, e con il conveniente ordinamento degli indici, l’equazione della sfera isotropa di $E_{n,\pi}$ si scrive $\sum_{i=1}^{\pi} x_i^2 = \sum_{j=\pi+1}^n x_j^2$; e in particolare per $\pi = n-1$ [$\pi = 1$], $\sum_{i=1}^{n-1} x_i^2 = x_n^2$ [$x_1^2 = \sum_{i=2}^n x_i^2$]. Pensando le x come coordinate cartesiane ortogonali standard, le due ultime equazioni sono quelle di un (doppio) cono circolare “retto” con il vertice nell’origine e con (x_n) [(x_1)] come asse di simmetria rotazionale. In uno spazio euclideo [antieuclideo] non esistono pseudosfere [sfere] centrate nell’origine, e la sfera isotropa degenera nell’origine stessa.

Le trasformazioni tra componenti di un vettore x di uno spazio pseudoeuclideo dovute al passaggio da una base pseudortonormale (e) a una base pseudortonormale (e'), diciamo $x'_i = a_{ih}x_h$, che lasciano invariata la forma $Q(x)$ – per cui cioè $\sum x_i^2 \varepsilon(i) = \sum x'_i{}^2 \varepsilon(i)$ (con la già data definizione di ε) – si dicono **trasformazioni B-pseudortonormali**. I loro coefficienti (a) soddisfano alle

$$(19) \quad a_{ih}a_{ik}\varepsilon(i) = \delta_{hk}\varepsilon(h).$$

Nel caso euclideo, $\varepsilon(i) = 1$ per ogni i , e siamo ricondotti alle trasformazioni ortonormali (1.5, 2). Se invece ci si riferisce a trasformazioni della base, diciamo $e'_i = b_{ih}e_h$, la stessa richiesta di invarianza della forma $Q(x)$ porta alle

$$(20) \quad b_{hi}b_{ki}\varepsilon(i) = \delta_{hk}\varepsilon(h).$$

Il confronto tra le (19) e le (20) mostra che, in una trasformazione pseudortonormale, i coefficienti (a) sono sommati rispetto ai loro *primi* indici (i) (con i fattori $\varepsilon(i)$), e i coefficienti (b) sono sommati rispetto ai loro *secondi* indici (i) (con gli stessi fattori $\varepsilon(i)$). È anche immediato verificare che il quadrato del determinante delle (a) soggette alle (19) [delle (b) soggette alle (20)] è uguale a 1.

Infatti $\det\{a_{ih}\varepsilon(i)\} = \Pi_k \varepsilon(k) \det\{a_{ih}\}$; quindi dalle (19) segue $\det^2\{a_{ih}\} \Pi_k \varepsilon(k) = \Pi_k \varepsilon(k)$, o (semplificando la scrittura del determinante) $(^\circ) \det^2\{a\} = 1$. Similmente si prova che dalle (20) segue $\det^2\{b\} = 1$. D'altra parte la $e_i x_i = e'_j x'_j$ implica $\det\{a\} \det\{b\} = 1$, ovvero $(*)$ “ $\det\{a\} = 1$ ” $\Leftrightarrow \Leftrightarrow$ “ $\det\{b\} = 1$ ”. Dalla $(^\circ)$, si ha $\det\{a\} = 1$ oppure $\det\{a\} = -1$; quindi, per la $(*)$, $\det\{b\} = 1$ nel primo caso e $\det\{b\} = -1$ nel secondo. In conclusione, si possono distinguere trasformazioni pseudortonormali **equiverse** (con $\det(a) = \det(b) = 1$) e trasformazioni pseudortonormali **antiverse** (con $\det(a) = \det(b) = -1$). Le trasformazioni di Lorentz sono pseudortonormali equiverse.

Siano E_{n,π_1} con base $\{e_i\}$ e F_{n,π_2} con base $\{f_j\}$ due spazi pseudoeuclidei $n \geq 2$ -dim per i quali $\pi_1 + \pi_2 = n$, $\{e_i\}_{i=1+\pi_1} = \{f_j\}_{j=\pi_2+1+n}$ e $\{e_i\}_{i=\pi_1+1+n} = \{f_j\}_{j=1+\pi_2}$. Questi spazi hanno gli stessi elementi, e se la base dell'uno è pseudortonormale anche quella dell'altro lo è. In queste basi, è evidente che la forma quadratica pseudopitagorica (18) su un qualunque comune elemento x cambia segno passando da uno spazio all'altro. In questo senso, gli spazi E_{n,π_1} e F_{n,π_2} possono considerarsi “antiequivalenti”. In particolare, le forme pseudopitagoriche in x nei due spazi antiequivalenti $E_{n,n-1}$ e $F_{n,1}$, che si dicono **lorentziani in senso lato**, sono $x_1^2 + \dots + x_{n-1}^2 - x_n^2$ e rispettivamente $-x_1^2 - \dots - x_{n-1}^2 + x_n^2$, appunto uguali e opposte. Le geometrie di spazi antiequivalenti, e in particolare della coppia lorentziana $(E_{n,n-1}, F_{n,1})$, sono sostanzialmente le stesse, e la scelta dell'uno o dell'altro spazio come oggetto di studio è questione di gusti. Nel caso della relatività speciale ($n = 4$) si opta indifferentemente, ma una volta per tutte, per $E_{4,3}$ o per $F_{4,1}$, con le forme quadratiche pseudopitagoriche in x uguali a $x_1^2 + x_2^2 + x_3^2 - x_4^2$ e rispettivamente $-x_1^2 - x_2^2 - x_3^2 + x_4^2$ (la nostra presente scelta è andata alla prima forma).

Gli spazi pseudoeuclidei n -dim, includendovi anche quelli “impropriamente” tali (gli euclidei e gli antieuclidei), sono evidentemente di $n+1$ tipi, che sono tuttavia a due a due (salvo uno se n è pari) antiequivalenti nel senso appena detto. Contando coppie di spazi antiequivalenti come un solo spazio, si vede subito che gli spazi pseudoeuclidei n -dim “essenzialmente” distinti sono $1 + n/2$ se n è pari, e $(n+1)/2 = 1 + (n-1)/2$ se n è dispari.

2.3.2) LO SPAZIO DI MINKOWSKI (NOZIONI DI BASE)

Come abbiamo appena detto, uno spazio pseudoeuclideo n -dim del tipo $E_{n,n-1}$ (o del tipo equivalente $E_{n,1}$) si dice (spazio) di Minkowski; e se $n = 4$, (spazio) di Minkowski **in senso stretto**. Esamineremo ora alcune definizioni e proprietà fondamentali degli spazi di Minkowski.

Un SI K di uno spazio lineare n -dim si dice un suo **cuneo** se è “stabile” rispetto alla somma (cioè, se $x \in K$, $y \in K$ implica $x+y \in K$) e alla moltiplicazione per un reale *non negativo* α ($x \in K$ implica $\alpha x \in K$). Un cuneo di dimensione massimale ($\dim K = n$) si dice **cuneo solido** (agg.). Per $n > 1$, sia $\{e_1, \dots, e_m\}$, $2 \leq m \leq n$, una famiglia di m vettori linearmente indipendenti (o m -famiglia **libera**) di E . Il SI di E

$$(1) \quad A =: \{x \mid x = \sum_{i=1}^m x_i e_i, \sum_{i=1}^{m-1} x_i^2 - x_m^2 \leq 0, x_m \geq 0\}$$

e il SI $\mathring{A} =: A' \cup \{O\}$, dove A' è definito come A ma con $<$ al posto della \leq nella prima disuguaglianza, e O è al solito lo zero ($\in A$) di E , sono coni. Si noti che $\mathring{A} \subset A$. Se $m = n$, questi coni (solidi) si dicono **minkowskiani**: **chiuso** (A) e **aperto** (\mathring{A}). È anche utile introdurre il **doppio cono minkowskiano chiuso** $A \cup -A$, che denoteremo $\pm A$, e l'analogo **doppio cono aperto** $\mathring{A} \cup -\mathring{A}$, che similmente denoteremo $\pm \mathring{A}$.¹⁴

Facendo $n = m = 4$, ed identificando la base di E $\{e_1, \dots, e_4\}$ con quella formata dai 3 versori spaziali ortonormali standard e dal versore della lunghezza römeriana ct , la seconda disuguaglianza nella (1), $x_4 \geq 0$, significa $ct \geq 0$; quindi A e \mathring{A} sono “nel futuro” (coni minkowskiani (solidi) **forward**, $ct \geq 0$) e i loro opposti $-A$ e $-\mathring{A}$ sono “nel passato” (coni minkowskiani (solidi) **backward**, $ct \leq 0$) rispetto al loro vertice O . Il valore di $\sum_{i=1}^3 x_i^2 - x_4^2$ è invariante per trasformazioni di Lorentz sulle $x_{1 \leq i \leq 4}$, e quindi lo è anche l'equazione della comune frontiera dei due coni forward o backward.¹⁵ Fisicamente, l'evento “punto materiale P nel vertice di A (o di \mathring{A})” implica l'evento “le coordinate di P sono nel cono forward \mathring{A} se $t > 0$, e nel cono backward $-\mathring{A}$ se $t < 0$ ”.

Sappiamo che il quadrato formale della pseudodistanza tra il punto-istante (o quadrivettore) $x \equiv (x_1, \dots, x_4)$ e l'origine, cioè (in forma pseudopitagorica) $\sum_{i=1}^3 x_i^2 - x_4^2$ è invariante rispetto a trasformazioni di Lorentz, e indefinito. Se $\sum_{i=1}^3 x_i^2 - x_4^2$ è negativo, il (quadri)vettore x si dice di tipo **temporale** (o **time-like**); se lo stesso numero è positivo, il vettore x si dice di tipo **spaziale** (o **space-like**); se infine è nullo, il vettore x (supposto distinto da $\langle 0,0,0,0 \rangle \equiv$ vettore “zero”) si dice di tipo **luminale** (o **light-like**).¹⁶ I vettori time-like sono in uno dei due coni aperti (forward o backward); i vettori space-like sono fuori del doppio cono chiuso, ovvero in $E \setminus \pm A$; e infine i vettori

¹⁴ \mathring{A} non è “aperto” in senso topologico perché il suo vertice non è un punto interno. A , $-A$, \mathring{A} , $-\mathring{A}$ sono stabili rispetto alla somma (lo proveremo più oltre), ma non rispetto alla moltiplicazione per un reale generico; quindi essi non sono spazi lineari. Similmente gli insiemi $(E \setminus \pm A) \cup \{O\}$ e $(E \setminus \pm \mathring{A}) \cup \{O\}$ sono stabili rispetto alla moltiplicazione per un reale generico ma non rispetto alla somma; nemmeno essi, quindi, sono spazi lineari.

¹⁵ Spesso è proprio questa frontiera che viene detta cono, ciò che noi eviteremo di fare qui.

¹⁶ A chi scrive, gli aggettivi “temporale” per “time-like” e “spaziale” per “space-like” non sono mai sembrati del tutto soddisfacenti. Invece “luminale” per “light-like” ci appare adeguato, ma non è di uso comune. In accordo con quanto osservato nella precedente sottosezione, i matematici dicono “pseudoreale” per space-like, “pseudoimmaginario” per time-like e “isotropo” per luminale.

light-like sono sulla frontiera di A aut di $-A$ (o equivalentemente di \hat{A} aut di $-\hat{A}$). Quella dei vettori light-like è cioè una possibilità limite. Fisicamente, essa rappresenta lo spostamento di un fotone.

Una curva di equazione parametrica $x = x(\lambda)$, λ essendo un generico parametro reale variabile in un certo intervallo base, di classe C^1 ¹⁷, si dice **time-like** se il suo vettore tangente $dx/d\lambda$ è ovunque time-like, e similmente per le curve **space-like** e le curve (rette) **light-like**. Se tale curva descrive lo sviluppo di un processo o fenomeno fisico nel tempo, ad esempio la traiettoria di un punto materiale, ovviamente quella curva deve essere time-like, e in tal caso si dice **curva di universo** del fenomeno considerato. Se un vettore x è time-like, allora esiste un riferimento, raggiungibile mediante una trasformazione di Lorentz, in cui quel vettore ha componenti spaziali nulle (cioè si trova sull'asse (x_4) ; $|x_4| > 0$ è allora c volte il tempo proprio di x rispetto all'origine, e non esiste un riferimento in cui x abbia componente temporale nulla. Similmente se x è space-like, c'è un sistema di riferimento, sempre raggiungibile mediante una trasformazione di Lorentz, ove x è simultaneo (\equiv è sul $(n-1)$ -piano $x_4 = 0$). Un evento a componente temporale x_4 nulla non può essere connesso all'evento O (origine) da una relazione causale.

Lo spazio di Minkowski (in senso stretto) M è uno spazio R -lineare 4-dim in tutto simile a E_4 salvo il fatto che il prodotto scalare (o interno) di due suoi elementi è sostituito da un corrispondente **pseudoprodotto interno**, definito come la somma dei prodotti delle tre prime componenti omologhe (di indici 1, 2, 3, spaziali) *meno* il prodotto delle quarte componenti (di indice 4, temporali) rispetto ad una base pseudortonormale¹⁸. Si vede immediatamente che questo pseudoprodotto interno, che denoteremo qui con $(*)$, gode di tutte le proprietà formali del prodotto interno euclideo, salvo la natura indefinita del quadrato formale ad esso associato, contro quella positiva definita del secondo. Come sappiamo, esso è invariante, in valore e forma, rispetto a certe trasformazioni affini delle coordinate le cui associate omogenee sono le trasformazioni di Lorentz. Le trasformazioni di Lorentz “parallele” hanno tre parametri costanti; le scriveremo nella forma $x_i = a_{ij}X_j$, dove i, j variano su $(1, \dots, 4)$ e la somma sugli indici ripetuti va intesa nel senso $a_{i\eta}X_\eta - a_{i4}X_4$, (“pseudosomma”; qui e nel seguito gli indici greci corrono su $(1,2,3)$, indici latini su $(1, \dots, 4)$). Riferendoci per cominciare al caso delle trasformazioni speciali, abbiamo $a_{11} = -a_{44} \geq 1$, $a_{41} = -a_{14}$, $a_{41} = (a_{11}^2 - 1)^{1/2}$, dove si può assumere a_{11} come parametro α della trasformazione. Si verifica allora che, come deve essere, $x_1^2 - x_4^2 = X_1^2 - X_4^2$, indipendentemente dal valore di $\alpha \geq 1$. Il determinante della trasformazione è $-a_{11}a_{44} + a_{41}a_{14} = \alpha^2 - (\alpha^2 - 1) = 1$. Il caso generale si può

¹⁷ Il fatto che il nostro spazio $R^3 \times R$ abbia (pseudo)metrica indefinita, e quindi non sia topologizzabile ricorrendo ad essa, non impedisce di considerare una quaterna di funzioni reali (di un parametro reale) aventi una certa arbitraria classe di continuità. Basta usare allo scopo una conveniente topologia su R^4 , per esempio la topologia standard.

¹⁸ Continueremo a chiamare “spazio di Minkowski”, e a denotarlo con M , un generico spazio affine ad esso associato, da considerare il vero modello dello spazio-tempo relativistico speciale. Il contesto basterà in ogni caso a distinguere tra “punti-istanti” di tale spazio affine e “vettori” in senso stretto dello spazio lineare di Minkowski associato.

analizzare in modo simile, partendo dalle trasformazioni di Lorentz vettoriali riportate nell'App. Spec. 2.D. I sedici coefficienti a_{ij} risultano allora legati da tredici identità (nel caso precedente le identità tra i quattro coefficienti erano tre), per cui essi si possono esprimere in termini di soli tre parametri (per fare qualche esempio, alcune di queste identità sono $a_{14}^2 + a_{24}^2 + a_{34}^2 - a_{44}^2 = -1$, oppure $a_{12} = a_{21}$, oppure $a_{12} = a_{14}a_{24}/(1-a_{44})$, ecc.). Ciò è naturale, perché proiettando le trasformazioni vettoriali su assi coordinati spaziali *comuni* per gli osservatori s e S , si ottengono trasformazioni *parallele* omogenee, a tre parametri.

Per brevità, spesso nel seguito scriveremo p. per il prefisso “pseudo” davanti ai termini che lo vorrebbero (come in p.prodotto \equiv pseudoprodotto, ecc.). Un quadrivettore si ottiene associando ad un normale vettore spaziale una quarta componente temporale. Così ad esempio si può definire una **quadrivelocità propria** come $\gamma(v,c)$, il cui p.modulo quadrato risulta uguale a $-c^2$ (infatti $\gamma(v,c) \cdot \gamma(v,c) = \gamma^2 c^2 (v^2/c^2 - 1) = -c^2$)¹⁹; la quadrivelocità ha cioè p.modulo costante, e quindi il p.prodotto della quadrivelocità propria per la sua derivata rispetto a τ (o **quadriaccelerazione propria**) è nullo. Quanto a quest'ultima, essa risulta uguale a $\gamma^2(a + \gamma^2 a \cdot v v / c^2, a \cdot v \gamma^2 / c)$; alcuni semplici passaggi mostrano allora che effettivamente il p.prodotto soprannominato è nullo. Si potrebbe proseguire con questo tipo di considerazioni formali, ma a parte una certa loro eleganza, esse non insegnano granché di nuovo.

Come è elementarmente noto, dalle usuali coordinate cartesiane ortogonali del piano (x, y) si passa alle corrispondenti coordinate polari (ρ, θ) per le quali è $x = \rho \cos \theta$, $y = \rho \sin \theta$ (fa eccezione l'origine $x = y = 0$, che è punto singolare della trasformazione, e dove l'anomalia θ è indeterminata) date da $\rho = (x^2 + y^2)^{1/2}$ e $\theta = \arg(x, y)$ ($\arg(x, y)$ è la funzione di x e y , unicamente determinata per dato θ , per cui x e y sono proporzionali secondo un *comune fattore positivo* a $\cos \theta$ e rispettivamente a $\sin \theta$). Ci proponiamo di esaminare la simile trasformazione che fa passare dalle coordinate (ξ_1, ξ_4) del piano di Minkowski alle **coordinate pseudopolari** $|D|$ e ψ , secondo le:

$$(2_1) \quad \xi_1 = \pm |D| \operatorname{Ch} \psi, \quad \xi_4 = \pm |D| \operatorname{Sh} \psi,$$

se $D^2 > 0$, oppure

$$(2_2) \quad \xi_1 = \pm |D| \operatorname{Sh} \psi, \quad \xi_4 = \pm |D| \operatorname{Ch} \psi$$

se $D^2 < 0$. Il doppio segno nelle (2) dipende dal quadrante in cui il piano $(\xi_1, \xi_4) \setminus O$ ($O \equiv (0, 0)$) è suddiviso dalle due rette singolari $\xi_1 = \pm \xi_4$. Naturalmente risulta $\xi_1^2 - \xi_4^2 = |D|^2 = D^2$ se $D^2 > 0$ (D reale), $\xi_1^2 - \xi_4^2 = -|D|^2 = D^2$ se $D^2 < 0$ (D immaginario) e $\xi_1^2 - \xi_4^2 = 0$ se $D^2 = 0$. Nel quadrante destro vale la (2₁) col segno +, nel quadrante sinistro quella col segno -, e in entrambi $\xi_4/\xi_1 = \operatorname{Th} \psi$.

¹⁹ Lo p.modulo quadrato della velocità non è più calcolabile, come è naturale, se $|v| = c$, ma il suo limite per $|v| \rightarrow c$ è ancora $-c^2$.

Nel quadrante superiore vale la (2₂) col segno +, nel quadrante inferiore quella col segno –, e in entrambi $\xi_4/\xi_1 = \text{Cth}\psi$. Le rette che limitano i quadranti sono singolari perché il valore dello jacobiano della trasformazione, $\partial(\xi_1, \xi_4)/\partial(|D|, \psi)$, è $|D|$ o $-|D|$, e vi si annulla. Anche l'origine O è singolare per la stessa ragione (lo jacobiano della trasformazione inversa vi si annulla); e infatti ψ è indeterminata in O. Le curve $|D| = \text{cost} > 0$ sono le iperboli coniugate aventi le rette $\xi_1 = \pm \xi_4$ come asintoti, progressivamente più esterne all'aumentare di $|D|$. Le curve $\psi = \text{cost}$ sono le rette (non soltanto i raggi) lungo le quali è costante il rapporto ξ_1/ξ_4 se $\xi_4 \neq 0$, oppure, per $\psi = 0$, la retta coordinata (ξ_1), bisettrice dei quadranti destro e sinistro, se $\xi_4 = 0$. Avvolgendo l'origine in senso antiorario, la variabile ψ cresce da $-\infty$ a $+\infty$ nei quadranti destro e sinistro, e decresce da $+\infty$ a $-\infty$ nei quadranti superiore e inferiore, annullandosi sulle rette coordinate. Contemporaneamente $\text{Th}\psi$, che è monotona-crescente e continua, cresce da -1 a $+1$ nei quadranti destro e sinistro. Sulla retta singolare $\xi_1 = \xi_4$, $\text{Th}\psi = 1 = \text{Cth}\psi$, e sulla retta singolare $\xi_1 = -\xi_4$ $\text{Th}\psi = -1 = \text{Cth}\psi$. $\text{Cth}\psi$, che è monotona-decrescente e discontinua in $\psi = 0$, dove salta da $-\infty$ a $+\infty$, varia, sempre avvolgendo l'origine in senso antiorario, da $+1$ a -1 nei quadranti superiore e inferiore, saltando da $+\infty$ a $-\infty$ quando attraversa la retta coordinata (ξ_4). Infine $D^2 > 0$ nei quadranti destro e sinistro, $D^2 < 0$ nei quadranti superiore e inferiore, $D^2 = 0$ sulle rette singolari $\xi_1 = \pm \xi_4$, e $|D| = |\xi_1|$ per $\xi_4 = 0$, $|D| = |\xi_4|$ per $\xi_1 = 0$. Si noti ancora che mentre ad ogni coppia $(\xi_1, \xi_4) \neq O$ corrisponde un'unica coppia $(|D|, \psi)$, non è vero il contrario; proprio perché le rette $\xi_1 = \pm \xi_4$ sono singolari, ad ogni coppia $(|D|, \psi)$ corrispondono *due* coppie (ξ_1, ξ_4) , l'una opposta all'altra.

È facile generalizzare questa analisi della relazione tra coordinate del *piano* di Minkowski e le sue coordinate pseudopolari al caso generale dello spazio di Minkowski in senso stretto, introducendo le funzioni circolari di due addizionali angoli (reali) accanto a quelle iperboliche, per ottenere così le corrispondenti **coordinate pseudosferiche**. Anche questa possibilità si può estendere a generici spazi minkowskiani n-dim.

Rimarchiamo ora un fatto di grande importanza fisica, e cioè che è possibile una definizione *empirica*, indipendente dalle coordinate usate, del quadrato della pseudodistanza D^2 tra il punto-istante $P \neq O$ e O, quale che sia il suo segno (compreso il caso $D^2 = 0$). Disponiamo già di questa definizione nel caso in cui $D^2 < 0$: $-D^2$ è infatti c^2 volte il quadrato del tempo proprio misurato da un orologio in moto uniforme da O a $P = P(\xi, x_4)$, con $\xi \equiv (x_1, x_2, x_3)$, (se $x_4 = ct > 0$) oppure da $P(\xi, x_4)$ a O (se $x_4 = ct < 0$). Le cose non sono altrettanto semplici quando $D^2 > 0$. In questo caso, per ipotesi P è in $E \setminus \pm A$. Consideriamo la retta oraria passante per P di un orologio in moto uniforme con qualsiasi velocità v (ma ovviamente con $|v| < c$). Questa retta interseca la frontiera di $\pm A$ in due punti P_p (sulla falda del passato) e P_f (sulla falda del futuro). D'altra parte, $OP_p = OP - P_pP$ e $OP_f =$

= $OP + PP_f$, mentre OP_p e OP_f sono segmenti luminali, quindi di p.lunghezza nulla. Prendendo il quadrato * (p.prodotto per sé) delle due precedenti relazioni otteniamo

$$(3_p) \quad 0 = OP*OP + P_pP*P_pP - 2OP*P_pP,$$

e rispettivamente

$$(3_f) \quad 0 = OP*OP + PP_f*PP_f + 2OP*PP_f.$$

Basta allora tener presente che P_pP e PP_f sono paralleli ed equiversi, per cui $PP_f = kP_pP$ dove k è un numero strettamente positivo, e quindi anche $PP_f*PP_f = k^2P_pP*P_pP$. Combinando le due (3) (basta moltiplicare per k la (3_p) e sommarle la (3_f)), si ottiene

$$(4) \quad D^2 = OP*OP = (-P_pP*P_pP)^{1/2}(-PP_f*PP_f)^{1/2},$$

ovvero

$$(4') \quad D^2 = Q(OP) = [-Q(P_pP)]^{1/2}[-Q(PP_f)]^{1/2}.$$

Si osserverà che i radicandi sono positivi per la presenza dei segni $-$. Sempre dalle (3), si desume poi che nel caso particolare in cui $OP*P_pP$, e quindi anche $OP*PP_f$, si annulli (OP ortogonale a P_pP_f),

$$(4bis) \quad D^2 = OP*OP = -P_pP*P_pP = -PP_f*PP_f,$$

in accordo con la (4) stessa quando sia $P_pP*P_pP = PP_f*PP_f$. Anche questa (4bis) si può alternativamente scrivere

$$(4bis') \quad D^2 = Q(OP) = -Q(P_pP) = -Q(PP_f).$$

I due radicandi a 2° m. della (4) sono misurabili come c^2 volte i quadrati dei tempi propri dell'orologio in moto lungo la retta passante per P_p e P_f ; e per costruzione, il risultato $OP*OP$ è indipendente dalla velocità dell'orologio stesso, cioè dalla pendenza della sua retta oraria, comunque maggiore di 1 in valore assoluto.²⁰ Questa è dunque l'interpretazione di D^2 quando è positivo. Infine se $D^2 = 0$, D^2 si interpreta come c^2 volte il quadrato del tempo proprio di un orologio (virtuale) incollato ad un fotone (il tempo si arresta dunque in quell'orologio). La grandezza reale (di dimensione "lunghezza al quadrato") D^2 [ΔD^2], già definita in termini delle coordinate lorentziane [delle differenze Δ delle coordinate lorentziane omologhe], e invariante rispetto a trasformazioni di Lorentz (omogenee) [rispetto a trasformazioni di Poincaré] riceve così in ogni caso una precisa definizione operativa in termini di misure effettuate servendosi di fotoni e di un orologio opportunamente predisposto. Questo tipo di definizione si dice comunemente **cronometrizzazione alla Born** della grandezza in oggetto (nel caso presente di D^2 o ΔD^2).

L'algebra di $R^3 \times R$, con la metrica indefinita ΔD^2 , si dice **algebra di Minkowski** in onore del matematico tedesco che la studiò sistematicamente sotto lo stimolo della teoria della relatività

²⁰ Vale la pena di rimarcare che la verifica sperimentale di quest'ultimo fatto potrebbe considerarsi come una significativa, sebbene assai parziale, prova della validità del modello di Einstein-Minkowski dello spazio-tempo (in assenza di gravità).

speciale, e la cui opera contribuì anche a porre le basi della teoria geometrica della relatività generale.²¹ L'algebra di Minkowski ha anche fornito le basi della sua estensione a generici spazi pseudoeuclidei n-dim; ma ad essa si era già rivolta l'attenzione dei matematici, e in particolare di Hilbert, indipendentemente dalla relatività.

Converrà nominare come **separazione** (dall'origine) **di** x , o **magnitudine di** x , la pseudodistanza $|\sum_{i=1}^3 x_i^2 - x_4^2|^{1/2}$ e denotarla come $\text{sep}(x)$ (e come $\text{sep}^2(x)$ il suo quadrato $|\sum_{i=1}^3 x_i^2 - x_4^2|$). Tra i quadrivettori distinti dal vettore zero, quelli isotropi, e soltanto essi, hanno magnitudine nulla; tutti gli altri hanno magnitudine > 0 . Spesso i vettori isotropi vengono anche detti “nulli”, ma occorre allora ricordarsi di non confonderli con il vettore zero. Ovviamente $\text{sep}(x) \equiv \text{sep}(-x)$. Per brevità, in luogo di $\text{sep}(x-y)$ ($\equiv \text{sep}(y-x)$) si scrive anche $\text{sep}(x,y)$ ($\equiv \text{sep}(y,x)$), che diremo **separazione tra** x e y .

²¹ H. Minkowski, “Raum und Zeit”, Göttingen 1907; con lo stesso titolo, ma in una versione meno tecnica, negli atti dell'80° Congresso dei Fisici tedeschi, Colonia 21 settembre 1908. Dello stesso Minkowski, vedi anche “Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körper”, Gött. Nachr. 53 (1908), e ancora in Phys. Zeitschr. **10**, 104 (1909) e in Math. Annalen **68**, 472 (1910).

2.4) ALGEBRA E GEOMETRIA DELLO SPAZIO DI MINKOWSKI

2.4.1) PARTE PRIMA

Veniamo ora allo spazio di Minkowski in senso stretto M ($n = 4$) e alla sua geometria. Come sappiamo, M è uno spazio propriamente pseudoeuclideo $E_{4,3}$; se $\{e_1, \dots, e_4\}$ è una sua base pseudortonormale, converremo che $\{e_i\}$ (i pedici greci corrono su $(1, \dots, 3)$) siano i tre versori euclidei ($Q(e_i) = 1$), ed e_4 quello pseudoeuclideo ($Q(e_4) = -1$). Per maggior chiarezza, in questa sezione denoteremo generici quadrivettori (e loro componenti) di M con lettere maiuscole X, Y, \dots , con l'eccezione dei coefficienti a, b, \dots nella rappresentazione di rette (che sono appunto quadrivettori di M).

Ricordiamo che la componente X_i (i pedici italiani corrono su $(1, \dots, 4)$) di $X \in M$ è uguale a $\varepsilon(i)B(X, e_i)$, e risulta $B(X, Y) = X_1Y_1 - X_4Y_4$, $Q(X) = X_1X_1 - (X_4)^2$ (si sottintendono somme da 1 a 3 sugli indici greci ripetuti), invariante rispetto a trasformazioni di Lorentz. Ad R^4 , il generico elemento del quale è la quaterna ordinata $\langle X_i \rangle$, si può attribuire la struttura di $E_{4,3}$ *definendo* su di esso la relativa forma bilineare $B(X, Y)$ come $X_1Y_1 - X_4Y_4$. In questo modo B ha già automaticamente la struttura canonica pseudortonormale, e precisamente con $\varepsilon(1) = 1$, $\varepsilon(4) = -1$; quindi il suo indice è 3. La forma quadratica associata (indefinita) coincide con l'invariante canonico pseudopitagorico che abbiamo scritto come D^2 se si identificano le X_i con le componenti spaziali cartesiane ortogonali e X_4 con la componente temporale römeriana. Come sappiamo (vedi S.sez. 2.3.2), D^2 riceve una precisa definizione operativa nello spazio-tempo della relatività speciale come risultato di convenienti misure temporali (cronometrizzazione di D^2).

Passiamo ora ai più importanti teoremi della geometria di M . Tutti i (quadri)vettori cui ci si riferisce nel seguito sono tacitamente presupposti diversi dal quadrivettore "zero" (cioè si presuppone che la colonna (o riga) delle sue componenti abbia rango 1). Per cominciare, abbiamo il teorema

T1. «Due (quadri)vettori X e Y , dei quali X time-like e Y light-like (isotropo), non possono essere ortogonali.»

Dim. Converrà denotare come $\beta(X, Y)$ la 3-forma bilineare (non-degenere) X_1Y_1 , e con $\chi(X)$ la sua associata quadratica X_1X_1 , definita positiva, per le quali $2\beta(X, Y) = \chi(X+Y) - \chi(X) - \chi(Y)$. Per la 3-forma β vale la disuguaglianza (di Cauchy-Schwarz) $\beta^2(X, Y) \leq \chi(X)\chi(Y)$; l'uguaglianza si ha sse $Y = \lambda X$ per un reale $\lambda \neq 0$. Similmente scriveremo $\tau(X, Y)$ per la 1-forma bilineare X_4Y_4 , e $\theta(X)$ per

la sua associata quadratica (definita positiva) $(X_4)^2$. Evidentemente, $B(X,Y) = \beta(X,Y) - \tau(X,Y)$; l'ortogonalità di X e Y equivale dunque a

$$(1) \quad \beta(X,Y) = \tau(X,Y).$$

La natura time-like di X equivale a $\theta(X) > \chi(X)$ e quella light-like di Y a $\theta(Y) = \chi(Y)$; queste implicano

$$(2) \quad \theta(X)\theta(Y) > \chi(X)\chi(Y),$$

e dunque

$$(3) \quad \beta^2(X,Y) < \theta(X)\theta(Y).$$

Se per assurdo valesse la (1), allora $\tau^2(X,Y) < \theta(X)\theta(Y)$, cioè $(X_4Y_4)^2 < X_4^2Y_4^2$, contraddizione. #

Dal teorema (T1) segue nel modo più semplice, a maggior ragione, il teorema

T2. «Due vettori time-like non possono essere ortogonali.» Quindi un vettore ortogonale ad un vettore time-like è necessariamente space-like.

Si è già visto che un vettore light-like è ortogonale a se stesso. Ci si può chiedere se due vettori light-like X e Y *non collineari* possono essere ortogonali. In tal caso si avrebbe $(\alpha X + \beta Y) = 0$ per α e β arbitrari, e quest'ultima mostra che la risposta è negativa. Passiamo ora al teorema

T3. «I coni solidi forward \dot{A} e backward $-\dot{A}$ sono stabili rispetto alla somma.» Questo è già stato affermato senza prova al momento di definire quei coni; vogliamo ora verificarlo.

Dim. Dalla (3) si ha $\beta(X,Y) < [\theta(X)\theta(Y)]^{1/2}$, quindi $B(X,Y) < [\theta(X)\theta(Y)]^{1/2} - \tau(X,Y) \equiv |X_4Y_4| - X_4Y_4$. Ma per ipotesi X_4 e Y_4 hanno lo stesso segno, positivo nel caso forward, negativo in quello backward, e dunque il secondo membro della precedente disuguaglianza è in realtà zero, ovvero $B(X,Y) < 0$. D'altra parte $Q(X+Y) = Q(X) + Q(Y) + 2B(X,Y)$, e tutti i termini a secondo membro sono negativi, da cui la tesi. #

La precedente conclusione può essere estesa dicendo che «la somma di un vettore time-like e di uno light-like entrambi forward o entrambi backward è time-like, e quella di due vettori light-like, sempre entrambi forward o entrambi backward, è time-like o light-like». Riassumendo, si può affermare che anche A e $-A$ sono stabili rispetto alla somma.

La stima $B(X,Y) < 0$, per X e Y in \dot{A} (o in $-\dot{A}$), può essere migliorata: vale infatti il teorema

T4. «Se X e Y sono entrambi in \dot{A} (o in $-\dot{A}$), allora

$$(4) \quad B(X,Y) \leq -[Q(X)Q(Y)]^{1/2}.$$

Dim. Per ipotesi, $\theta(X) = \chi(X) - Q(X)$, $\theta(Y) = \chi(Y) - Q(Y)$, quindi $B(X,Y) = \beta(X,Y) - [(\chi(X) - Q(X))(\chi(Y) - Q(Y))]^{1/2}$. D'altra parte (Cauchy-Schwarz):

$$(5) \quad \beta(X,Y) \leq [\chi(X)\chi(Y)]^{1/2}$$

(ove l'uguaglianza si ha sse X e Y sono collineari-equiversi, cioè collineari con fattore > 0), per cui $B(X,Y) \leq [\chi(X)\chi(Y)]^{1/2} - [(\chi(X) - Q(X))(\chi(Y) - Q(Y))]^{1/2}$. Il secondo membro di quest'ultima è negativo (confermando la stima precedente di $B(X,Y)$), e il suo valore assoluto è $\geq [Q(X)Q(Y)]^{1/2}$ (avendosi uguaglianza sse $\chi(X) = \chi(Y) = 0$)¹. La tesi è così dimostrata. #

Il ragionamento che porta alla (4) resta valido se X e/o Y sono light-like, nel qual caso essa diventa

$$(4bis) \quad B(X,Y) \leq 0.$$

(T3) può così essere esteso ad una coppia di vettori entrambi in A (o in $-A$). Due vettori time- o light-like entrambi in A (o in $-A$) si dicono **equicroni**. Nella (4) il segno = vale sse vale nella (5) quindi sse $Y_i = \lambda X_i$ con $\lambda > 0$. Ma si vede subito, allora, che anche $Y_4 = \lambda X_4$; ovvero il segno = vale sse $Y = \lambda X$ con $\lambda > 0$. In particolare l'uguaglianza nella (4bis) vale sse X è light-like e $Y = \lambda X$, con $\lambda > 0$. Quadrando la (4), si ha una disuguaglianza di tipo Cauchy-Schwarz ma con \geq al posto di \leq , vale a dire $B^2(X,Y) \geq Q(X)Q(Y)$. Quest'ultima è peraltro più permissiva della (4), perché se vale per X e Y vale anche per X e $-Y$; ovvero, si applica a generici vettori time-like e/o light-like, non necessariamente equicroni.

In generale (cioè per generiche forme pseudoeuclidee $B \leftrightarrow Q$ su uno spazio R -lineare E finito-dim), l'ortogonalità di $x \in E$ e $y \in E$ è equivalente alla condizione

$$(6) \quad Q(x+y) = Q(x) + Q(y) \quad ^2.$$

Tenuto conto di questo, e tornando a M , abbiamo il teorema

T5. «L'ortogonalità di X (time-like) e Y (space-like) equivale al fatto che la loro somma è light-like.» La dimostrazione è immediata. Poiché l'ortogonalità è banalmente stabile rispetto alla moltiplicazione per reali, segue il seguente

C1. Criterio di ortogonalità tra un vettore time-like e un vettore space-like: « X (time-like) e Y (space-like) sono ortogonali sse, avendoli "normalizzati", cioè avendoli resi pseudounitario e rispettivamente unitario mediante moltiplicazione per convenienti fattori (il supporre questi fattori > 0 non infirma la generalità), la loro somma è light-like.»

Applicando la (6) a due vettori X, Y space-like ortogonali, si vede subito che la loro somma è space-like. Segue il

¹ Questo asserto diventa del tutto evidente ricordando che $B(X,Y)$, $Q(X)$ e $Q(Y)$ sono invarianti rispetto a trasformazioni di Lorentz. Se si usa, come è sempre possibile, un riferimento in cui $\chi(Y) = 0$, il valore assoluto in questione si riduce a $[(\chi(X) - Q(X))(-Q(Y))]^{1/2} \geq [(-Q(X))(-Q(Y))]^{1/2}$, qed.

² Poiché in generale $Q(x+y) - Q(x-y) = 4B(x,y)$, se x e y sono ortogonali $Q(x+y)$ e $Q(x-y)$ sono uguali.; quindi la (6) può risciversi con $Q(x-y)$ al posto di $Q(x+y)$. Il contenuto di alcuni teoremi che seguono può dunque riformularsi sostituendo "differenza" a "somma". Per inciso, si vede subito, inoltre, che $Q(x+y) + Q(x-y) = 2(Q(x) + Q(y))$; quest'ultima identità è detta da alcuni **del parallelogramma**.

C2. Criterio di ortogonalità tra due vettori space-like: «X e Y, entrambi space-like, sono ortogonali sse, avendoli normalizzati, risulta $Q(X+Y) = 2$.» (questo non è altro che il teorema di Pitagora).

Abbiamo ancora un ovvio

C3. Criterio di ortogonalità tra un vettore space-like e un vettore light-like: «X (space-like) e Y (light-like) sono ortogonali sse, avendo normalizzato X, risulta $Q(X+Y) = 1$ » (cioè, sse la loro somma è space-like unitaria).

Veniamo ora ad esaminare la questione della disuguaglianza triangolare, cominciando dal caso di due vettori entrambi time-like e ortocroni. Abbiamo innanzitutto il teorema

T6. « Due vettori X e Y entrambi time-like e ortocroni soddisfano alla

$$(7) \quad \text{sep}(X+Y) \geq \text{sep}(X) + \text{sep}(Y)».$$

Dim. In ogni caso, abbiamo $Q(X+Y) = Q(X) + Q(Y) + 2B(X,Y)$. Sotto le ipotesi, e tenendo conto della (4), $\text{sep}^2(X+Y) = \text{sep}^2(X) + \text{sep}^2(Y) - 2B(X,Y) \geq \text{sep}^2(X) + \text{sep}^2(Y) + 2\text{sep}(X)\text{sep}(Y) = [\text{sep}(X) + \text{sep}(Y)]^2$, qed. #

Segue il teorema

T6bis. «T6 si estende a due vettori X e Y time- o light-like equicroni, possibilmente modificando in modo ovvio la (7)»

Dim. Se X e/o Y sono light-like, basta usare la (4bis) in luogo della (4); si giunge così ancora alla (7), avendovi modificato il 2° membro in $\text{sep}(X)$ o $\text{sep}(Y)$ o zero, a seconda che Y, o X, o entrambi, siano light-like. # ³

Supponendo X e Y time-like e *anticroni*, possiamo riscrivere la precedente (7) in una forma più espressiva ponendovi $-Y$ in luogo di Y, e ottenendo:

$$(7bis) \quad \text{sep}(X-Y) \equiv \text{sep}(X,Y) \geq \text{sep}(X) + \text{sep}(Y)$$

(**disuguaglianza triangolare inversa**), da eventualmente semplificare al solito modo se uno o entrambi i vettori sono light-like. La (7bis) non dipende da dove abbiamo posto l'origine del riferimento. Denotando questa origine come Z, otteniamo:

$$(8) \quad \text{sep}(X,Y) \geq \text{sep}(X,Z) + \text{sep}(Z,Y),$$

valida per tre punti X, Y, Z tali che $Z-X$ e $Y-Z$ siano entrambi in A (o in $-A$), e per il resto arbitrari. Questa è la forma più usuale in cui è presentata la disuguaglianza triangolare inversa.

Consideriamo adesso una situazione complementare alla precedente, per la quale sussiste il teorema

³ Se X e Y sono entrambi light-like, allora la (7) è un'eguaglianza (del tipo $0 = 0$) anche se X e Y sono semplicemente collineari (e non necessariamente ortocroni).

T7. «Se X e Y sono entrambi space-like, e la loro somma e la loro differenza sono space- o light-like (cioè se $Q(X) > 0$, $Q(Y) > 0$, $Q(X+Y) \geq 0$, $Q(X-Y) \geq 0$), allora

$$(9) \quad \text{sep}(X+Y) \leq \text{sep}(X) + \text{sep}(Y)».$$

Dim. Il fatto che $X - Y$ (quindi anche $Y - X$) sia space- o light-like implica che $0 \leq Q(X-Y) = Q(X) + Q(Y) - 2B(X,Y)$, ossia che $B(X,Y) \leq Q(X) + Q(Y)$, ove i due termini a 2° membro sono strettamente positivi. Moltiplicando X per $\sqrt{\mu}$ (con $\mu > 0$) e Y per $1/\sqrt{\mu}$, il 1° membro non cambia, quindi $B(X,Y) \leq \mu Q(X) + \mu^{-1}Q(Y)$. Considerato come funzione di μ , il 2° membro di questa ha un *minimo* per $\mu = [Q(Y)/Q(X)]^{1/2}$, in corrispondenza del quale vale $2[Q(X)Q(Y)]^{1/2}$. Si conclude che $Q(X+Y) = Q(X) + Q(Y) + 2B(X,Y) \leq Q(X) + Q(Y) + 2[Q(X)Q(Y)]^{1/2}$. Ma $Q(X+Y) \geq 0$ per ipotesi, e quindi può scriversi come $\text{sep}^2(X+Y)$, mentre il 2° membro è uguale a $[(Q(X))^{1/2} + (Q(Y))^{1/2}]^2$, dal che scende immediatamente la (9). Si rimarca che la dimostrazione *non* vale se X e/o Y sono light-like. #

Le ipotesi che legittimano la (9) restano invariate se vi si scambia Y con $-Y$; quindi la (9) può riscriversi come

$$(9\text{bis}) \quad \text{sep}(X,Y) \leq \text{sep}(X) + \text{sep}(Y);$$

(disuguaglianza triangolare diretta);

ovvero, sostituendo Z all'origine,

$$(10) \quad \text{sep}(X,Y) \leq \text{sep}(X,Z) + \text{sep}(Z,Y),$$

per tre punti X, Y, Z per i quali $Q(X-Z) > 0$, $Q(Y-Z) > 0$, $Q(X-Y) \geq 0$ e $Q(X-Z+Y-Z) \geq 0$. La (10) è la forma più usuale in cui viene presentata la disuguaglianza triangolare diretta.

Siano adesso X, Y, Z tre punti *distinti* qualsiasi che soddisfano la

$$(11) \quad \text{sep}(X,Y) = \text{sep}(X,Z) + \text{sep}(Y,Z)$$

(uguaglianza triangolare). Ci si chiede quando essa sia soddisfatta. Supponiamo $Z - X$ e $Y - X$ collineari, $Z-X = \lambda(Y-X)$ ($\Rightarrow \lambda \neq 0$), e di conseguenza $Y-Z = (1-\lambda)(Y-X)$ ($\Rightarrow \lambda \neq 1$). Allora $\text{sep}(X,Z) = |\lambda|\text{sep}(X,Y)$ e $\text{sep}(Y,Z) = |1-\lambda|\text{sep}(X,Y)$, per cui la (11) richiede $|\lambda| = |1-\lambda|$, ossia $0 < \lambda < 1$ (i valori $\lambda = 0$ e $\lambda = 1$ non sono ammessi, come già più sopra rilevato). Quindi la (11) è soddisfatta se X, Z, Y sono collineari e Z è “tra” X e Y . Un'altra possibilità è che X, Y, Z (distinti) siano sulla stessa retta isotropa (la dimostrazione è banale). Vale il teorema, che ci limitiamo ad enunciare:

T8. «L'uguaglianza (11) è soddisfatta, per X, Y, Z distinti, soltanto nei due casi sopra considerati; quindi in tutti gli altri casi vale una disuguaglianza (diretta o inversa) *stretta*».

(T8) ci permette di caratterizzare il segmento (rettilineo) tra X e Y , con $Y-X$ non light-like, come il luogo di tutti e soli i punti Z per i quali vale l'*uguaglianza triangolare* (11). Il punto Z si

dice allora **allineato con X e Y, e tra essi**. Tre punti che non siano sulla stessa retta isotropa si dicono **allineati** se per essi vale l'uguaglianza triangolare (11), indipendentemente da quale dei tre sia tra gli altri due. Con l'eccezione del caso di punti X, Y sulla stessa retta isotropa, abbiamo così, in M , una caratterizzazione della nozione di **segmento tra X e Y**, o anche di **retta per X, Y**, completamente indipendente dalla coordinatazione, ed espressa in soli termini di separazione, che come sappiamo è cronometrizzabile (se X e Y non sono sulla stessa retta isotropa). Tali segmenti o rette possono eventualmente essere orientati partendo dalla nozione dell'“essere tra”, vedi Cap. 1. Quanto al segmento isotropo tra X e Y (possibilmente orientato), esso coincide con la corrispondente traiettoria luminale, e dunque è ancora fisicamente determinato in modo unico. Poiché abbiamo già dato la definizione parametrica di questi oggetti in un generico spazio lineare, il precedente asserto va giustificato; il che non è difficile (solo un po' noioso), e non viene qui fatto esplicitamente per brevità (cfr. l'analogo problema nello spazio euclideo).

2.4.2) PARTE SECONDA

Consideriamo una successione finita di $N+1$ punti distinti di M , $H = {}^0X, {}^1X, \dots, {}^{N-1}X, {}^NX = K$, supponendo che ogni segmento ${}^iX - {}^{i-1}X$ (con $1 \leq i \leq N$) appartenga al cono solido A (oppure a $-A$); allora anche $K-H$ gli appartiene (cfr. 2.4.1, (T3)), e risulta

$$(1) \quad \text{sep}(H,K) \geq \sum_{i=1}^N \text{sep}({}^{i-1}X, {}^iX).$$

Il caso che $K-H$ sia isotropo non è escluso; ma allora, si constata facilmente, ogni segmento nella somma deve essere a sua volta isotropo, e la (1) diventa un'uguaglianza.

Similmente, consideriamo un'analogha successione di punti sotto l'ipotesi che ogni segmento ${}^iX - {}^{i-1}X$ appartenga a $[E \pm \mathring{A}] \cup \{O\}$, e che anche $K-H$ gli appartenga; allora vale la disuguaglianza (1) rovesciata, cioè con \leq al posto di \geq :

$$(2) \quad \text{sep}(H,K) \leq \sum_{i=1}^N \text{sep}({}^{i-1}X, {}^iX).$$

Ancora, la possibilità che $K-H$ sia isotropo non è esclusa; ma allora ogni segmento deve essere a sua volta isotropo, e la (2) diventa ancora una uguaglianza. In ogni caso, la possibilità che la (1) oppure la (2) si riducano ad uguaglianze equivale all'allineamento orientato dei punti della successione, ovvero, secondo una notazione già introdotta nel Cap 1, con riferimento a uno spazio euclideo e per $N = 2$, a $[H, {}^1X, \dots, {}^{N-1}X, K]$.

Sia adesso Γ una curva di classe C^2 tra H e K del tipo $X = X(\lambda)$, con $0 \leq \lambda \leq 1$, orientata secondo le λ crescenti. Supporremo innanzitutto che il vettore tangente di Γ , $\tau =: dX/d\lambda = \tau(\lambda)$ (che

è C^1 per ipotesi) sia sempre $\neq 0$, e che $\tau(\lambda)$ non sia mai isotropo lungo Γ . Allora la **separazione differenziale** $\text{sep}(X(\lambda+d\lambda), X(\lambda))$ è positiva (se $d\lambda > 0$), e $\int_{\lambda=0}^1 \text{sep}(X(\lambda+d\lambda), X(\lambda))$ è ben definito. Si imporranno a Γ limitazioni analoghe, ma leggermente più restrittive (in quanto si deve evitare che $\tau(\lambda)$ sia isotropo), a quelle precedenti a proposito delle spezzate nella (1) e rispettivamente nella (2), e che formuleremo come segue:

a) se $K - H$ è time-like,

$$(1\text{bis}) \quad \forall \lambda \text{ in } [0,1], \tau(\lambda) \in \mathring{A} \text{ (oppure } \in -\mathring{A}\text{);}$$

b) se $K - H$ è space-like,

$$(2\text{bis}) \quad \forall \lambda \text{ in } [0,1], \tau(\lambda) \in E \setminus \pm A.$$

Per $\Delta\lambda$ abbastanza piccolo, e per qualunque λ in $[0,1]$, esiste (finito) il limite, per $\Delta\lambda \rightarrow 0$, del rapporto incrementale $\text{sep}(X(\lambda+\Delta\lambda), X(\lambda))/\Delta\lambda$, che diremo $w_\Gamma(\lambda)$, per ipotesi di classe C^1 , e che risulta > 0 . Introducendo l'integrale (tra 0 e 1) $\int w_\Gamma(\lambda)d\lambda$, è allora intuitivamente evidente (e si può dimostrare) che, comunque si scelga Γ sotto i vincoli convenuti in (1bis) o (2bis) a seconda dei casi, e sempre sottintendendo che gli integrali siano tra 0 e 1,

$$(3) \quad \text{sep}(H,K) > \int w_\Gamma(\lambda)d\lambda$$

se $K-H$ è time-like, e

$$(4) \quad \text{sep}(H,K) < \int w_\Gamma(\lambda)d\lambda$$

se $K-H$ è space-like. Si notino i segni ($>, <$) in luogo dei (\geq, \leq) nelle (3,4).

Siamo ormai in grado di formulare il seguente problema variazionale.⁴ Siano H e K dati arbitrariamente purché con $K - H$ non light-like, e consideriamo una curva Γ di classe C^2 sotto le condizioni espresse in (1bis) oppure (2bis) a seconda della natura di $K-H$. L'insieme delle siffatte curve tra H e K si dirà **insieme ammissibile rispetto a $K-H$** , o **($K-H$)-ammissibile**. Per ogni Γ nell'insieme ($K-H$)-ammissibile, il segmento rettilineo da H a K massimizza ($K-H$ time-like) o minimizza ($K-H$ space-like) l'integrale $\int w_\Gamma(\lambda)d\lambda$. In altre parole, al variare di Γ nell'insieme ($K-H$)-ammissibile, tale segmento è una estrema del problema variazionale $\delta \int w_\Gamma(\lambda)d\lambda = 0$; e si può dimostrare che tale estrema è unica. Abbiamo insomma una terza, e particolarmente significativa, caratterizzazione del segmento rettilineo orientato $K-H$. In effetti le equazioni di Eulero (o Euler (Leonhard, Basilea 1707, S. Pietroburgo 1783)) e Lagrange di tale problema variazionale sono del tipo $d^2X/d\lambda^2 = 0$. Integrandole tenendo conto delle condizioni ai limiti, esse

⁴ Ci affidiamo in parte all'intuizione e in parte alle presunte preconoscenze del lettore nei due paragrafi che seguono. Per una rassegna accettabilmente organica del calcolo delle variazioni, si vedano i capitoli 6 e 7.

danno immediatamente $X(\lambda) = H + (K-H)\lambda$. Si ottengono così le equazioni parametriche del segmento rettilineo orientato (secondo le λ crescenti) da $H (= X(\lambda=0))$ a $K (= X(\lambda=1))$.

Tornando al caso generale, il parametro λ si dice **separazione normalizzata**, di $X(\lambda)$ da H , lungo Γ . L'integrale rispetto a X di $w_\Gamma(X(\lambda))$ da 0 a λ , con $0 \leq \lambda \leq 1$, diciamo $s(\lambda)$, di classe C^2 , si dice **pseudolunghezza** (o **misura**) della porzione di Γ compresa tra H e $X(\lambda)$; il suo valore per $\lambda = 1$ è la pseudolunghezza di Γ . Infine, risulta $ds/d\lambda = w_\Gamma$. Sotto le condizioni convenute, nulla vieta di usare s come parametro naturale lungo Γ ; in questo caso la sopravvissuta equazione variazionale si scrive semplicemente $\delta \int ds = 0$. Se $K-H$ e Γ sono time-like e forward, il parametro s non è altro che c volte il tempo proprio τ di un orologio che viaggia lungo la curva oraria Γ , regolato in modo che segni 0 in H . L'equazione di Eulero-Lagrange associata al problema, $d^2X/d\tau^2 = 0$, traduce precisamente la legge dinamica del punto materiale sulla curva d'universo Γ in assenza di forza. Si ritrova così la legge d'inerzia per quel punto, la cui traiettoria è descritta da $X = X(\tau) = a\tau + h$, dove a e h sono due vettori costanti, sotto il vincolo $Q(a) < 0$, che si determinano mediante le solite condizioni accessorie. Infine se $K - H$ è light-like, il segmento rettilineo che ha K e H per estremi soddisfa ancora, ovviamente, la $d^2X/d\tau^2 = 0$, ma questa *non* può più considerarsi equazione di EL di un problema variazionale del tipo $\delta \int ds = 0$. $K - H$ va cioè pensato come limite di segmenti rettilinei $K' - H$ space-like, oppure time-like (i quali sono tutti estremali della $\delta \int ds = 0$) allorché $K' \rightarrow K$.

Dobbiamo ancora introdurre la nozione di **piede**, o **proiezione ortogonale** Z' su una retta *non light-like* q di un punto Z fuori di essa. Converrà supporre q passante per l'origine O e definirla mediante un suo punto Y ($Q(Y) \neq 0$). Per definizione, è $B(Z-Z', Y) = 0$, ove per ipotesi $Z \neq Z'$; ricerchiamo un reale λ per cui $Z' = \lambda Y$. Se $\lambda = 0$, il piede Z' di Z sulla retta q è l'origine O , come è giusto perché allora $B(Z, Y)$ si annulla. Escludiamo questo caso supponendo $\lambda \neq 0$: poiché $Z = Z' + (Z-Z')$, risulta $Q(Z) = \lambda^2 Q(Y) + Q(Z-\lambda Y)$. Ma $Q(Z-\lambda Y) = -2\lambda B(Y, Z) + Q(Z) + \lambda^2 Q(Y)$, e quindi

$$(5) \quad \lambda[\lambda Q(Y) - B(Y, Z)] = 0,$$

cioè, poiché abbiamo supposto $\lambda \neq 0$;

$$(6) \quad Z' = YB(Y, Z)/Q(Y) \quad ^5.$$

Quest'ultima dà $Z' = O$ se $B(Y, Z) = 0$, come deve essere. Se X è un punto corrente di q , risulta $Q(Z-X) = Q(Z'-X) + Q(Z'-Z)$; dunque, se q è space-like, $Q(Z-X) \geq Q(Z'-Z)$, e se q è time-like, $Q(Z-X) \leq Q(Z'-Z)$. Nel primo caso (q space-like), se $Z'-Z$ è space-like lo è anche $Z-X$, e

⁵ Si noti che se, in deroga alla restrizione convenuta in apertura, $Q(Y) \neq 0$, si facesse $Q(Y) = 0$ nella (5) (con $\lambda \neq 0$), allora proprio secondo la (5) $B(Y, Z) = 0$, e la (6) diventerebbe inutilizzabile. In realtà, se p è light-like, non ha senso ricercare il piede su di essa di Z (che è fuori di p), perché in questo caso $\text{sep}(Z, X) = \text{cost}$ al variare di X su p .

$\text{sep}(Z,X) \geq \text{sep}(Z,Z') \forall X \neq Z'$, cioè $X = Z'$ è *minimo assoluto* (> 0) per $\text{sep}(Z,X)$. Analogamente, se $Z'-Z$ è light-like, $X = Z'$ è *minimo assoluto* ($= 0$) per $\text{sep}(Z,X)$. Se poi $Z'-Z$ è time-like, allora $Q(Z-X)$ (che è negativo per $X = Z'$), aumenta all'allontanarsi di X da Z' , si annulla per $Q(Z'-X) = Q(Z'-Z)$, e poi continua ad aumentare con valori positivi; quindi $\text{sep}(X,Z)$, positiva per $X = Z'$, diminuisce fino al punto ove si annulla, e poi torna ad aumentare, ossia $X = Z'$ è un *massimo relativo* per $\text{sep}(X,Z)$. Nel secondo caso (q time-like), sappiamo che $Z'-Z$ è space-like; quando X si allontana da Z' , $Q(Z-X)$ diminuisce, si annulla per $Q(Z'-X) = -Q(Z-Z')$ e continua a diminuire con valori negativi. Quindi $X = Z'$ è *massimo assoluto* (> 0) per $\text{sep}(Z,X)$. Abbiamo così una *caratterizzazione differenziale* per il piede Z' di Z su q (non light-like) applicabile in ogni caso.

Esamineremo ora la nozione di parallelismo, in particolare tra rette orientate, nello spazio di Minkowski M . Siano p e q due generiche rette orientate di equazioni parametriche $X = X(\lambda) = a\lambda + H$ e rispettivamente $Y = Y(\mu) = b\mu + K$, dove a, b, H, K , sono (quadri)vettori, i primi due distinti da zero, e λ, μ sono parametri reali. Per definizione, diremo che p e q sono **parallele equiverse** se $a = b$, e **parallele antiverse** se $a = -b$, e semplicemente **parallele** se $a = b$ aut $a = -b$. Si verifica subito che il parallelismo equiverso e il parallelismo sono relazioni di equivalenza. Se le due rette sono parallele ed hanno un punto in comune, allora sono gli stessi luoghi geometrici. Dall'ipotesi scende infatti che esistono valori $\underline{\lambda}$ e $\underline{\mu}$ dei parametri tali che $a\underline{\lambda} + H = \pm a\underline{\mu} + K$; per cui, eliminando ad es. K , segue che l'equazione di q è $Y(\mu) = a[\mu - (\underline{\mu} - \underline{\lambda})] + H$, oppure $Y(\mu) = -a[\mu - (\underline{\mu} + \underline{\lambda})] + H$; vale a dire, a parte l'orientamento rispetto al parametro, cambia soltanto il valore di questo per cui X o Y è uguale a H , che è $\lambda = 0$ per p , e $\mu = \underline{\mu} - \underline{\lambda}$, oppure $\mu = \underline{\mu} + \underline{\lambda}$, per q . Quando poi $a \neq \pm b$, o le rette si intersecano, ossia esistono valori dei parametri che danno lo stesso punto, o questi valori non esistono, e allora esse si dicono **sghembe**.

Ricercheremo ora una caratterizzazione del parallelismo in termini di separazione. A questo scopo, cominceremo con l'escludere che a e b , definiti come sopra, siano vettori isotropi, e (ovviamente) che siano ortogonali. Senza limitazioni di generalità, supporremo anche che $|Q(a)| = |Q(b)| = 1$. Se $Y(\mu) \in q$, sia $Y' = Y'(\lambda)$ il suo piede su p ; possiamo dunque eliminare μ in favore di λ (ad es.) mediante la richiesta $B(a,\delta) = 0$, ove $\delta =: Y - Y'$. Il risultato è

$$(7) \quad \mu = B^{-1}(a,b)[\lambda Q(a) + B(a,\eta)],$$

ove abbiamo posto $\eta =: H - K$ (al solito, per brevità abbiamo scritto $B^{-1}(\quad, \quad)$ per $(B(\quad, \quad))^{-1}$, ecc.) In definitiva si trova che $\delta = \alpha\lambda + \beta$, con

$$(8) \quad \alpha =: B^{-1}(a,b)Q(a)b - a,$$

$$(9) \quad \beta =: B(a,\eta)B^{-1}(a,b)b - \eta.$$

Richiederemo ora che $Q(\delta)$ non dipenda da λ ; in questo caso anche $\text{sep}(Y, Y')$ non ne dipende, e si dirà **separazione ortogonale di q da p**. Studiamo come questa richiesta influisce sul possibile parallelismo tra p e q. Innanzitutto risulta

$$(10) \quad Q(\delta) = \lambda^2 Q(\alpha) + 2\lambda B(\alpha, \beta) + Q(\beta),$$

per cui dobbiamo imporre

$$(11) \quad Q(\alpha) = Q(a)[Q(a)Q(b)B^{-2}(a, b) - 1] = 0,$$

$$(12) \quad B(\alpha, \beta) = B(a, \eta) - B^{-1}(a, b)Q(a)B(b, \eta) = 0.$$

Avendo supposto a e b non isotropi, la (11) mostra che $Q(a)$ e $Q(b)$ devono avere lo stesso segno, quindi che $Q(a)Q(b) = 1$, ossia $Q(a) = Q(b)$, e che $B^{-2}(a, b) = 1$. Allora la (12) potrà anche scriversi come

$$(13) \quad B(a, \eta) - B(a, b)Q(a)B(b, \eta) = 0,$$

perché $B^{-1}(a, b) = B(a, b)$. Si noti che η deve essere diverso da zero, perché altrimenti le due rette in esame si intersecherebbero. Moltiplicandola per $B(a, b)Q(b)$, la (13) si può equivalentemente scrivere come

$$(13\text{bis}) \quad B(b, \eta) - B(a, b)Q(b)B(a, \eta) = 0,$$

che è la sua versione con l'interscambio di a con b. In definitiva il sistema (11,12) (o (11,13)) equivale al suo simmetrico rispetto allo scambio di a con b.

Riferendoci ad es. alla (13), è immediato verificare che essa è soddisfatta per $a = b$ e per $a = -b$, cioè che il parallelismo implica la costanza di $Q(Y - Y')$ (o equivalentemente di $Q(X - X')$, dove $X \in p$ e X' è il suo piede su q). Tuttavia non è detto che l'implicazione inversa valga in generale. Mostriamo ora che $a = b$ consegue dall'ipotesi che le rette p e q, orientate secondo i parametri λ e μ crescenti, siano entrambe time-like, equicrone e abbiano $Q(\delta)$ costante. Sempre per a e b pseudounitari, abbiamo

$$(14) \quad Q(a) = Q(b) = -1$$

(che implica $Q(a)Q(b) > 0$),

$$(15) \quad B^2(a, b) = 1, \text{ e}$$

$$(16) \quad a_4 b_4 > 1.$$

La tesi è dunque che dalle (14,15,16) consegue l'uguaglianza $a = b$. Come sappiamo, per due vettori time-like e equicroni come a e b deve essere $B(a, b) < 0$; e quindi la (15) implica che

$$(17) \quad B(a, b) = -1.$$

A parte un possibile riordinamento dei tre indici 1, 2, 3, la più generale espressione di siffatti a e b, supposti per fissare le idee entrambi forward, è la seguente:

$$(18_a) \quad a_1 = \text{Sh}\chi \sin\theta \cos\varphi, \quad a_2 = \text{Sh}\chi \sin\theta \sin\varphi, \quad a_3 = \text{Sh}\chi \cos\theta, \quad a_4 = \text{Ch}\chi,$$

e similmente,

$$(18_b) \quad b_1 = \text{Sh}\chi' \sin\theta' \cos\varphi', \quad b_2 = \text{Sh}\chi' \sin\theta' \sin\varphi', \quad b_3 = \text{Sh}\chi' \cos\theta', \quad b_4 = \text{Ch}\chi'.$$

Qui θ e θ' sono colatitudini (quindi variano in $[0, \pi]$), φ e φ' sono longitudini (quindi variano in $[0, 2\pi)$), e χ e χ' variano in $(-\infty, +\infty)$. È immediato verificare che le (18) verificano le due (14) e la (16) indipendentemente dai valori dei sei parametri, per il momento arbitrari, χ , θ , φ , χ' , θ' , φ' . Resta dunque da imporre la (17). Questo vuol dire che

$$(19) \quad \text{Ch}\chi \text{Ch}\chi' - \text{Sh}\chi \text{Sh}\chi' [\cos\theta \cos\theta' + \sin\theta \sin\theta' \cos(\varphi - \varphi')] = 1.$$

È facile dimostrare che il quadrato del contenuto delle [] è comunque ≤ 1 ⁶; quindi tale contenuto può scriversi come $\cos\psi$, per un ψ unicamente determinato in $[0, \pi]$. Trascriveremo di conseguenza la (19) come

$$(19\text{bis}) \quad 1 = \text{Ch}\chi \text{Ch}\chi' - \text{Sh}\chi \text{Sh}\chi' \cos\psi.$$

Ci sono tre e soltanto tre casi in cui questa (19bis) è soddisfatta, e cioè (i) se $\chi = \chi' = 0$; (ii) se $\chi = \chi'$ e $\psi = 0$; (iii) se $\chi = -\chi'$ e $\psi = \pi$. Alla luce delle (18), in tutti e tre i casi l'unica possibilità è $a = b$ (lasciamo al lettore di verificarlo). Se poi a e b sono entrambi backward, basterà porre un segno meno nei 2ⁱ membri delle quarte delle (18) e procedere in modo simile. La tesi è così provata. Infine sappiamo che nel caso delle rette time-like equicrone il vettore di separazione $\delta = Y - Y' = -(X - X')$ è space-like.

Possiamo ormai affrontare l'ultima questione di cui ci occupiamo in questa breve rassegna sulla geometria dello spazio-tempo propriamente pseudoeuclideo: quella cioè della preparazione di una base pseudortonormale. Consideriamo un fascio di parallele time-like equicrone forward, quindi di equazioni parametriche del tipo $X = X(\lambda) = a\lambda + H$, dove a è comune a tutte le rette del fascio, mentre λ (crescente dal passato al futuro) e H sono specifici per ogni retta, e dove potrà ancora assumersi $Q(a) = -1$ senza limitare la generalità. Evidentemente, la generica parallela del fascio può interpretarsi come linea oraria di un corrispondente orologio normale standard, il cui tempo proprio τ , o meglio una grandezza adimensionale ad esso proporzionale, può usarsi come parametro λ . Per rendere le formule che seguono più espressive, continueremo tuttavia a scrivere τ per questo parametro adimensionale proporzionale al tempo proprio degli orologi. Siano ora p e q due parallele del fascio, di equazioni parametriche $X = X(\tau) = a\tau + H$ e rispettivamente $Y = Y(\tau') = a\tau' + K$, e sia al solito $Y' = Y'(\tau)$ il piede su p di $Y(\tau') \in q$. In forza dell'ortogonalità tra a e $Y(\tau') - Y'(\tau)$, come sappiamo risulta $\tau' = \tau - B(a, \eta)$ (con $\eta = H - K$). Sembrerebbe a prima vista

⁶ Scrivendolo come $Lx^2 + 2Mx + N$, con $L =: \sin^2\theta \sin^2\theta' \geq 0$, $M =: \cos\theta \cos\theta' \sin\theta \sin\theta'$, $N =: \cos^2\theta \cos^2\theta' \geq 0$, $x =: \cos(\varphi - \varphi')$, si vede subito che il suo massimo per x in $[-1, 1]$ vale $L + 2|M| + N = L + 2\sqrt{LN} + N = (\sqrt{L} + \sqrt{N})^2$. Quest'ultimo è comunque il quadrato di un coseno (posto $A =: \cos\theta \cos\theta'$, $C =: \sin\theta \sin\theta'$, basta esaminare separatamente le quattro possibilità $(A \geq 0, C \geq 0)$, $(A \geq 0, C \leq 0)$, $(A \leq 0, C \geq 0)$, $(A \leq 0, C \leq 0)$) e quindi ≤ 1 , qed.

sensato, a questo punto, pensare di sincronizzare gli orologi O (su p) e O' (su q) in modo che $\tau' = 0$ corrisponda a $\tau = 0$, richiedendo quindi che $\eta = X(\tau=0) - Y(\tau'=0)$ sia ortogonale ad a , ossia che $X(\tau=0)$ sia il piede su p di $Y(\tau'=0)$ ($\in q$), o viceversa che $Y(\tau'=0)$ sia il piede su q di $X(\tau=0)$ ($\in p$). Ciò è tuttavia impossibile, perché la velocità di propagazione di qualunque segnale è finita, e non si può comunicare ad O' di regolarsi su $\tau' = 0$ “esattamente quando” O segna $\tau = 0$ (o viceversa). È invece possibile che, al suo tempo proprio τ , O invii un fotone ad O' , che lo riceve in $\Omega =: O'(\tau')$ (posizione ($\in q$) di O' al suo tempo proprio τ') e immediatamente lo ritrasmette a O , che a sua volta lo riceve in $O(\tau+\Delta\tau)$ (posizione ($\in p$) di O al suo tempo proprio $\tau + \Delta\tau$, con $\Delta\tau > 0$). Quindi se Ω' è il piede su p di Ω , $\Omega - \Omega' = \delta$, vettore della separazione ortogonale tra le due parallele. Sappiamo poi che in questo caso deve aversi

$$(20) \quad Q(O(\tau+\Delta\tau)-\Omega') = Q(\Omega'-O(\tau)) = -Q(\delta).$$

Ma $\Omega' = O(\tau_1)$ per qualche τ_1 , e dunque $O(\tau+\Delta\tau) - \Omega' = a(\tau+\Delta\tau-\tau_1)$ e similmente $\Omega'-O(\tau) = a(\tau_1-\tau)$, per cui la prima delle due equazioni (20) equivale a $(\tau+\Delta\tau-\tau_1)^2 = (\tau_1-\tau)^2$, ovvero a $\tau_1 - \tau = \Delta\tau/2$. D'altra parte sappiamo che $Q(\delta) = Q(\beta)$ quando valgono le (11) e (12), ove β è dato dalla (9); questa ora diventa $\beta = B(a,\eta)Q(a)a - \eta$, quindi $Q(\beta) = Q(\eta) - B^2(a,\eta)Q(a)$. Ma abbiamo appena visto che $B(a,\eta) = 0$ equivale all'eguaglianza dei tempi propri τ e τ' in punti (di p e rispettivamente di q) che sono l'uno il piede dell'altro. Facendo dunque $B(a,\eta) = 0$, otteniamo innanzitutto $Q(\delta) = Q(\eta)$ (il che è ovvio, essendo allora addirittura $\delta = -\eta$), e inoltre che $\Omega \equiv O'(\tau')$ ha lo stesso tempo proprio di Ω' , ossia che $\tau' = \tau_1 = \tau + \Delta\tau/2$. Si ritrova così esattamente quel “criterio di sincronizzazione” di orologi in punti diversi dello spazio che è stato introdotto su base intuitiva all'inizio di questo capitolo.⁷ Sappiamo inoltre, adesso, che avendo sincronizzato con quella procedura la rete rigida di orologi corrispondente al fascio di parallele che sono le loro linee d'universo nello spazio di Minkowski, se sezioniamo detto fascio col 3-piano σ di equazione $\tau = \text{costante}$ (\equiv spazio fisico “sincrono”), σ risulta ortogonale al fascio stesso. Altrimenti detto, abbiamo “separato” lo spazio fisico sincrono (3-piano euclideo) dal tempo. Ciò corrisponde alla classica scomposizione di uno spazio lineare con prodotto interno nella somma diretta di due suoi sottospazi ortogonali. La “rigidità” della rete spaziale di orologi sincronizzati può essere convalidata istante per istante verificando che $Q(\delta) = (\Delta\tau/2)^2$, come si deduce dalla seconda equazione (20),⁸ si mantiene costante lungo ogni coppia di parallele; anche se essa rigidità è di per

⁷ Si prova facilmente che la relazione tra due orologi in S e in s di essere sincronizzati in base a questo criterio è simmetrica e transitiva; quindi che è una relazione di equivalenza, avendo dato per convenzionalmente vera la sua riflessività.

⁸ Infatti da $\tau_1 - \tau = \Delta\tau/2$ e dalla seconda (20) abbiamo $Q(\delta) = -(\Delta\tau/2)^2 Q(a) = (\Delta\tau/2)^2$. Quest'ultima relazione ci assicura che un segnale luminoso emesso da una sorgente su una certa retta d'universo raggiunge i ricevitori sulle rette

sé automaticamente assicurata dal fatto che tutti gli orologi del fascio si muovono del *comune* moto inerziale specificato da a .

Quanto alla coordinatazione dello spazio euclideo (sincrono), si tratta di un problema già discusso nel Cap 1. Se in particolare si vuole un riferimento cartesiano ortonormale, si richiederà alla relativa base $\{e_i\}_{i=1,2,3}$ di soddisfare alle $B(e_i, e_k) = \delta_{ik}$. Ponendo per brevità $\eta_{ik} = e_i - e_k$, ciò si otterrà verificando che $Q(\eta_{ik}) = Q(e_i) + Q(e_k) = 2$, ovvero che $sep(e_i, e_k) = 2^{1/2} \forall (i, k)$. Inoltre per costruzione risulta $B(e_i, e_4) = 0 \forall i$, e $B(e_4, e_4) = Q(e_4) = -1$, avendo ora scritto come e_4 il vettore unitario a del fascio di parallele. Si vede subito, infine, che per ogni coppia di vettori X, Y , l'invariante $B(X, Y)$ ha la struttura canonica $X_1 Y_1 - X_4 Y_4$ nel riferimento così costruito, se X_i e Y_i ($i = 1, \dots, 4$) sono le componenti di X e di Y , cioè se $X = e_i X_i$, $Y = e_i Y_i$ (somme su i).

Per orientare lo spazio sincrono, basterà prescrivere un ordine sugli elementi "spaziali" e_i della base, ad es. e_1, e_2, e_3 . Lo spazio M diventa così (convenzionalmente e) separatamente orientato nei suoi componenti ortogonali (temporale e spaziale) di cui è la somma diretta. Rileviamo anche che, *orientamento a parte*, l'intera operazione di coordinatazione p.ortonormale è stata ancora una volta eseguita facendo esclusivo uso di orologi normali e messaggi luminali, cioè è stata completamente cronometrizzata.

d'universo ad essa parallele e aventi la stessa separazione ortogonale da essa con comune ritardo $\Delta t/2$ (espresso ad es. in secondi) pari a quella separazione ortogonale (espressa in unità $c \times$ secondi).

2.5 INTRODUZIONE ALL'ELETTROMAGNETISMO MAXWELLIANO

In forza della sua stretta connessione con la relatività speciale, ci è parso opportuno illustrare i fondamenti primi dell'Elettromagnetismo (EM) classico già al livello di questo secondo capitolo, con la presente Sez. 2.5. (Per un migliore inquadramento della materia, e se lo riterrà opportuno, il lettore potrà associare al suo studio quello della App. Spec. 2.H.) Ulteriori informazioni sull'EM si troveranno più avanti, soprattutto nelle S.sez. 5.2.4 e 7.1.2. Come è ovvio, su un piano logico una parte significativa di quanto segue nella sezione ha carattere induttivo.

2.5.1) LE EQUAZIONI DI MAXWELL-LORENTZ

L'esperienza mostra che tra i normali punti materiali, univocamente e permanentemente caratterizzati dalla loro massa inerziale di quiete m^0 (strettamente positiva), ve ne sono di similmente caratterizzati da un addizionale numero reale *con segno*, cioè la loro **carica elettrica** $q \neq 0$; a differenza della massa, un *invariante relativistico*. In quanto dotato di carica q , il punto materiale P, secondo l'osservatore *inerziale di quiete istantanea* \mathfrak{s}^0 rispetto a P stesso, è soggetto ad una forza f^0 , o **forza elettrica** di quiete, uguale al prodotto di q per il cosiddetto **campo elettrico** di quiete e^0 :

$$(1) \quad f^0 =: qe^0.$$

Mentre q è qui considerata come un'osservabile "diretta", ed f^0 come una osservabile "indiretta" attraverso i suoi effetti dinamici su P, e^0 è una proprietà del punto-istante dello spazio-tempo occupato da P che impareremo a *calcolare* in termini di altre osservabili dirette.

Naturalmente l'uguaglianza (1) vale in certi sistemi delle unità di misura usate per misurare le forze, le cariche ed i campi elettrici; in un sistema di unità arbitrarie si ha una semplice proporzionalità. La storia delle unità di misura elettromagnetiche è notevolmente articolata, e non ne parleremo. La scelta di un sistema ottimale di unità dipende infatti dalle prevalenti applicazioni della teoria, e ancora oggi oscilla tra almeno un paio di opzioni; ma nonostante la sua grande importanza pratica, essa non ne ha alcuna sul piano dei fondamenti. A titolo di esempio, noi useremo qui due sistemi di unità, quello cosiddetto "gaussiano" per illustrare l'elettromagnetismo (EM) convettivo nel vuoto, e quello cosiddetto internazionale, o "di Giorgi" (1935), per illustrare l'EM nei mezzi materiali, in entrambi i casi "razionalizzati".¹

¹ L'uso di unità razionalizzate è vantaggiosa se si parte dal sistema di equazioni di Maxwell (vedi (11a, 11b) e (2.5.2, 21a, 21b)), perché permette di evitare la comparsa di un 4π nella maggior parte delle relazioni di interesse (a

In forza della (1), la legge dinamica “istantanea” in \mathfrak{s}^0 per P è espressa dalla

$$(2) \quad m^0 a^0 = q e^0 \equiv f^0,$$

dove a^0 è l’accelerazione di P secondo \mathfrak{s}^0 .²

Vogliamo adesso tradurre la (2) in un altro sistema inerziale \mathfrak{s} (che per fissare le idee possiamo pensare come quello di laboratorio), e nel quale P ha coordinate $(x(t), t)$ e velocità $d_t x \equiv v = v(t)$. Utilizziamo a questo fine la (2.D, 4ter) e la (2D, 5bis), cioè:

$$(3) \quad \gamma a = \gamma^{-1} [a^0 + (\gamma^{-1} - 1) a^0 \cdot v v / v^2]$$

e rispettivamente

$$(4) \quad f = \gamma^{-1} f^0 + (1 - \gamma^{-1}) f^0 \cdot v v / v^2$$

dove $\gamma = \gamma(v^2)$ è il solito coefficiente $(1 - v^2/c^2)^{-1/2}$, mentre $a = a(t)$ sono l’accelerazione di, e rispettivamente la forza agente su, P, sempre secondo \mathfrak{s} . Si noti che sia la (3) che la (4) si riducono ad identità ponendovi $v = 0$ e a^0, f^0 in luogo di a e rispettivamente di f (infatti, allora $\mathfrak{s} \equiv \mathfrak{s}^0$ e $\gamma = 1$).

Analogamente a quanto abbiamo già rilevato nella discussione della legge dinamica newtoniana, anche in questo caso non è f ad essere direttamente osservabile/misurabile, ma piuttosto i suoi effetti attraverso le (3,4). Si tratta dunque di descrivere il presente modello \mathcal{M} dell’accelerazione di P (cfr. la discussione nella Sez. 2.1) con il quale si esprime f come funzione di $x(t)$, di $v(t)$ e di t .

Le equazioni di Maxwell-Lorentz (11) riportate più avanti sono invarianti in forma rispetto a trasformazioni di Lorentz se i vettori del campo si trasformano secondo le successive (12a ÷ 12d). In particolare secondo la (12a) deve essere

$$(5) \quad e^0 = \gamma(e + v \times h/c) + (1 - \gamma)e \cdot v v / v^2$$

(quindi $e^0 \cdot v = e \cdot v$) dove e e h sono il campo elettrico, e rispettivamente il campo magnetico, secondo \mathfrak{s} , lungo la traiettoria di P, quindi in $(x(t), t)$. Si noti che anche la (5) dà luogo ad una identità ponendovi $v = 0$ e e^0 in luogo di e . Le funzioni $e = e(x, t)$ e $h = h(x, t)$ nell’aperto di spazio-tempo di interesse sono per il momento da pensare come *date*.

Combinando le (1,4,5) si trova con alcuni passaggi

$$(6) \quad f = q(e + v \times h/c).$$

Questa forza, che è detta **forza di Lorentz**, è come preannunciato funzione di t lungo la traiettoria di P. D’altra parte, combinando le (1,2,4) si trova

cominciare dalle stesse equazioni di Maxwell). Ovviamente questo numero comparirà allora in altre relazioni, ma di interesse minore dal punto di vista prescelto.

² È inteso che il vettore e^0 nella (2) dipende soltanto dal punto-istante ove si trova P, e non da P. Questo significa che se al punto P si sostituisce un punto P’ di massa di quiete $m^{0’}$ generalmente diversa da m^0 e di carica $q’$ generalmente diversa da q , ma coincidente con P, la legge dinamica per quel P’ diventa, denotandone con $a^{0’}$ l’accelerazione, $m^{0’} a^{0’} = q’ e^0$.

$$(7) \quad f = m^0[\gamma^{-1}a^0 + (1-\gamma^{-1})a^0 \cdot vv/v^2],$$

che riproduce la (2) per $v = 0$, $f = f^0$. Calcoliamo adesso $d_t(\gamma v) = \gamma d_t v + v d_t \gamma = \gamma a + \gamma^2 \gamma a \cdot vv/c^2$; quindi, in forza della (3), e con qualche passaggio, $d_t(\gamma v) = \gamma^{-1}a^0 + [\gamma^{-1}(\gamma^{-1}-1) + v^2/c^2]a^0 \cdot vv/v^2 = f/m^0$ (per la (7)). In definitiva,

$$(8) \quad m^0 d_t(\gamma v) = f.$$

La (8) è la legge dinamica relativistica per P; né questo può destare sorpresa, dal momento che tale legge fu a suo tempo (vedi (2.2.2, 6)) indotta alla luce di un risultato della teoria elettromagnetica, cfr. la (2.2.2, 4).

Il moto di P nel riferimento \mathfrak{s} è a questo punto determinato dalla (8) con la f espressa dalla (6), per date condizioni accessorie; o equivalentemente, sempre per le date condizioni accessorie, dal sistema normale del 1° ordine nelle due incognite x , v

$$(9_1) \quad d_t x = v \equiv \Psi_1(x, v, t),$$

$$(9_2) \quad d_t v = q/(\gamma m^0)[e(x, t) + v \times h(x, t)/c] \cdot [\mathbf{1} + \gamma^2 vv/v^2]^{-1} \equiv \Psi_2(x, v, t),$$

dove la matrice la cui inversa figura a 2° membro della (9₂) è non singolare (il suo determinante valendo $1 + \gamma^2 > 0$) ed elementarmente invertibile. Quindi esiste una e una sola soluzione $x = x(t)$, $v = v(t)$ per le date $e(x, t)$, $h(x, t)$ e sotto le date condizioni accessorie.

Dobbiamo ora rimuovere il fatto di avere temporaneamente assunto le funzioni $e(x, t)$ e $h(x, t)$ come date. Il percorso logico che porta alla determinazione di tali funzioni, *sotto certi dati addizionali direttamente osservabili*, non è immediato, e la sua precisa descrizione costituisce il primo e più profondo cambiamento nei fondamenti della fisica avvenuto dai tempi di Newton in qua (1873). Innanzitutto \mathfrak{s} *osserva/misura* un quadrivettore (\equiv vettore a quattro componenti, tre spaziali e una temporale) $(\rho u, \rho c) \equiv \rho(u, c)$, funzione abbastanza regolare di (x, t) in un dominio di queste coordinate che converrà supporre cilindrico, cioè della forma $\Omega \times \Delta t$, con $\Omega \equiv$ una regione spaziale aperta di contorno $\partial\Omega$ convenientemente regolare, e $\Delta t \equiv$ un intervallo temporale. (Anche se non strettamente necessario, converrà inoltre supporre Ω semplicemente connesso.) Nel quadrivettore $\rho(u, c)$, ρ si interpreta come **densità di carica elettrica** (carica elettrica nell'unità di volume di un continuo carico) ed u come **velocità euleriana** di tale continuo, entrambe pensate come funzioni di (x, t) direttamente osservabili/misurabili da \mathfrak{s} .³ Le funzioni indipendenti nel quadrivettore $\rho(u, c)$ sono tuttavia soltanto tre, perché in $\Omega \times \Delta t$ deve essere soddisfatta l'equazione "di continuità" (o "di conservazione")

$$(10) \quad \nabla_x \cdot (\rho u) + \partial_t \rho = 0,$$

³ Non a caso abbiamo usato il simbolo u , e non v , per questa velocità: $v = v(t)$, riferita al punto P, è a tutti gli effetti una velocità lagrangiana, mentre $u = u(x, t)$ è appunto una velocità euleriana (vedi App. Spec. 2.G).

dove ∇_x denota l'operatore **gradiente** nello spazio x . La (10) esprime la conservazione della carica totale, sperimentalmente verificabile in $\Omega \times \Delta t$. Nel modello \mathcal{M} , $\rho u/c$ e ρ devono a questo punto essere usate come “sorgenti” (termini liberi) di un sistema differenzial-parziale lineare simmetrico-iperbolico nei due campi vettoriali $e = e(x,t)$, e $h = h(x,t)$ (dimensionalmente omogenei nella versione (11) di quel sistema riportata più sotto), e che si dicono **campo elettrico** e rispettivamente **campo magnetico** (secondo \mathfrak{s} , e in $(x,t) \in \Omega \times \Delta t$). Questo sistema differenziale è il **sistema di Maxwell-Lorentz** nel vuoto (ma generalmente in presenza di densità di carica), e si scrive come segue (al solito, \times denota il prodotto vettoriale standard destro):

$$(11a) \quad \nabla_x \times e + \partial_t h/c = 0, \quad \nabla_x \cdot h = 0,$$

$$(11b) \quad \nabla_x \times h - \partial_t e/c = \rho u/c, \quad \nabla_x \cdot e = \rho. \quad ^4$$

L'EM da esse governato, per i dati $\rho u/c$, ρ (termini liberi sotto la (10)) e sotto le convenienti condizioni accessorie (iniziale e al contorno, tipicamente lineari e omogenee), si dice “convettivo nel vuoto”; infatti la sede del campo è il vuoto, il mezzo continuo è carico con densità $\rho = \rho(x,t)$ e velocità euleriana $u = u(x,t)$, $(x,t) \in \Omega \times \Delta t$, e la sola densità di corrente è quella di convezione ρu . Le unità usate nelle (11) sono quelle gaussiane. È immediato verificare che le (11b) *implicano* il vincolo (10), che deve quindi essere soddisfatto per usare in esse $\rho u/c$ e ρ come sorgenti. Quanto alle seconde delle (11a) e (11b), esse si riducono entrambe ad una condizione iniziale: su $\nabla_x \cdot h$ in forza della prima (11a) (se ne prenda la divergenza), e rispettivamente su $\nabla_x \cdot e - \rho$ in forza della prima (11b) (ancora, se ne prenda la divergenza tenendo presente la (10)). In definitiva le equazioni scalari *indipendenti* del sistema di Maxwell-Lorentz (11) sono sei, quante le incognite scalari e , h .

Le (11) sono valide in un generico riferimento *inerziale* (x,t) , quindi in uno spazio-tempo libero da massa-energia. Questo è a rigore incompatibile con i principi della relatività speciale, perché alla esistenza di un campo EM si associano comunque una densità di energia e di suo flusso (cfr. il teorema di Poynting di cui nella S.sez. 2.5.2). La difficoltà è analoga a quella che, in relatività speciale, obbliga a supporre un punto materiale test di massa abbastanza piccola per non perturbare la metrica “piatta” dello spazio-tempo pseudoeuclideo-lorentziano, e si aggira supponendo il campo EM abbastanza debole; una limitazione estremamente blanda, dal punto di vista quantitativo, nella esperienza comune.

⁴ Le (11) mettono in evidenza una stretta similitudine, o una speciale forma di simmetria, tra campo elettrico e campo magnetico. È chiaro che una simmetria in senso stretto tra e ed h potrebbe esserci soltanto se esistesse una **densità di carica magnetica** v da affiancare alla densità di carica elettrica ρ . La nota risposta è che le cariche magnetiche esistono, ma si presentano sempre a coppie di segno opposto, di valore assoluto tendente all'infinito e a distanza tendente a zero, di modo che il prodotto tra le due quantità resti finito (**dipoli**). In queste condizioni, che sono normalmente accettate nella fisica macroscopica, è naturale che la densità di carica magnetica sia zero. Volendo descrivere la pseudosimmetria esistente tra le (11a) e le (11b) completate con $\rho u/c$ e rispettivamente v nei secondi membri, si osserverà che si passa dalle une alle altre, ad esempio, con gli scambi $e \leftrightarrow h$, $\partial_t \leftrightarrow -\partial_t$ e $\rho \leftrightarrow v$.

Resta da esaminare il problema del comportamento del sistema di Maxwell-Lorentz a fronte di trasformazioni di Lorentz. Affinché il modello sia autoconsistente, dovremo richiedere che le (11) restino formalmente invariate passando dal riferimento inerziale \mathfrak{s} ad un altro riferimento inerziale \mathfrak{S} in moto con velocità $-\Upsilon$ rispetto al riferimento \mathfrak{s} ⁵; il che avviene effettivamente, come si dimostra con calcoli un po' noiosi ma banali, sse i campi (e, h, ρ, u) in \mathfrak{s} e gli omologhi (E, H, R, U) in \mathfrak{S} sono legati gli uni agli altri mediante le seguenti formule di trasformazione:

$$(12a) \quad e = gE + (1-g)E \cdot \Upsilon \Upsilon / \Upsilon^2 + g \Upsilon \times H/c,$$

$$(12b) \quad h = gH + (1-g)H \cdot \Upsilon \Upsilon / \Upsilon^2 - g \Upsilon \times E/c,$$

ove $g = (1 - \Upsilon^2/c^2)^{-1/2}$; oltre alla formula di trasformazione (2.D, 2), che ora si scriverà

$$(12c) \quad u = (1 - U \cdot \Upsilon / c^2)^{-1} \{ U g^{-1} + [(1 - g^{-1}) U \cdot \Upsilon / \Upsilon^2 - 1] \Upsilon \},$$

con $g = (1 - \Upsilon^2/c^2)^{-1/2}$. A queste (12a ÷ 12c) si aggiunge la

$$(12d) \quad \rho (1 - u^2/c^2)^{1/2} = R (1 - U^2/c^2)^{1/2},$$

che esprime l'invarianza della carica contenuta in un piccolo volume in moto, in quanto questo subisce la contrazione di Lorentz. Dall'invarianza in forma delle (11) consegue quella della (10), che diventa dunque $\nabla_X \cdot (UR) + \partial_T R = 0$. (Questo si potrebbe anche verificare in modo diretto usando le trasformazioni di Lorentz tra $(\nabla_X \cdot (1/c)\partial_T)$ e $(\nabla_X \cdot (1/c)\partial_t)$.)

Ci è ormai familiare che le (12a ÷ 12c) si invertono scambiando le variabili maiuscole con le omologhe minuscole e scrivendo $-\Upsilon$ in luogo di Υ . In particolare dalla (12a) si ha

$$(12a') \quad E = ge + (1-g)e \cdot \Upsilon \Upsilon / \Upsilon^2 - g \Upsilon \times h/c.$$

Se in questa si identifica \mathfrak{S} con il sistema di quiete \mathfrak{s}^0 , e si sostituisce $-\Upsilon$ (velocità di \mathfrak{s}^0 rispetto a \mathfrak{s}) con v (velocità di P rispetto a \mathfrak{s}), e quindi g con γ , si ritrova la (5). Si comprende bene che l'invarianza in forma delle (11) sotto le (12) è fondamentale, e conferma che le trasformazioni di Lorentz, cardine della relatività speciale, sono anche una vera e propria chiave di volta della teoria elettromagnetica classica. Usando la solita regola, l'inversa della (12b) risulta poi essere

$$(12b') \quad H = gh + (1-g)h \cdot \Upsilon \Upsilon / \Upsilon^2 + g \Upsilon \times e/c;$$

e ancora identificando \mathfrak{S} con il sistema di quiete \mathfrak{s}^0 e sostituendo $-\Upsilon$ con v , si conclude che in P esiste un campo magnetico di quiete (rispetto a P), cioè $h^0 = \gamma h + (1-\gamma)h \cdot vv/v^2 - \gamma v \times e/c$; ma questo campo magnetico h^0 non esercita alcuna forza su P, perché P è per definizione istantaneamente fermo in \mathfrak{s}^0 . Nei limiti di un modello "aperto" \mathcal{M} come il presente, in cui cioè il campo sorgente $\rho(u/c, 1)$ nelle (11) è trattato come un dato direttamente osservabile, arbitrario sotto il vincolo (10), e per le date condizioni accessorie del sistema (11) che assicurano l'esistenza/unicità delle sue

⁵ Abbiamo scritto $-\Upsilon$ invece di Υ per tener fede alla convenzione di denotare con Υ la velocità di \mathfrak{s} (minuscolo) rispetto a \mathfrak{S} (maiuscolo).

soluzioni, il quadro di base della dinamica relativistica di una carica-massa puntiforme in $\Omega \times \Delta t$ è così completo.⁶

Un corrispondente modello “chiuso” \mathcal{M}^* deve invece tener conto del fatto che il campo sorgente è a sua volta *sostanzialmente influenzato* dalla forza di Lorentz. Alla luce della (6), in questo mezzo continuo materiale-carico deve esistere infatti, in $\Omega \times \Delta t$, una **densità di forza di Lorentz**

$$(13) \quad \kappa = \rho(e + u \times h/c),$$

alla quale si associa la **densità di potenza di Lorentz** $\pi =: \kappa \cdot u = \rho e \cdot u$. È allora ben naturale che la densità di forza (13) determini (a meno di condizioni accessorie) il moto del mezzo in quanto dotato di massa con una certa densità μ , e quindi ancora ρ e u attraverso la legge dinamica. (Il vincolo (10) continua ad essere una conseguenza automatica delle (11b) come nel modello aperto \mathcal{M} .) La situazione si complica allora significativamente, perché accanto a ρ e u entrano in gioco altri campi, e quanto meno quello della densità di massa μ . Il più semplice, e di non grande interesse pratico, dei casi pensabili è quello in cui il mezzo non interagisce meccanicamente con se stesso, e quindi non è sede di un campo di forze interne (“mezzo materiale-carico disgregato”): allora si dovranno aggiungere al modello (11) altre cinque equazioni scalari, perché alle sei (e, h) si aggiungono ora le cinque incognite scalari ρ , u e μ . In approssimazione classica (cioè se il moto è non-relativistico) μ si tratta come un invariante; la conservazione di massa è espressa allora dalla equazione di continuità $\nabla_x \cdot (\mu u) + \partial_t \mu = 0$, mentre l’equazione di moto è $\mu[\nabla_x u^2/2 + (\nabla_x \times u) \times u + \partial_t u] = \kappa$, con κ data dalla (13) (Il contenuto delle [] è una possibile rappresentazione del campo euleriano delle accelerazioni.) Manca dunque una relazione scalare, che è tipicamente un “vincolo di stato” finito che leghi ρ e μ , diciamo della forma $\mathcal{F}(x, t | \rho, \mu) = 0$. Sorvoliamo per il momento sulla generalizzazione relativistica di queste equazioni. Anche nella approssimazione classica, il problema è di pertinenza della **(elettro)magnetodinamica dei mezzi continui in movimento**⁷.

Alcuni commenti a questo quadro sono opportuni. Nel modello aperto \mathcal{M} , da una parte abbiamo un *campo* di cariche (**cariche di campo**, “field charges”), $\rho(u/c, 1)$, che costituisce la sorgente del campo EM (e, h) , e dall’altra la forza di Lorentz esercitata sulla carica-massa puntiforme q (**carica-massa di prova**, o “test”), secondo la (6). Vi è in ciò un’ovvia dissimmetria:

⁶ Per ragioni di coerenza notazionale, in tutta questa sezione si usano le notazioni “minuscole” del campo elettrico e magnetico h visti da \mathfrak{s} , riservando le maiuscole agli stessi campi visti da \mathfrak{S} . La notazione standard è invece quella maiuscola, cioè E, H . Lo stesso varrà più oltre (S.sez. 2.5.2) a proposito dei campi di induzione d e b in mezzi materiali.

⁷ Il caso di un mezzo fluido è quello di maggior interesse pratico. Gli sforzi nel fluido hanno tipicamente natura viscosa. La disciplina fisica (e fisico-matematica) che si occupa di questi problemi è la **Magnetofluidodinamica**, ove il prefisso “elettro” è caduto non tanto per brevità, quanto perché fluidi o mezzi anche modestamente conduttori risultano in condizioni usuali elettricamente “quasi-neutri”, con densità di corrente quasi esclusivamente di conduzione (vedi S.sez. 2.5.2). Evidentemente il sistema delle equazioni risultanti è non-lineare anche nel semplicissimo caso di mezzo disgregato.

perché non tener conto, infatti, della forza (di Lorentz) esercitata dalla carica di prova sulle cariche di campo? Al solito, si rimedia a questa banale incongruenza del modello supponendo q asintoticamente piccola (ed è questo il senso, come si è già osservato, della locuzione “di prova”); nonché, volendone mantenere inalterato il moto, diminuendo nello stesso rapporto anche la sua massa. Evidentemente questa dissimetria manca invece nel modello chiuso \mathcal{M}^* : il mezzo di densità di massa μ , densità di carica ρ e velocità euleriana u è al contempo *attivo* (= sorgente di un’azione EM) e *passivo* (soggetto alla densità di forza EM κ). Possiamo radicalizzare la situazione considerata in \mathcal{M}^* immaginando un numero finito – anche soltanto due – di cariche-masse puntiformi in interazione EM: in questo caso ciascuna di esse produce un campo EM che agisce sull’altra determinandone completamente, sotto le condizioni accessorie, il moto. Nasce tuttavia, in connessione con questo modello chiuso di un numero finito di cariche-masse puntiformi in interazione EM, un’altra incongruenza: perché non tener conto dell’azione che una carica-massa puntiforme, pensata come sorgente di un campo EM, esercita su se stessa in quanto soggetta alla forza di Lorentz associata a quel campo? Questo problema ha fatto a lungo discutere; la sua radice, dal punto di vista in cui ci siamo posti, sta nell’aver considerato cariche-masse puntiformi, cioè singolari, *laddove il modello chiuso di Maxwell-Lorentz ne prevede una distribuzione continua*.⁸

Nell’ottica che abbiamo appena illustrato emerge infine l’idea che il campo EM sia a tutti gli effetti soltanto una sorta di “mediatore matematico” per descrivere come “le cariche-masse influenzano il moto delle cariche-masse”, e sia dunque *destituito di ogni realtà fisica*. Non a caso non abbiamo dato (né avremmo potuto farlo in modo logico) alcun protocollo operativo di osservazione/misura *diretta* del campo EM stesso. Anche se può sembrare a prima vista un po’ eversiva (basta non essersi mai soffermati a considerare la situazione da un punto di vista *radicalmente operativa*⁹), questa idea è corretta finché non si trattino l’energia e il momento EM come direttamente misurabili, e risale, con le dovute precisazioni, a Gauss¹⁰. Si immagini per

⁸ Può essere di interesse esaminare qui sotto quali condizioni (necessarie e sufficienti) il sistema di Maxwell-Lorentz (11) si mantiene formalmente invariato, e quindi mantiene la sua validità, a fronte di una **coniugazione CPT** (inversione del segno della Carica, (cioè $\rho \rightarrow -\rho$), di quello della Parità dell’orientamento (cioè $\times \rightarrow -\times$), e di quello del Tempo (cioè $\partial_t \rightarrow -\partial_t$). Si trova facilmente che tale invarianza sussiste sse, oltre al campo ρ , si invertono anche gli altri tre campi e , h ed u . Lo stesso problema può essere considerato con riferimento a coniugazioni parziali, ad esempio di tipo CP, oppure CT, oppure P, .. e via dicendo. Il quadro completo di queste possibilità è il seguente. CPT \Leftrightarrow “inversione di e , h ed u ” (oltre a quella di ρ , che ovviamente si accompagna sempre a C), come anzidetto; CP \Leftrightarrow “inversione di e ”; CT \Leftrightarrow “inversione di e ed u ”; PT \Leftrightarrow “inversione di u ”; C \Leftrightarrow “inversione di e e di h ”; P \Leftrightarrow “inversione di h ”; T \Leftrightarrow “inversione di u e di h ”. L’inversione di u , che sempre si accompagna ad una coniugazione di tipo T, è ben giustificata alla luce del significato fisico di u stessa. Infine si verifica che la forza di Lorentz $q(e + v \times h/c)$ resta invariata alle condizioni sopra specificate, beninteso se v si trasforma allo stesso modo di u .

⁹ Sarà necessario, in particolare, riflettere su cosa si debba intendere per “grandezze direttamente misurabili”. Emergerà allora che tali grandezze sono quelle sommabili, e quindi moltiplicabili per un intero, per sovrapposizione o contiguità, e soltanto esse.

¹⁰ Beninteso, Gauss si riferiva al campo gravitazionale e non a quello elettromagnetico (egli morì infatti ben diciotto anni prima della pubblicazione della teoria di Maxwell (1873)).

un momento, infatti, che in luogo del sistema differenziale parziale di Maxwell ed associate condizioni accessorie ci sia una relazione invertibile “in modo banale” rispetto al campo EM (tale cioè da poter esprimere quest’ultimo immediatamente e senza difficoltà come funzione lineare delle sorgenti): basterà allora trascrivere questa funzione nella (13) ed il campo EM non sarà nemmeno percepito come ingrediente della teoria. È chiaro che l’essere in realtà il campo EM un complicato funzionale integrale (lineare) delle sorgenti non cambia la sostanza logica del discorso. Del resto anche Sommerfeld (Arnold, Königsberg 1868, Monaco di Baviera 1951) sostenne che il campo EM non è che una mera «Rechnengrösse» (per così dirla); e più o meno la stessa tesi è stata ripresa in tempi più recenti da A. Fokker, da J. Wheeler, da R. Feynman, e da altri. Infine si può affermare che *in linea di principio* un problema analogo investe *ogni teoria di campo*.

2.5.2) LA TEORIA EM NEL LINGUAGGIO DEL CALCOLO TENSORIALE

Così come è stata descritta nella sottosezione precedente, la teoria EM convettiva (sia nel modello aperto \mathcal{M} che in quello chiuso \mathcal{M}^*) si fonda sulla legge assiomatica (2.5.1, 1) e sulla validità (del pari assunta assiomaticamente) delle equazioni di Maxwell-Lorentz (2.5.1, 11) ed associate regole di trasformazione (2.5.1, 12). I concetti/strumenti strettamente matematici presupposti da questa impostazione – non contando la teoria dei sistemi differenziale parziali lineari iperbolici – sono abbastanza a portata di mano, limitandosi a: (i) il concetto di campo scalare (come ρ) o vettoriale (come u , e , h) in uno spazio 4-dimensionale pseudoeuclideo con tre coordinate spaziali $x^{1 \leq i \leq 3}$ più una coordinata temporale $x^4 = ct$, e metrica pseudopitagorica $\text{diag}(1,1,1,-1)$; e (ii) il concetto di trasformazione di Lorentz (o più in generale di Poincaré), che lascia invariata la forma quadratica pseudopitagorica nelle $\Delta x^{1 \leq i \leq 4}$ – se le x^i sono cartesiane ortogonali – secondo la $\sum_{i=1}^3 (\Delta x^i)^2 - (\Delta x^4)^2 = \sum_{i=1}^3 (\Delta x'^i)^2 - (\Delta x'^4)^2$ al passaggio dal riferimento inerziale (x) al riferimento inerziale (x'). Nel seguito continueremo ad usare indici latini variabili su 1, ..., 4 e indici greci variabili su 1, ..., 3.

Gli strumenti appena nominati sono dunque sufficienti alla assiomatizzazione dell’EM convettivo nel vuoto e ai suoi sviluppi come teoria fisico-matematica standard; ma ciò non esclude che altri strumenti, e in particolare un’algebra meno elementare di quella dei campi scalari e vettoriali, ne permettano una presentazione più compatta e più aderente alla sua struttura profonda. In effetti, le moderne trattazioni dell’EM – sia nel vuoto che in mezzi materiali – fanno largo uso del linguaggio tensoriale. Dei campi tensoriali su generiche varietà metriche, di cui i campi scalari/vettoriali su spazi euclidei e pseudoeuclidei sono l’accezione più semplice, ci occuperemo

diffusamente e sistematicamente a partire dal prossimo capitolo. D'altra parte, se ci si limita a spazi pseudoeuclidei (o comunque a spazi in cui esiste una metrica indefinitamente *costante*), la "derivazione" di campi tensoriali coincide con quella ordinaria; mentre l'algebra tensoriale si riduce a poche nozioni elementari che diamo per note al lettore.

Come vedremo in un momento, lo strumento formale inventato da Ricci (o Ricci-Curbastro, Gregorio, Lugo It. 1853, Bologna 1925) e Levi-Civita (Tullio, Padova 1873, Roma 1941) permette effettivamente una presentazione della teoria EM assai più sintetica ed elegante che non rinunciandovi, e forse ne rivela meglio alcuni aspetti. Ma esso non è indispensabile; e di fatto, se applicato all'elettromagnetismo coniugato alla relatività speciale (come si è visto, un'unione irrinunciabile), sembra scadere al ruolo di una specie di stenografia. Si ha peraltro la sensazione che dietro questo ruolo inessenziale si nasconda qualcosa di più profondo; il che è vero, ma si rivela soltanto nell'EM nello spazio-tempo pseudoriemanniano della relatività generale, "incurvato" dal campo di materia-energia, e quindi dallo stesso campo EM. Ad ogni buon conto, riporteremo qui appresso una sintesi della formulazione tensoriale dell'EM convettivo nel vuoto; come pure, più oltre in questa sottosezione, della formulazione tensoriale dell'EM in mezzi materiali. Resta inteso che il lettore potrà momentaneamente rinunciare al seguito della sottosezione per tornarvi più tardi, quando avrà acquistato sufficiente familiarità con le nozioni generali illustrate nei Capp. 3 e 4.

Nel seguito aboliremo per brevità l'espressione "campo di" laddove ci vorrebbe; ad esempio dicendo "quadrivelocità" invece di "campo di quadrivelocità", e simili. Definiamo anzitutto la **densità di quadricorrente convettiva** $j^i =: \rho(u^i, c)$, dove $\rho = \rho(x, t)$ è la densità di carica del mezzo e $u^i = u^i(x, t)$ ne è la velocità euleriana. Definiamo poi l'operatore "naturalmente" covariante $\partial_i =: (\partial_i, (1/c)\partial_t)$, dove ∂_i è la componente $(\cdot)_i$ di ∇_x . La 4-divergenza ∂_{ij}^i , un invariante scalare perché le j^i sono per definizione controvarianti, vale allora $\partial_{ij}^i = \partial_i(\rho u^i) + \partial_t \rho$.¹¹ Quindi la

$$(1) \quad \partial_{ij}^i = 0$$

equivale al vincolo di compatibilità (2.5.1, 10).

¹¹ Volendo, l'invariante ∂_{ij}^i può scriversi $g^{ik}\partial_{jk}$, con $g^{ik} \equiv g_{ik} = \text{diag}(1, 1, 1, -1) = \delta^{ik}\varepsilon(i)$, $\varepsilon(1, 2, 3) = 1$, $\varepsilon(4) = -1$. Il passaggio alle componenti covarianti j_k rende opposta la 4^a di esse, lasciando invariate le prime tre, per modo che $j_k = \rho(u_k, -c)$. Dunque è ancora $g^{ik}\partial_{jk} = \partial_i(\rho u_i) - (1/c)\partial_t(-c\rho) \equiv \partial_i(\rho u_i) + \partial_t \rho$. (Ci siamo permessi questa precisazione presumendo che il lettore non abbia ancora sufficiente familiarità con il calcolo tensoriale.) Se poi la coordinata römeriana ct è sostituita con ict ($i^2 = -1$), la definizione di j^i va modificata in $j^i =: \rho(u^i, ic)$, per modo che $\partial_{ij}^i = (1/(ic))\partial_t(ic\rho) = \partial_t \rho$ permanga invariata; ma la metrica diventa allora formalmente euclidea, $g^{ik} = \delta^{ik}$, e possiamo operare senza altre preoccupazioni. Benché artificiosa, questa seconda opzione ("passaggio a coordinata römeriana immaginaria") è ormai prevalente in relatività, perché conduce ad un "modus operandi" più semplice e sicuro. Il fatto che in realtà nell'algebra psudoeuclidea sia $g_{44} = g^{44} = -1$ viene così delegato alle nuove definizioni delle quarte componenti. Per meglio spiegarci, siano $a_i =: (a_i, a_4)$ e $b_i =: (b_i, b_4)$ due vettori 4-dim dello spazio pseudoeuclideo-lorentziano: possiamo ad es. costruire l'invariante scalare 4-dim $a^i b_i =: g^{ik} a_i b_k$, con $g^{ik} = \delta^{ik}\varepsilon(i)$ (quindi uguale a $a_i b_i - a_4 b_4$). Giungiamo tuttavia allo stesso risultato se poniamo $a_i =: (a_i, ia_4)$, $b_i =: (b_i, ib_4)$, e usiamo la stessa regola formale $a^i b_i =: g^{ik} a_i b_k$ con la metrica pitagorica $g^{ik} = \delta^{ik}$. Torneremo più oltre sulla questione, che non è così banale come può sembrare a prima vista.

Introdurremo ora un (campo di) **primo 2-tensore elettromagnetico antisimmetrico** ψ_{ik} (in $\Omega \times \Delta t$), quindi a sole sei componenti indipendenti, mediante le seguenti:

$$(2) \quad \psi_{\iota, \iota+1} =: h_{\iota+2}, \quad \psi_{4\iota} =: -e_{\iota},$$

con $\iota = 1, 2, 3$, le prime tre da leggere mod3. Ribadiamo che, con la scelta del segno del tensore metrico pseudopitagorico qui adottata, l'abbassamento o l'innalzamento di un indice greco lascia invariato il valore della componente vettoriale (o tensoriale) che ne è affetta (ad es., $a^{\iota} \equiv a_{\iota}$), mentre quello dell'indice 4 ne inverte il segno ($a^4 \equiv -a_4$). Alla luce di ciò, definiremo le ψ^{ik} , o in generale le componenti controvarianti di un 2-tensore *generico*, come identiche alle sue componenti covarianti salvo quelle sulla quarta riga aut colonna (ma non quella di indici (⁴⁴), invariata), che vengono invertite di segno. Si può allora immediatamente calcolare l'invariante quadratico $\psi_{ik}\psi^{ik}$, che risulta uguale a $2(h^2 - e^2)$. L'invariante lineare, $\psi_i^i = g_{ik}\psi^{ik}$, è invece banalmente nullo per la simmetria del primo fattore e l'antisimmetria del secondo.

Ciò premesso, si può verificare che le

$$(3) \quad \partial_k \psi^{ik} = j^i$$

equivalgono alle (2.5.1, 11b). Le (1) si possono allora considerare una immediata conseguenza di tali (3), dell'antisimmetria di ψ^{ik} e della simmetria delle derivate seconde ∂^2_{ik} .

Dobbiamo ora riprodurre le (2.5.1, 11a). Allo scopo si introduce un **secondo 2-tensore elettromagnetico antisimmetrico** χ_{ik} mediante le

$$(4) \quad \chi_{\iota, \iota+1} =: e_{\iota+2}, \quad \chi_{4\iota} =: h_{\iota},$$

formalmente identiche alle (2) (e quindi da leggere anch'esse mod3) se nei 2ⁱ membri di queste ultime si sostituiscono h con e, e -e con h. Similmente si può allora verificare che le

$$(5) \quad \partial_k \chi^{ik} = 0,$$

dove χ^{ik} è definito secondo quanto più sopra descritto in termini di χ_{ik} , equivalgono alle (2.5.1, 11a). Come prevedibile, l'invariante quadratico $\chi_{ik}\chi^{ik}$ risulta opposto a $\psi_{ik}\psi^{ik}$, cioè $\chi_{ik}\chi^{ik} = 2(e^2 - h^2)$, mentre l'invariante lineare $\chi_i^i = g_{ik}\chi^{ik}$ è ancora banalmente nullo. Infine l'invariante quadratico $\psi_{ik}\chi^{ik}$ vale $4e \cdot h$.

Si vede subito che i due 2-tensori elettromagnetici introdotti non sono indipendenti, ma tra loro biunivocamente legati dalla

$$(6) \quad \psi_{ik} = -(1/2)Y_{ikjh}\chi^{jh},$$

o dalla inversa

$$(7) \quad \chi_{ik} = -(1/2)Y_{ikjh}\psi^{jh} \quad ^{12}$$

¹² Che la (7) sia effettivamente l'inversa della (6) si può verificare mediante una sostituzione diretta di $\chi^{ik} = -(1/2)Y^{ikjh}\psi_{jh}$ nella (6). Nel Cap. 3, S.sez. 3.2.2, sarà dimostrato nell'ambito di considerazioni sistematiche che

dove Υ_{ikjh} è il simbolo completamente antisimmetrico unitario di ordine 4, con ${}_{1234}$ come permutazione fondamentale (cioè, Υ_{ikjh} è uguale a +1 [-1] se ${}_{ikjh}$ è una permutazione pari [dispari] di ${}_{1234}$, e altrimenti è uguale a 0). Il 2-tensore $\chi_{(2)}$ si dice **coniugato** o **duale** del 2-tensore $\psi_{(2)}$. (Si tenga tuttavia presente che il coniugato del coniugato non è il tensore di partenza, ma il suo opposto. ¹³)

Sotto trasformazioni di Lorentz parallele, le componenti ψ_{ik} e χ_{ik} si trasformano secondo leggi doppiamente covarianti, cioè come quelle valide per prodotti ordinati di due vettori (ad es. secondo le $\psi'_{ik} = \partial x^j / \partial x'^i \partial x^h / \partial x'^k \psi_{jh}$), nel passaggio dal riferimento lorentziano (x) al riferimento lorentziano (x'). Con un po' di pazienza, si verifica allora che queste leggi covarianti *equivalgono alle leggi di trasformazione* (2.5.1, 12a, 12b).

Introduciamo adesso la **densità di quadriforza di Lorentz** $\kappa^i =: (\kappa^i, \pi/c)$, dove $\kappa^i = \kappa_i$ sono le componenti spaziali di $\kappa = \rho(e + u \times h/c)$ (densità di forza di Lorentz, cfr. (2.5.1, 13)) e $\pi = c\kappa^4$ è la relativa **densità di potenza di Lorentz** $\kappa \cdot u = \rho e \cdot u$. Si verifica allora che

$$(8) \quad \kappa_i = \psi_{ikj}{}^k,$$

essendo $\kappa_i = (\kappa_i, \kappa_4 = -\pi/c)$. Con questo *tutte* le equazioni dell'elettromagnetismo convettivo nel vuoto (ma in presenza di densità di carica) ricevono una formulazione tensoriale: le (3) sotto il vincolo di compatibilità (1), le (5) e le (8).

Pur offrendone una versione molto più compatta, quanto precede è soltanto una modalità alternativa di presentare la teoria EM illustrata nella precedente sottosezione. Non può tuttavia sfuggire allo studioso che la presente versione tensoriale (1,3,5,8) dell'EM convettivo sembra andare più direttamente al cuore della materia, ed è presumibilmente da privilegiare se ci si pone dal punto di vista dei fondamenti. Questa impressione si rafforzerà con gli sviluppi che esporremo appresso.

Come nella meccanica classica dei continui materiali la densità di forza uguaglia la divergenza di un 2-tensore simmetrico 3-dim (il tensore degli sforzi), così anche nel caso dell'elettromagnetismo convettivo la densità di quadriforza di Lorentz κ_i può rappresentarsi come la divergenza di un 2-tensore simmetrico E_{ik} 4-dim, il **2-tensore degli sforzi EM** (o **tensore energetico EM**), secondo la

$$(9) \quad \kappa_i = \partial^k E_{ik},$$

$\Upsilon_{iklm} \Upsilon^{jhlm} = 2\delta_{ik}{}^{jh}$, dove a 2° membro figura il cosiddetto “simbolo di Kronecker generalizzato” del 4° ordine. Questa funzione d'indici (in realtà un 4-tensore in forma mista) vale $(-1)^p$ se i due indici in basso sono diversi tra loro e i due indici in alto sono una loro permutazione di parità p, mentre vale 0 in tutti gli altri casi. È allora chiaro che si ha identicamente $\beta_{ik} = (1/2)\delta_{ik}{}^{jh}\beta_{jh}$ per qualsiasi 2-tensore antisimmetrico β_{ik} .

¹³ Così come è stata qui definita, l'operazione di coniugazione corrisponde alla sostituzione di h con e e di e con -h. Quindi il coniugato del coniugato è semplicemente l'opposto.

dove ∂^k sta come sappiamo per ($\partial^k \equiv \partial_k, -(1/c)\partial_t$); o se si preferisce, secondo la

$$(9') \quad \kappa^i = \partial_k E^{ik}.$$

Qui le E^{ik} sono le componenti controvarianti del tensore degli sforzi EM, cioè $E^{1k} = E_{1k}$, $E^{14} = -E_{14}$, $E^{44} = E_{44}$, secondo la regola generale. Naturalmente il sistema (9) (con le κ_i date dalle (8)) ammette infinite soluzioni. Dimostriamo che una di queste soluzioni è

$$(10) \quad E_{ik} =: \psi_{ih}\psi_{ik}^h + g_{ik}\psi_{jh}\psi^{jh}/4,$$

dove è al solito $g_{ik} = \text{diag}(1,1,1,-1)$. Prendendo la ∂^k della (10), abbiamo infatti: $\partial^k E_{ik} = \psi_{ih}\partial_k\psi^{hk} + \partial_k\psi_{ih}\psi^{hk} + \partial_i(\psi_{jh}\psi^{jh})/4$, perché la simultanea dislocazione dell'indice k nei primi due termini a 2° membro li lascia invariati, e $g_{ik}\partial^k = \partial_i$. Ora $\partial_k\psi^{hk} = j^h$ (eq. (3)) e $\psi_{ih}j^h = \kappa_i$ (eq. (8)); quindi

$$(11) \quad \partial^k E_{ik} = \kappa_i + (\partial_k\psi_i^h + \partial_i\psi_k^h + \partial^h\psi_{ki})\psi_h^k/2,$$

perché il 1° e il 3° addendo nella parentesi danno lo stesso contributo. Si verifica senza difficoltà che la somma ciclica tra parentesi nel 2° membro è nulla in forza delle (2.5.1, 11a). Basta infatti porre $h = 1, k = 2, i = 3$ per avere, in forza della definizione (1), $\partial_1\psi_{23} + \partial_2\psi_{31} + \partial_3\psi_{12} = \nabla_x \cdot \mathbf{h} = 0$; e porre $h = 4, k = 2, i = 3$ per avere $\partial_2\psi_{34} + \partial_3\psi_{42} + \partial_4\psi_{23} = \partial_2\mathbf{e}_3 - \partial_3\mathbf{e}_2 + \partial_t\mathbf{h}_1/c = (\nabla_x \times \mathbf{e})_1 + \partial_t\mathbf{h}_1/c = 0$ (e simili). Oltre ad aver provato la tesi (9), abbiamo così al contempo verificato che le equazioni di Maxwell-Lorentz omogenee (2.5.1, 11a) possono scriversi nella forma alternativa $\partial_k\psi_i^h + \partial_i\psi_k^h + \partial^h\psi_{ki} = 0$ per ogni $k, i, h = 1, 2, 3, 4$; o se si preferisce, nella forma

$$(12) \quad \Upsilon^{ihkj}\partial_i\psi_{hk} = 0,$$

dove Υ^{ihkj} è ancora il simbolo completamente antisimmetrico unitario di ordine 4, ora in funzione controvariante (ad esempio, per $j = 4$ il 1° membro della (12) è uguale a $\nabla_x \cdot \mathbf{h}$). Si noti ancora che il tensore degli sforzi EM E_{ik} ha traccia (\equiv invariante lineare $g^{ik}E_{ik}$) uguale a zero: infatti, essendo $g^{ik}g_{ik} = 4$, abbiamo:

$$(13) \quad g^{ik}E_{ik} = \psi_{ih}\psi^{hi} + \psi_{jh}\psi^{jh} = 0.$$

Se calcoliamo le componenti E_{ik} , troviamo (tenendo presente, in particolare, che $\psi_{jh}\psi^{jh} = 2(h^2 - e^2)$):

$$(14_1) \quad E_{44} = -(e^2 + h^2)/2 \equiv -w;$$

$$(14_2) \quad E_{41} = (\mathbf{e} \times \mathbf{h})_1 \equiv r_1/c,$$

$$(14_3) \quad E_{11} = -(e^2 + h^2 - 2e_1^2 - 2h_1^2)/2,$$

$$(14_4) \quad E_{12} = e_1e_2 + h_1h_2,$$

e simili.

Per $i = 4$, la (9) moltiplicata per c dà:

$$(15) \quad -\pi = c\kappa_4 = c\partial^k E_{4k} = c(\partial^k E_{4k} + \partial^4 E_{44}) = \nabla_x \cdot \mathbf{r} + \partial_t w,$$

dove $r = c(e \times h)$ in accordo con la (14₂). La (15) è suscettibile di una naturale *interpretazione energetica*: poiché π è la densità di potenza di Lorentz, cioè il lavoro fatto dal campo EM sulle cariche per unità di volume e di tempo, si può interpretare w come **densità di energia EM** ed r come suo **flusso**. La (15) afferma allora che la variazione per unità di tempo dell'energia EM per unità di volume è uguale all'energia EM *entrante* dal relativo contorno, meno il lavoro fatto dal campo EM sulle cariche (sempre per unità di volume e di tempo). Questo è anche meglio espresso dalla versione integrata su un dominio spaziale fisso ω di contorno regolare $\partial\omega$ della (15), cioè dalla

$$(15') \quad d_t \int_{\omega} w d\omega = \int_{\partial\omega} (-n) \cdot r d\partial\omega - \int_{\omega} \rho u \cdot e d\omega,$$

dove $-n$ è la normale a $\partial\omega$ orientata verso l'interno. Per brevità, d'ora in avanti sottintenderemo il pedice x nell'operatore gradiente ∇_x , perché il rischio di confonderlo con ∇_x è ormai trascurabile. La (15) deve anche potersi derivare *direttamente* (cioè senza far uso esplicito del calcolo tensoriale) dalle equazioni di Maxwell-Lorentz. In effetti basta moltiplicare scalarmente la prima (2.5.1, 11a) per h e la prima (2.5.1, 11b) per e e sottrarre membro a membro; tenendo conto della identità vettoriale

$$(16) \quad \nabla \cdot (v \times w) = w \cdot (\nabla \times v) - v \cdot (\nabla \times w)$$

valida per generici vettori v e w , si ottiene subito la (15).

Il bilancio (15) costituisce il noto **teorema di Poynting**¹⁴ ed ha la tipica struttura di una **equazione di conservazione nel continuo**: vale a dire, se A è la densità (quantità per unità di volume) della grandezza (non necessariamente scalare) che si conserva, F il suo flusso e Q la sua produzione per unità di volume e di tempo, si ha

$$(17) \quad \nabla \cdot F + \partial_t A = Q;$$

o in forma integrata,

$$(17') \quad \int_{\partial\omega} n \cdot F d\partial\omega + d_t \int_{\omega} A d\omega = \int_{\omega} Q d\omega,$$

ove n è sempre orientate verso l'esterno.¹⁵

Per $i = \iota (= 1, \dots, 3)$, in notazione senza indici la (9) risulta essere

$$(18) \quad -\kappa = \nabla \cdot T + \partial_t r / c^2,$$

¹⁴ Poynting (John, Monton Ingh. 1852, Birmingham 1914), noto soprattutto per il suo teorema (15), ma anche per anni di esperimenti intesi a determinare la densità media della Terra, e per la misura della costante gravitazionale. Il vettore r (vedi (14₂)) è universalmente nominato come **vettore di Poynting**.

¹⁵ Come vedremo più diffusamente, equazioni di conservazione nel continuo giocano un ruolo fondamentale in fisica matematica. Ad esempio nella dinamica *classica* dei continui materiali, la conservazione della densità di momento cinetico è espressa dalla $\partial_t G + \nabla \cdot T = f$, dove G è appunto la densità di momento cinetico μu (μ = densità di massa, u = velocità euleriana della materia), T è il 2-tensore simmetrico del flusso di momento cinetico $\mu u u$, e f è la densità di forza attiva sul mezzo, tutte funzioni di (x, t) . La stessa equazione di conservazione della massa [della carica] ha ovviamente questa struttura: denotando con u la velocità euleriana del continuo materiale [del continuo carico] e mettendo in conto l'esistenza di una possibile sorgente di massa M [di carica C] si ha $\partial_t \mu + \nabla \cdot (\mu u) = M$ [$\partial_t \rho + \nabla \cdot (\rho u) = C$].

dove

$$(19) \quad T =: [(e^2 + h^2)\mathbf{1} - 2ee - 2hh]/2 = -E,$$

e $\mathbf{1}$ è la diade unitaria 3-dimensionale (la $T = -E$ è la notazione senza indici della $T_{ik} = -E_{ik}$).

Anche la (18) ha una evidente struttura di equazione di conservazione nel continuo: la grandezza che si conserva è ora la “densità di momento EM” r/c^2 , T (2-tensore simmetrico che si dice **tensore di Maxwell**) è il suo flusso, e κ è la densità di forza di Lorentz, esercitata dal campo EM sul continuo carico. Il parallelo con la (15) è trasparente: r/c^2 gioca il ruolo di w , T quello di r , e κ quello di π . Ancora, come la (15) anche la (18) deve potersi derivare direttamente dalle equazioni di Maxwell-Lorentz. Basta infatti moltiplicare vettorialmente la prima (2.5.1, 11a) per e , la prima (2.5.1, 11b) per h e sommare tenendo conto della identità vettoriale

$$(20) \quad \nabla \cdot (\mathbf{1}a^2/2 - aa) = a \times \nabla \times a - a\nabla \cdot a,$$

valida per un generico vettore a ; eliminando nel risultato le divergenze di e e di h mediante le equazioni scalari (2.5.1, 11) si ritrova la (18).

2.5.3) L'ELETTROMAGNETISMO NEI CONTINUI MATERIALI

Il modello (2.5.1, 11) dell'elettromagnetismo classico nel vuoto è quello di primario interesse in fisica matematica. Tuttavia l'attenzione dei fisici e soprattutto degli ingegneri si rivolge anche in maggior misura all'elettromagnetismo nei continui materiali. In tali continui si deve adottare un modello più generale di quello delle (2.5.1, 11) per tener conto delle “polarizzazioni”, elettrica e magnetica, che vi occorrono sotto l'influenza dei campi elettrico e rispettivamente magnetico.¹⁶ Vale a dire, ai campi elettrico e magnetico si devono associare le loro “induzioni”: l'**induzione elettrica** d e rispettivamente l'**induzione magnetica** b . Il **sistema di Maxwell** nel continuo materiale carico considerato è allora costituito dalle seguenti equazioni differenzialparziali lineari, che scriviamo in unità Giorgi,

$$(1a) \quad \nabla \times e^+ + \partial_t b = 0, \quad \nabla \cdot b = 0,$$

$$(1b) \quad \nabla \times h^+ - \partial_t d = j, \quad \nabla \cdot d = \rho,$$

dove j è una **densità di corrente elettrica**, solitamente dominata dalla sua parte “conduttiva” e legata a ρ dall'equazione di conservazione

$$(2) \quad \nabla \cdot j + \partial_t \rho = 0.$$

¹⁶ Rinviamo a qualunque testo standard di elettromagnetismo per quanto attiene ai modelli fenomenologici della polarizzazione elettrica e magnetica.

Nelle (1) i campi elettrico e magnetico sono stati contrassegnati con un $^+$ in quanto dimensionalmente e numericamente diversi da quelli senza $^+$ che compaiono nelle (2.5.1, 11), pur restando ad essi proporzionali (in altre parole, cambia soltanto la loro unità di misura). Tenuto conto della (2), le equazioni scalari del sistema (1) conseguono dalle corrispondenti equazioni vettoriali a meno di una condizione iniziale, esattamente come le equazioni scalari del sistema (2.5.1, 11) conseguono dalle loro controparti vettoriali.

Alle (1) vanno associate convenienti relazioni fenomenologiche “di stato”, o “costitutive”, che legano i campi alle loro induzioni, quindi del tipo

$$(3_1) \quad d^0 = d^0(e^{+0}|x^0, t^0),$$

$$(3_2) \quad b^0 = b^0(h^{+0}|x^0, t^0).$$

Come indicano i soprascritti (0), queste (3) valgono *nel sistema inerziale di quiete* rispetto al mezzo materiale, e in prima approssimazione sono lineari in e^{+0} e rispettivamente in h^{+0} . Se le (3) sono lineari, e il mezzo materiale è omogeneo-isotropo e in quiete, il sistema (1, 3), con sorgenti j e ρ legate dalla (2), è sostanzialmente simile al sistema (2.5.1, 11). Confrontando la (2) con la (2.5.1, 10), j prende il posto di ρu del modello convettivo; ed è ancora immediato verificare che la (2) stessa è una condizione di compatibilità dei termini liberi j e ρ rispetto alle (1b), come la (2.5.1, 10) lo è rispetto alle (2.5.1, 11b). Il sistema delle (1, 3) consta di dodici equazioni scalari indipendenti nelle altrettante incognite e^+ , d , h^+ , b .¹⁷ Infine alle relazioni fenomenologiche (3₁, 3₂) spesso se ne aggiunge una terza, secondo la quale la densità di corrente j è somma di una parte non-autonoma i che nel riferimento di quiete è data da

$$(3_3) \quad i^0 = i^0(e^{+0}|x^0, t^0),$$

e in prima approssimazione come $i^0 = \sigma e^{+0}$ (σ essendo un fattore di proporzionalità di congrua dimensione che si dice **conducibilità elettrica**) e di una parte autonoma j_{conv} convettiva. In tal caso i veri termini liberi del sistema sono ρ e j_{conv} . Naturalmente la condizione (2), che è conseguenza del sistema (1), resta formalmente invariata, ma coinvolge *direttamente* il campo EM.

Nel vuoto, dove non c'è polarizzazione né conduzione, si ha un vincolo di semplice proporzionalità (con fattore positivo) tra e^+ e d e tra h^+ e b ,¹⁸ esclusivamente connessa all'uso di diverse unità di misura, diciamo $d = \epsilon_0 e^+$ e $b = \mu_0 h^+$, dove ϵ_0 e μ_0 sono costanti positive che si

¹⁷ Così come lo abbiamo presentato, il sistema delle (1, 3) suggerisce una naturale associazione dei quattro vettori e^+ , h^+ , d , b nelle coppie (e^+ , h^+) (“campi”) e (d , b) (“induzioni”). Una scelta più logica è tuttavia quella di associarli nelle coppie (e^+ , b) e (h^+ , d), perché la prima coppia di vettori compare soltanto nelle equazioni *omogenee* (1a) e la seconda soltanto, e in modo analogo, in quelle *non-omogenee* (1b). In conseguenza di ciò, la (3₂) viene preferibilmente presentata nella forma $h^{+0} = h^{+0}(b^0|x^0, t^0)$; e in particolare in un mezzo isotropo-omogeneo, nella forma $h^{+0} = \mu^{-1}b^0$, con μ = permeabilità magnetica del mezzo (vedi più sotto).

¹⁸ Si sono qui soppressi i superscritti 0 in e ed h , perché trattandosi del vuoto non ha senso parlare di “riferimento di quiete” rispetto al mezzo.

dicono **permeabilità elettrica**, e rispettivamente **permeabilità magnetica** del vuoto, ¹⁹ (il pedice (o) significa appunto “del vuoto”). Le (1) vettoriali si riscrivono in questo caso come:

$$(4a) \quad \nabla \times \mathbf{e}^+ + \mu_0 \partial_t \mathbf{h}^+ = 0,$$

$$(4b) \quad \nabla \times \mathbf{h}^+ - \varepsilon_0 \partial_t \mathbf{e}^+ = \mathbf{j}.$$

In forza del modo in cui ε_0 e μ_0 entrano nelle (4), è necessario identificare il prodotto $(\varepsilon_0 \mu_0)^{-1/2}$ con la velocità della luce; facendo i relativi conteggi (vedi la nota (¹⁹)), questa risulta uguale a $2,998 \cdot 10^8 \text{ ms}^{-1}$. ²⁰ Poiché deve essere $\nabla \times \mathbf{h}^+ - \varepsilon_0 \partial_t \mathbf{e}^+ = \mathbf{j} = \rho \mathbf{u} = c \nabla \times \mathbf{h} - \partial_t \mathbf{e}$, è naturale porre $\mathbf{h}^+ = c \mathbf{h} = \mathbf{h}(\varepsilon_0 \mu_0)^{-1/2}$ e $\mathbf{e}^+ = \mathbf{e}/\varepsilon_0$. Con queste posizioni, e scrivendo $\rho \mathbf{u}$ in luogo di \mathbf{j} , le (4) vengono a coincidere con le equazioni vettoriali (2.5.1, 11).

Più in generale, cioè in un generico mezzo materiale in quiete, omogeneo-isotropo e a polarizzazione lineare, avremo relazioni costitutive (3₁, 3₂) del tipo $\mathbf{d}^0 = \varepsilon \mathbf{e}^{+0}$, $\mathbf{b}^0 = \mu \mathbf{h}^{+0}$, dove le permeabilità ε e μ sono costanti positive che con il loro prodotto $\varepsilon \mu$ danno ancora l'inverso del quadrato della velocità della luce nel mezzo, non minore di $1/c^2$. Più in generale ancora, in luogo delle permeabilità scalari si hanno corrispondenti tensori doppi simmetrici (possibilmente funzioni date del posto e del tempo): ad esempio, secondo la $\mathbf{d}_i^0 = \sum_{\kappa=1}^3 \varepsilon_{i\kappa} \mathbf{e}^{+0}_{\kappa}$. Mezzi per i quali vale questa approssimazione, e che per la definizione non sono in generale né isotropi né omogenei, si dicono ancora **a polarizzazione lineare**. ²¹ Lo stesso tipo di generalizzazione si applica infine alla (*) $\mathbf{i}^0 = \sigma \mathbf{e}^{+0}$: la densità di corrente di conduzione \mathbf{i}^0 è allora una trasformata lineare di \mathbf{e}^{+0} attraverso un 2-tensore spaziale simmetrico di conducibilità $\sigma_{i\kappa}$, secondo la $\mathbf{i}_i^0 = \sum_{\kappa=1}^3 \sigma_{i\kappa} \mathbf{e}^{+0}_{\kappa}$. La (*) è universalmente nota come **legge (locale, scalare) di Ohm** (Georg, Erlangen Germ. 1789, Monaco di Baviera 1854).

Buona parte degli sviluppi che abbiamo illustrato a proposito delle (2.5.1, 11) si possono adattare alle (1) senza particolari difficoltà finché le permeabilità del mezzo ε e μ sono supposte costanti e positive (cioè finché il mezzo considerato è supposto (a polarizzazione) lineare-positiva,

¹⁹ In unità Giorgi, risulta $\varepsilon_0 = 8,854 \cdot 10^{-12} \text{ F(farad)m}^{-1}$, e $\mu_0 = 1,257 \cdot 10^{-6} \text{ H(henry)m}^{-1}$, quindi $\varepsilon_0 \mu_0 = 1,113 \cdot 10^{-17} \text{ FHm}^{-2}$. Ma dimensionalmente, i FH sono s² (secondi al quadrato); e infatti è immediato verificare, sulla base delle (4), che $\varepsilon_0 \mu_0$ è dimensionalmente l'inverso del quadrato di una velocità.

²⁰ Infatti prendendo $\nabla \times$ della (4a) [della omogenea associata alla (4b)], usando l'identità operatoriale $\nabla \times \nabla \times = \nabla \nabla \cdot - \nabla^2$, e ricordando che $\nabla \cdot \mathbf{h}^+ = 0$ [che $\nabla \cdot \mathbf{e}^+ = 0$ in assenza di cariche], si ha subito $(\nabla^2 - \varepsilon_0 \mu_0 \partial_t^2) \mathbf{e}^+ = 0$ [($\nabla^2 - \varepsilon_0 \mu_0 \partial_t^2$) $\mathbf{h}^+ = 0$]. Queste equazioni d'alembertiane provano che $\varepsilon_0 \mu_0$ è l'inverso del quadrato della velocità di propagazione di un campo EM nel vuoto, cioè proprio c^{-2} .

²¹ In mezzi cosiddetti “conduttori”, di conducibilità σ e in quiete rispetto al riferimento, e per frequenze non troppo elevate, la parte conduttiva $\sigma \mathbf{e}^+$ della densità di corrente \mathbf{j} domina largamente sul termine $\varepsilon \partial_t \mathbf{e}^+$ nella (4b); infatti la conducibilità di tali mezzi è tipicamente $10^7 \text{ S(siemens)m}^{-1}$. Allora il termine $-\partial_t \mathbf{d}$ nella (1b) può essere ignorato, e quindi $\nabla \cdot \mathbf{j}$, nonché ρ a meno di una condizione iniziale, vengono posti uguali a zero. (Lo stesso risultato si ottiene dicendo che nei conduttori è *formalmente* $\mathbf{d} = 0$.) Questo è quanto avviene sistematicamente, ad esempio, nell'ingegneria dei generatori e dei motori elettrici: le frequenze fondamentali tipiche, in tale fondamentale campo di applicazioni, non superano infatti di norma qualche centinaio di hertz.

omogeneo e isotropo), e non vi è conduzione. Tuttavia questa generalizzazione presenta aspetti non univoci per quanto riguarda le azioni dei campi.

Studiamo innanzitutto la possibilità di tradurre le (1) nel linguaggio tensoriale come abbiamo fatto con le (2.5.1, 11). Introduciamo ora due distinti 2-tensori elettromagnetici antisimmetrici, diciamo β_{ik} e γ_{ik} . Il primo è definito dalle (leggi ι mod3)

$$(5) \quad \beta_{\iota, \iota+1} = b_{\iota+2}, \quad \beta_{4\iota} = -e^+_{\iota}/c$$

che coincidono con le (1) ponendo β in luogo di ψ , b in luogo di h e e^+/c in luogo di e ; e il secondo dalle

$$(6) \quad \gamma_{\iota, \iota+1} = h^+_{\iota+2}, \quad \gamma_{4\iota} = -cd_{\iota},$$

che coincidono con le (5) ponendovi γ in luogo di β , h^+ in luogo di b e cd in luogo di e^+/c . Si verifica facilmente, allora, che le (1a) si possono scrivere come

$$(7) \quad \Upsilon^{hijk} \partial_i \beta_{jk} = 0,$$

e similmente le (1b) come

$$(8) \quad \partial^k \gamma_{ik} = j_i.$$

(Volendo, le (7, 8) si potrebbero riscrivere in termini dei 2-tensori coniugati di β e di γ .) Dalle (8) si traggono poi immediatamente le (2.5.2, 1) cioè $\partial^i j_i = \partial_j^i = 0$. Fin qui il tentativo è dunque riuscito.

Dobbiamo ora generalizzare ai mezzi materiali l'espressione della densità di forza di Lorentz (2.5.1, 13). La legge

$$(9) \quad \kappa = e^+ \rho + j \times b$$

è plausibile e corretta entro gli errori osservativi, e si riduce alla (2.5.1, 13) nel vuoto. Sostituendo infatti ρ a j , $\mu_0 h^+$ a b , ch a h^+ e e/ϵ_0 a e^+ , otteniamo (tenendo presente che stiamo usando unità Giorgi) $\kappa_{\text{Giorgi}} = e\rho/\epsilon_0 + \rho u \times \mu_0 ch = (\rho/\epsilon_0)(e + u \times h/c) = (1/\epsilon_0)\kappa_{\text{Gauss}}$. Cambia dunque, e nel modo giusto, soltanto l'unità di misura. È immediato verificare che la (9) si riproduce ponendo

$$(10_1) \quad \kappa_{\iota} = \beta_{ikj}^k;$$

infatti, ad es. per $\iota=1$, otteniamo $\kappa_1 = \beta_{1kj}^k = \beta_{12j}^2 + \beta_{13j}^3 + \beta_{14j}^4 = b_3j^2 - b_2j^3 + e^+_{1\rho} = (j \times b)_1 + e^+_{1\rho}$. Questo induce a completare la (10₁) con la

$$(10_2) \quad \kappa_4 = \beta_{4kj}^k = -e^+_{kj}/c \equiv -e^+ \cdot j/c.$$

Anche questa si riduce, nel vuoto e a meno dell'unità di misura, alla densità di potenza di Lorentz, secondo l'analogia $\kappa_{4\text{Giorgi}} = (1/\epsilon_0)(-e \cdot \rho u/c) = (1/\epsilon_0)(-\pi/c) = (1/\epsilon_0)\kappa_{4\text{Gauss}}$. Le (10₁, 10₂) si riassumono nella

$$(10_3) \quad \kappa_{\iota} = \beta_{ikj}^k,$$

per $i = 1, \dots, 4$. In conclusione si può supporre che le (10₃) diano la densità di forza-potenza di Lorentz in un mezzo materiale (in unità Giorgi, cioè in N(newton)m⁻³ ²²).

Le difficoltà nascono se cerchiamo di esprimere la $\kappa_{i\text{Giorgi}}$ (che denoteremo ormai semplicemente κ_i) come divergenza di un conveniente 2-tensore energetico 4-dim EM, diciamolo ancora E_{ik} , secondo la $\kappa_i = \partial^k E_{ik}$; cioè, analogamente a come è stato possibile nel vuoto mediante la (9). A prima vista, un candidato naturale al ruolo di E_{ik} sembra essere

$$(11) \quad E_{ik} =: \beta_{ih}\gamma^h_k + g_{ik}\beta_{jh}\gamma^{jh}/4,$$

cfr. le (2.5.2, 10); ma di fatto soltanto il secondo termine di tale **2-tensore di Minkowski** è simmetrico, e questo sembra poco accettabile in vista delle richieste della teoria della relatività generale. Inoltre il tensore di Minkowski non soddisfa la (2.5.2, 9), ma piuttosto la $\partial^k E_{ik} = \kappa_i + \eta_i$, ove

$$(12) \quad 4\eta_i =: \partial_i\gamma_{jh}\beta^{jh} - \partial_i\beta_{jh}\gamma^{jh}$$

(ovviamente le componenti controvarianti dei 2-tensori antisimmetrici β e γ si definiscono con la solita regola in termini delle relative componenti covarianti). Si è allora ripiegato sul simmetrizzato del 2-tensore di Minkowski, $E_{(ik)}$; tuttavia nemmeno questo soddisfa la (2.5.2, 9), ma piuttosto la $\partial^k E_{(ik)} = \kappa_i + \eta_i + v_i$, ove

$$(13) \quad 2v_i =: \partial_h(\beta^{jh}\gamma_{ij} - \gamma^{jh}\beta_{ij}).$$

Come il tensore energetico EM nel vuoto, sia il 2-tensore di Minkowski che il suo simmetrizzato hanno invariante lineare nullo, secondo la $g^{ik}E_{ik} = g^{ik}E_{(ik)} = \beta_{ik}\gamma^{ki} + \beta_{ik}\gamma^{ik} = 0$; ma essi falliscono entrambi l'obiettivo di dare la densità di quadriforza (10₃) con la loro divergenza. Questa difficoltà ha dato luogo ad un lungo dibattito durante la prima metà dello scorso secolo. Benché esso non possa giudicarsi di importanza primaria dal punto di vista dei *fondamenti* della fisica macroscopica, nel suo sviluppo si giunse a mettere in dubbio la correttezza della stessa (10₃). Accettando questo ordine di idee, il problema diventa quello di identificare (in termini di campo EM) il 2-tensore energetico *simmetrico* che con la sua divergenza fornisce quella che si ritiene essere la corretta densità di quadriforza di Lorentz nei mezzi materiali. Abraham (Max, Danzica 1875, Monaco di Baviera 1922 ²³) ha proposto un 2-tensore energetico simmetrico; la cui divergenza soddisfa la (10₃) soltanto a meno di un termine contenente $(1 - c^2\varepsilon\mu)$ (un numero molto

²² In unità gaussiane una forza ha dimensioni Q^2L^{-2} (dove $Q \equiv$ carica); si pensi ad es. alla legge di Coulomb che dà la forza F tra due cariche q_1 e q_2 a distanza r , $F = q_1q_2/r^2$, cioè $F_{\text{Gauss}} =_{\text{dim}} Q^2L^{-2}$. In unità Giorgi, una forza F ha dimensioni MLT^{-2} , cioè $F_{\text{Giorgi}} =_{\text{dim}} MLT^{-2}$, mentre una permeabilità elettrica ha dimensioni $Q^2M^{-1}L^{-3}T^2$. Secondo le formule del testo, $F_{\text{Giorgi}} =_{\text{dim}} (1/\varepsilon)F_{\text{Gauss}}$, dove ε è una permeabilità elettrica, quindi $F_{\text{Giorgi}} =_{\text{dim}} Q^{-2}ML^3T^{-2}Q^2L^{-2} = MLT^{-2}$, come deve essere.

²³ M. Abraham fu noto soprattutto come didatta: dopo un tirocinio a Berlino con M. Planck, insegnò a Göttingen, a Milano e a Stoccarda. Ne accenniamo qui perché egli fu autore, in collaborazione con A. Föppl, di un trattato di EM classico sul quale hanno studiato generazioni di fisici ("Theorie der Elektrizität", 1904, aggiornato da R. Becker a partire dal 1930, e tradotto in molte lingue tra cui l'italiano).

piccolo) come fattore. Poco più tardi si è peraltro fatta luce una diversa idea: infatti la relatività generale impone la simmetria del tensore energetico *totale*, per cui il carattere non simmetrico del tensore di Minkowski può essere accettato con la riserva di compensarlo mediante il tensore energetico relativo ad altro subsistema facente parte del sistema “chiuso” globale di interesse (dopotutto non sembra legittimo considerare come sistema “chiuso” un campo EM in un mezzo materiale).²⁴ Va anche detto che la misura precisa della densità di quadriforza di Lorentz in un mezzo materiale è tecnicamente assai delicata, e le alternative in gioco differiscono per quantità relative molto piccole.

Tornando ai tensori di Minkowski e al suo simmetrizzato, diventa a questo punto legittimo chiedersi quale sia il loro interesse nella teoria EM nei mezzi materiali. La risposta è semplice, e sta nel fatto che, in un mezzo materiale omogeneo-isotropo *in quiete rispetto al riferimento*, si ha $\eta_i = 0$ e $v_i = 0$, e quindi $\partial^k E_{ik} = \partial^k E_{(ik)} = \kappa_i$. (Naturalmente questo vale anche se il mezzo è in moto rettilineo uniforme rispetto al riferimento, perché allora esso sarà in quiete rispetto ad un altro riferimento, trasformato di Lorentz del precedente.) Per quanto consta all'autore, l'orientamento oggi prevalente sull'annosa questione è quello di accettare la (10₃) e il tensore di Minkowski. Ciò in quanto la mancanza di simmetria del tensore di Minkowski può essere *comunque* corretta aggiungendogli un conveniente tensore antisimmetrico e privo di divergenza (un 2-tensore generico ha *due* divergenze, ma se è antisimmetrico l'una è opposta all'altra). Concludiamo tornando a rimarcare che, dal punto di vista dei fondamenti, l'elettromagnetismo di interesse primario resta quello convettivo nel vuoto, nel quale il tensore energetico EM è “naturalmente” simmetrico e soddisfa la (2.5.2, 9).

Il teorema di Poynting per mezzi materiali si ottiene a partire dalle (1) con procedura analoga a quella usata per ricavare la (2.5.2, 15) a partire dalle (2.5.1, 11); cioè, moltiplicando scalarmente la prima (1a) per h e la prima (1b) per e , sottraendo membro a membro e utilizzando l'identità (2.5.2, 16). Si ricava così:

$$(14) \quad -e^+ \cdot j = e^+ \cdot \partial_t d + h^+ \cdot \partial_t b + \nabla \cdot (e^+ \times h^+).$$

Per j dato, le relazioni costitutive (3) restano sullo sfondo, nel senso che sono necessarie per risolvere il problema nelle dodici incognite e^+ , d , h^+ , b . Se applichiamo la (14) al vuoto, abbiamo al solito $e^+ = e/\epsilon_0$, $h^+ = hc$, $d = \epsilon_0 e^+$, $b = \mu_0 h^+$, $j = \rho u$. Sostituendo, moltiplicando per ϵ_0 e ricordando che $c^2 \epsilon_0 \mu_0 = 1$, si ritrova subito la (2.5.2, 15). Secondo la (14), la somma dei primi due termini è la ∂_t della densità di energia EM, $e^+ \times h^+$ è l'espressione attuale del vettore di Poynting, e $e^+ \cdot j$ è l'energia ceduta al mezzo dal campo EM per unità di tempo e di volume.

²⁴ Il primo lavoro in cui si sostiene questo punto di vista è presumibilmente quello dovuto a Jg. Tamm, Journ. Physics URSS, **1**, 439 (1939).

Le relazioni costitutive entrano invece sostanzialmente nel calcolo della densità di forza di Lorentz κ . In un mezzo omogeneo-isotropo e a polarizzazione lineare, *in quiete rispetto al riferimento*, si ha

$$(15) \quad \kappa = \rho e^+ + \mathbf{j} \times \mathbf{b} = -\nabla \cdot \mathbf{s} - \partial_t(\mathbf{d} \times \mathbf{b}),$$

dove \mathbf{s} è un 2-tensore *spaziale simmetrico* le cui componenti tipiche sono

$$(16a) \quad s_{11} = (\mathbf{e}^+ \cdot \mathbf{d}/2 - e^+_1 d_1) + (\mathbf{h}^+ \cdot \mathbf{b}/2 - h^+_1 b_1),$$

$$(16b) \quad s_{12} = -e^+_1 d_2 - h^+_1 b_2;$$

o se si preferisce, in notazione senza indici che ne mette in evidenza la stretta parentela con il tensore di Maxwell (2.5.2, 19),

$$(17) \quad \mathbf{s} = \epsilon(\mathbf{e}^{+2} \mathbf{1} - 2\mathbf{e}^+ \mathbf{e}^+)/2 + \mu(\mathbf{h}^{+2} \mathbf{1} - 2\mathbf{h}^+ \mathbf{h}^+)/2.$$

Possiamo terminare qui questa rassegna dei fondamenti della teoria EM classica, un capitolo centrale in tutta la fisica matematica dei mezzi continui, vuoto compreso. Come gli abbiamo già segnalato, il lettore troverà più avanti (v. S.sezz. 5.2.4 e 7.1.2) altre importanti informazioni sull'elettromagnetismo classico.

APP. SPEC. CAP. 2

APP. 2.A SPOSTAMENTI RIGIDI E CINEMATICA CLASSICA

Propedeutica ad una illustrazione dei fondamenti della Cinematica Relativa classica del punto è la definizione di **spostamento rigido** di un sistema (discreto o continuo) di punti.¹ Si consideri una terna ortogonale $\{e_i\}$ di *versori*; la componente di e_i rispetto ad una simile terna ortogonale $\{E_k\}$ di versori è $E_k \cdot e_i$, che denoteremo α_{ki} . Quindi $e_i = \alpha_{ki} E_k$ (somma da 1 a 3 sulle coppie di indici ripetuti) e $E_k = \alpha_{ki} e_i$. Il fatto che $\{e_i\}$ e $\{E_k\}$ sono terne ortonormali equivale alla

$$(1) \quad \alpha_{ki} \alpha_{hi} = \delta_{kh} = \alpha_{ik} \alpha_{ih}.$$

Ricordiamo che una matrice α soddisfacente alle (1) si dice **ortonormale**. Poiché in forza della definizione l'inversa di una matrice ortonormale è uguale alla sua trasposta, il determinante della prima (o della seconda) ha valore assoluto 1. Una matrice ortonormale si dice **congruente** se il suo determinante è +1, e **anticongruente** se è -1.² Sia O la comune origine dei versori della terna $\{E_k\}$ e o quella dei versori della terna $\{e_i\}$. Sia P un punto definito da $P - O = E_k X_k$ (cioè, X_k siano le coordinate cartesiane ortogonali di P rispetto alla terna $\{E_k\}$), e sia P' analogamente definito da $P' - o = e_i x_i$. Interpretiamo ora P' come nuova posizione di P: lo "spostamento" $P' - P$ è quindi dato da

$$(2) \quad P' - P = o + e_i x_i - O - E_k X_k.$$

Lo spostamento di un sistema di punti P è detto **rigido** se gli spostamenti di ciascuno di essi sono dati dalla (2) in cui si faccia $x_i = X_i$ (per $i = 1,2,3$), e per qualche $\{e_i\}$ definita come sopra; cioè, se

$$(3) \quad P' - P = (o-O) + X_i E_k (\alpha_{ki} - \delta_{ki})$$

per qualche matrice ortonormale α . Se in particolare $\alpha_{ki} = \delta_{ki}$ (si noti che la matrice identità è automaticamente congruente) allora $P' - P = (o-O)$, e lo spostamento si dice **traslazione**; se $o = O$ lo spostamento si dice invece **rotazione** (attorno al polo O), congruente o anticongruente a seconda

¹ Le considerazioni che seguono attorno allo spostamento rigido sono puramente geometriche, e avrebbero potuto essere anticipate al Cap 1. La loro immediata applicazione è tuttavia alla nozione di moto, e per questa ragione esse sono state integrate in questo capitolo. Aggiungiamo che una parte di quanto segue può essere generalizzato senza problemi ad un numero arbitrario (invece di 3) di dimensioni.

² Con ogni evidenza, una matrice ortonormale congruente non può trasformarsi *per continuità* in una simile matrice anticongruente, e viceversa. Una matrice ortonormale è automaticamente non singolare; ma più in generale, sarebbe già non singolare una matrice che trasforma una base qualsiasi in una base qualsiasi. Ciò segue dal fatto che un sistema lineare omogeneo di n equazioni in n incognite ha la sola soluzione banale sse la matrice del sistema è non singolare.

del segno di $\det \alpha$. Le traslazioni sono ovviamente congruenti. Infine, (ricordiamo che) le rotazioni congruenti si dicono anche **rotazioni proprie** (o semplicemente **rotazioni** se il senso è chiaro).

Si dimostra immediatamente che la distanza (pitagorica) tra coppie arbitrarie di punti appartenenti ad un dato sistema di punti, soggetto ad uno spostamento rigido, non varia a seguito di tale spostamento. Siano infatti P_1, P_2 due punti arbitrari del sistema, ${}_1X_k$ e ${}_2X_k$ le loro coordinate in $\{E\}$, quindi:

$$(4_1) \quad P_2 - P_1 = ({}_2X_k - {}_1X_k)E_k.$$

I corrispondenti punti dopo lo spostamento rigido siano P'_1 e P'_2 , quindi

$$(4_2) \quad P'_2 - P'_1 = P_2 - P_1 + ({}_2X_i - {}_1X_i)(\alpha_{ki} - \delta_{ki})E_k = ({}_2X_i - {}_1X_i)\alpha_{ki}E_k.$$

Dalla (4₁), si ha $(P_2 - P_1)^2 = \sum_i ({}_2X_i - {}_1X_i)^2$, e dalla (4₂), $(P'_2 - P'_1)^2 = ({}_2X_i - {}_1X_i)\alpha_{ki}({}_2X_j - {}_1X_j)\alpha_{hj} E_k \cdot E_h =$
 ((*) essendo l'usuale prodotto interno) $= ({}_2X_i - {}_1X_i)\alpha_{ki}({}_2X_j - {}_1X_j)\alpha_{kj} = \sum_i ({}_2X_i - {}_1X_i)^2$, qed.

Questa proprietà di invarianza delle distanze tra coppie arbitrarie di punti del sistema in oggetto è anche caratteristica: se è rispettata, lo spostamento del sistema è automaticamente rigido. Questo si può dimostrare considerando lo spostamento di una terna di punti non allineati. Siano P_0, P_1, P_2 questi tre punti, e P_0 sia spostato in P'_0 (la scelta di questo spostamento corrisponde a quella di tre parametri liberi); fissato P'_0 , P_1 è spostato in P'_1 , che deve essere sulla sfera di centro P'_0 e raggio P_1P_0 (due parametri liberi); fissato P'_1 , P_2 è spostato in P'_2 , che deve essere sul cerchio intersezione delle due sfere centrate in P'_0 e P'_1 e raggi P_2P_0 e rispettivamente P_2P_1 (un parametro libero). Fissato P'_2 , un quarto punto P_3 non-coplanare ai primi tre si sposta in una posizione unicamente determinata P'_3 se si aggiunge una condizione di congruenza o di anticongruenza; e ogni punto successivo si sposta in una posizione ormai unicamente determinata. I parametri liberi dell'intera operazione sono sei, come era da attendersi. Se si pensa a P_0 come origine di una terna $\{E\}$, e si sceglie $P_1 - P_0$ come E_1 e $P_2 - P_0$ come E_2 , si può assumere $e_1 =: P'_1 - P'_0$ e $e_2 =: P'_2 - P'_0$; e_3 è allora completamente determinato dalla congruenza o anticongruenza, e l'asserto è dimostrato avendo posto $\alpha_{ki} = e_i \cdot E_k$, ove α è automaticamente ortonormale.

Le usuali trattazioni della Cinematica Relativa (classica) del punto introducono all'espressione dell'**atto di moto**³ **rotatorio** (cioè con un punto fisso) di un sistema rigido partendo da quella di un suo *piccolo* spostamento (rotatorio), passando poi al rapporto tra questo spostamento e la (piccola) variazione Δt del suo parametro t , e trascurando, come è lecito, gli infinitesimi di ordine superiore al primo nel calcolo del limite di questo rapporto per $\Delta t \rightarrow 0$. Se P è il generico

³ L'atto di moto di un sistema di punti è l'insieme delle velocità di quei punti, ciascuna etichettata dal punto al quale si riferisce.

punto del sistema rigido in moto rotatorio attorno a O (fisso), e $d_t P \equiv d_t(P-O)$ la sua velocità, l'atto di moto del sistema è espresso dalla

$$(5) \quad d_t P = \omega \times (P-O),$$

dove ω è la **velocità angolare** o **spin** del sistema stesso, e \times denota il prodotto vettoriale standard.

Ponendo nella (5) i tre elementi di una base $\{e_i\}$ solidale al sistema in luogo di $P - O$, si hanno le **formule di Poisson**

$$(6) \quad d_t e_i = \omega \times e_i,$$

per $i = 1, 2, 3$. Supponendo la base $\{e_i\}$ ortonormale destra (cioè $e_i \times e_{i+1} = e_{i+2}$, ove i pedici sono da leggere mod3), le (2) si invertono rispetto a ω secondo la

$$(7) \quad \omega = \sum_i (d_t e_i \cdot e_{i+1}) e_{i+2},$$

dove i va da 1 a 3. (Si è sottinteso che le funzioni $e_i(t)$ siano di CdC 1 \equiv di classe C^1 .)

Evidentemente, la (6) e la (7) sono equivalenti, l'una implicando l'altra.

In questo ordine di idee, la rappresentazione di uno spostamento rotatorio del sistema è sostanzialmente affidata all'intuizione geometrico-sintetica.⁴ Vi è tuttavia un modo diverso, più soddisfacente dal punto di vista concettuale, di pervenire alle formule di Poisson. Con la $\{e_i\}$ definita come sopra (ortonormale destra) si parta dalla identità (per $\{e_i\}$ di classe C^1):

$$(8) \quad d_t e_3 \equiv (d_t e_3 \cdot e_i) e_i,$$

In realtà il termine con $i = 3$ manca nella somma a 2° membro perché $d_t e_3 \cdot e_3 = 0$ segue da $e_3 \cdot e_3 = 1$; la somma consta quindi di due soli termini, $(d_t e_3 \cdot e_1) e_1 = (d_t e_3 \cdot e_1) e_2 \times e_3$ e $(d_t e_3 \cdot e_2) e_2 = (d_t e_2 \cdot e_3) e_1 \times e_3$ (l'ultima uguaglianza segue dalla $d_t e_3 \cdot e_2 + d_t e_2 \cdot e_3 = 0$, a sua volta conseguenza della $e_2 \cdot e_3 = 0$).

Quindi

$$(9_3) \quad d_t e_3 = [\sum_i (d_t e_i \cdot e_{i+1}) e_{i+2}] \times e_3 \equiv \omega_{(e)} \times e_3,$$

avendo denotato con $\omega_{(e)}$ la somma tra []; e in generale,

$$(9_i) \quad d_t e_i = \omega_{(e)} \times e_i,$$

per $i = 1, 2, 3$.

Fino a questo punto, $\omega_{(e)}$ è una abbreviazione notazionale con la quale si rappresenta un vettore a priori dipendente dalla base ortonormale $\{e_i\}$ prescelta. Si può tuttavia provare che $\omega_{(e)}$ non varia passando dalla base $\{e_i\}$ ad una generica base $\{e'_i\}$ (ortonormale e congruente) ad essa solidale, cioè tale che $e_i = \beta_{ij} e'_j$, dove β_{ij} sono gli elementi di una matrice β ortonormale congruente

⁴ Il teorema centrale (dimostrato con mezzi di questo tipo) è dovuto a Eulero, e recita: «nello spostamento rotatorio di un sistema rigido, oltre al polo O della rotazione esiste un punto $J \neq O$ che rimane fisso». Per la definizione di rigidità (invarianza di tutte le distanze tra punti), l'intera retta passante per J e O rimane allora fissa. Tradotto in termini algebrici, il teorema di Eulero afferma che una 3×3 -matrice ortonormale congruente diversa dall'identità ha esattamente uno dei suoi tre autovalori uguale a 1. Il corrispondente autovettore, definito a meno di un fattore, è l'asse dello spostamento rotatorio.

e costante rispetto a t . Si ha allora infatti $d_t e_i = \beta_{ij} d_t e'_j = \omega_{(e)} \times \beta_{ih} e'_h$. Moltiplicando quest'ultima per β_{ik} e sommando rispetto a i ,

$$(10) \quad \beta_{ik} \beta_{ij} d_t e'_j = \omega_{(e)} \times \beta_{ik} \beta_{ih} e'_h,$$

ovvero

$$(10') \quad d_t e'_k = \omega_{(e)} \times e'_k.$$

Ma quest'ultima equivale a

$$(11) \quad \omega_{(e)} = \sum_i (d_t e'_i \cdot e'_{i+1}) e'_{i+2} \equiv \omega_{(e')},$$

come preannunciato. Si verifica senza difficoltà che la relazione della base $\{e(t)\}$ alla base $\{e'(t)\}$: “ $e_i(t) = \beta_{ij} e'_j(t)$ per qualche matrice $\{\beta_{ij}\}$ ortonormale/congruente e costante” è una equivalenza; la corrispondente classe di equivalenza può quindi identificarsi con il moto rotatorio in oggetto, che è completamente individuato da un qualsiasi rappresentante della classe, ad es. da $\{e(t)\}$. Il vettore $\omega_{(e)} = \omega_{(e)}(t)$ è univocamente determinato da quel rappresentante $\{e(t)\}$ mediante la (7), ed è invariante nella classe di equivalenza; ovvero è un invariante (di fatto, caratteristico) dell'atto di moto rotatorio considerato. Poiché lo spin dipende soltanto dal moto rotatorio, e non dalla specifica terna ortonormale che lo rappresenta, in esso si abolirà il pedice che denota quella terna (fermo restando che ω dipende in generale da t).

La base ortonormale destra $\{e(t)\}$ si può a sua volta definire mediante la

$$(12_1) \quad e_j(t) =: \alpha_{hj}(t) E_h,$$

dove $\alpha_{hj}(t)$ sono gli elementi di una matrice ortonormale congruente funzione data di classe C^1 di t , ed $\{E_h\}$ è una base ortonormale fissa (destra). La (12₁) si inverte immediatamente nella

$$(12_2) \quad E_h = \alpha_{hj}(t) e_j(t).$$

Dalla $\alpha_{hi} = E_h \cdot e_i$, si ha anche:

$$(13_1) \quad d_t \alpha_{hi} = E_h \cdot d_t e_i,$$

ovvero

$$(13_2) \quad d_t e_i = d_t \alpha_{hi} E_h,$$

come si trae immediatamente anche dalla (12₁). La (12₁) e la (13₂) consentono così di costruire ω a partire della matrice α e della sua derivata $d_t \alpha$, per la data $\{E_h\}$ fissata una volta per tutte. Scrivendo per maggior chiarezza $_{h,i+1}$ in luogo di $_{h,i+1}$, e simili, il risultato è che

$$(14) \quad \omega = \sum_i d_t \alpha_{hi} \alpha_{h,i+1} \alpha_{s,i+2} E_s,$$

ovvero anche, in forza della (11),

$$(14') \quad \omega = \sum_i d_t \alpha_{qk} \mu_{ki} \alpha_{qp} \mu_{p,i+1} \alpha_{sr} \mu_{r,i+2} E_s$$

per qualunque matrice ortonormale congruente μ . Infatti, pensando μ come costante e ponendo $\gamma_{qi} =: \alpha_{qk}\mu_{ki}$, l'espressione a 2° membro della (10') diventa $\sum_i d_t \gamma_{hi} \gamma_{h,i+1} \gamma_{s,i+2} E_s$, cioè lo spin $\omega_{(e')}$ calcolato mediante la terna $\{e'_k =: \mu_{ik}e_i = \gamma_{hk}E_h\}$, solidale alla terna $\{e\}$; ma come abbiamo visto $\omega_{(e')} = \omega_{(e)} = \omega$, da cui la tesi.

Siano ora X_h le coordinate del punto P nella base *fissa* $\{E_h\}$ e x_i quelle nelle base *mobile* $\{e_i(t)\}$. Per la definizione, si ha

$$(15) \quad X_h = \alpha_{hi}x_i;$$

supponendo la matrice α e la colonna x di classe C^1 e derivando,

$$(16) \quad d_t X_h = \alpha_{hi} d_t x_i + d_t \alpha_{hi} x_i;$$

e supponendo α e x di classe C^2 e derivando una seconda volta,

$$(17) \quad d_t^2 X_h = \alpha_{hi} d_t^2 x_i + 2 d_t \alpha_{hi} d_t x_i + d_t^2 \alpha_{hi} x_i.$$

Nella (16), il 1° termine a 2° membro è la componente su E_h della **velocità relativa** v_{rel} (nel riferimento mobile) di $P = X_h E_h$, mentre il 2° termine è la componente su E_h della sua **velocità di trascinamento** V_{tr} . Secondo la (5), deve essere

$$(5') \quad V_{tr} = \omega \times (P-O),$$

ossia $d_t \alpha_{hi} x_i = \omega \times (P-O) \cdot E_h$. La correttezza di questa relazione può essere verificata direttamente.

Se sviluppiamo la (13₁) tenendo conto delle formule di Poisson e della (12₂), otteniamo infatti:

$$d_t \alpha_{hi} = d_t e_i \cdot E_h = (\omega \times e_i) \cdot \alpha_{hj} e_j = \omega_{i+2} \alpha_{h,i+1} - \omega_{i+1} \alpha_{h,i+2}, \text{ quindi: } d_t \alpha_{hi} x_i = \sum_i (\omega_{i+2} \alpha_{h,i+1} - \omega_{i+1} \alpha_{h,i+2}) x_i = \\ = \sum_i (\omega_i x_{i+1} - \omega_{i+1} x_i) \alpha_{h,i+2} = \omega \times (P-O) \cdot \alpha_{hj} e_j = \omega \times (P-O) \cdot E_h = V_{tr} \cdot E_h, \text{ qed.}$$

Analogamente, nel 2° membro della (17) abbiamo la componente su E_h dell'accelerazione relativa a_{rel} (1° termine), poi quella dell'**accelerazione di Coriolis** A_C (2° termine) e infine quella dell'**accelerazione di trascinamento**, A_{tr} (3° termine). Con procedure analoghe si può dimostrare che

$$(18) \quad A_C = 2\omega \times v_{rel}.$$

D'altra parte, alla luce della (5),

$$(19) \quad A_{tr} = d_t \omega \times (P-O) + \omega \times d_t P = d_t \omega \times (P-O) + \omega \times (\omega \times (P-O)).$$

La componente di questa su E_h deve dunque essere uguale a $d_t^2 \alpha_{hi} x_i$; e volendo, anche questo fatto può essere verificato in modo diretto. Insomma, da una parte abbiamo le (15), (16) e (17), e dall'altra le equivalenti

$$(15bis) \quad P = P - O = X_h E_h,$$

$$(16bis) \quad d_t P = d_t X_h E_h = v_{rel} + V_{tr},$$

con V_{tr} data dalla (5'),

$$(17bis) \quad d_t^2 P = d_t^2 X_h E_h = a_{rel} + A_C + A_{tr},$$

con A_C e A_{tr} date dalla (18) e rispettivamente dalla (19). Qui, ricordiamo, la base mobile è in moto rotatorio attorno al polo O fisso.

Poiché le matrici ortogonali congruenti formano un gruppo (generalmente non commutativo) sotto l'operazione prodotto-matriciale, gli **spostamenti omorotatori** (\equiv rotatori con polo fisso comune) formano anch'essi un gruppo, isomorfo al primo. A uno spostamento rotatorio infinitesimo corrisponde una matrice ortonormale del tipo $\mathbf{1} + \varepsilon$, dove $\mathbf{1}$ è la matrice identità e ε è una matrice infinitesima *antisimmetrica*.⁵ Il prodotto di una matrice antisimmetrica con una colonna (di componenti di un vettore polare⁶) corrisponde al prodotto vettoriale di un vettore assiale per quel vettore polare. Nel presente caso, avendolo diviso per il tempo in cui lo spostamento rotatorio infinitesimo occorre, il vettore assiale risultante non è altro che lo spin del sistema. Componendo due siffatti spostamenti omorotatori infinitesimi, diciamo $\mathbf{1} + \varepsilon_1$ e $\mathbf{1} + \varepsilon_2$ (nell'ordine) si ottiene (a meno di infinitesimi del 2° ordine) $(\mathbf{1} + \varepsilon_1) \cdot (\mathbf{1} + \varepsilon_2) = \mathbf{1} + \varepsilon_1 + \varepsilon_2$, e quindi $(\mathbf{1} + \varepsilon_1) \cdot (\mathbf{1} + \varepsilon_2) = (\mathbf{1} + \varepsilon_2) \cdot (\mathbf{1} + \varepsilon_1)$: il gruppo degli spostamenti omorotatori infinitesimi è commutativo. Poiché lo spin di un moto rotatorio è lineare in $d_t \varepsilon$, si comprende che lo spin del moto composto di due moti omorotatori (in un certo ordine) è la somma di uno **spin relativo** e di uno **spin di trascinamento**. Il risultato di questo ragionamento informale (che può comunque essere reso rigoroso) è verificabile in modo diretto come segue. Il primo moto rotatorio sia quello della terna ortonormale destra mobile $\{e\}$ rispetto all'analoga terna fissa $\{E\}$, descritto dalla $e_i = \alpha_{hi} E_h$; e il secondo moto rotatorio sia quello della terna $\{f\}$ rispetto alla terna $\{e\}$, descritto dalla $f_j = \beta_{ij} e_i$ (α e β essendo due matrici ortogonali congruenti funzioni di classe C^2 di t), quindi $f_j = \gamma_{hj} E_h$, con $\gamma_{hj} =: \alpha_{hi} \beta_{ij}$. Lo spin del primo moto (di trascinamento), diciamo ${}^\alpha \omega$, è dato da ${}^\alpha \omega = \sum_i d_t \alpha_{hi} \alpha_{h,i+1} \alpha_{s,i+2} E_s \equiv \sum_i d_t \alpha_{qk} \beta_{ki} \alpha_{qp} \beta_{p,i+1} \alpha_{sr} \beta_{r,i+2} E_s$, ove come è lecito si è posto β , ortonormale congruente, al posto della generica μ (cfr. eq (14')); mentre quello del secondo moto (relativo), diciamo ${}^\beta \omega$, è ${}^\beta \omega = \sum_i d_t \beta_{hi} \beta_{h,i+1} \beta_{s,i+2} e_s$. D'altra parte lo spin del moto risultante è

$$(20) \quad \omega = \sum_i d_t \gamma_{hi} \gamma_{h,i+1} \gamma_{s,i+2} E_s = \sum_i (d_t \alpha_{qi} \beta_{ki} + \alpha_{qk} d_t \beta_{ki}) \alpha_{qp} \beta_{p,i+1} \alpha_{sr} \beta_{r,i+2} E_s.$$

Nell'ultima somma, il primo addendo (in $d_t \alpha$) si riconosce subito come ${}^\alpha \omega$, e il secondo (in $d_t \beta$), tenuto conto della ortogonalità di α e della $\alpha_{sr} E_s = e_r$, come ${}^\beta \omega$; cioè, $\omega = {}^\alpha \omega + {}^\beta \omega$, come si voleva dimostrare. Si noti che ${}^\alpha \omega \cdot E_s$ è espressa in termini di α esattamente come ${}^\beta \omega \cdot e_s$ è espressa in termini

⁵ Infatti la richiesta di ortogonalità su $\mathbf{1} + \varepsilon$ implica $(\mathbf{1} + \varepsilon)^{-1} = \mathbf{1} + {}^t \varepsilon$ (dove t indica trasposizione), e d'altra parte $(\mathbf{1} + \varepsilon)^{-1} = \mathbf{1} - \varepsilon$, da cui la tesi ${}^t \varepsilon = -\varepsilon$.

⁶ Ricordiamo che un vettore polare è un normale vettore "intrinseco", mentre un vettore assiale cambia segno a fronte di una trasformazione anticongruente della base. Il prodotto vettoriale di due vettori polari o assiali è un vettore assiale, mentre quello di un vettore assiale per un vettore polare è un vettore polare. La corrispondenza di una matrice antisimmetrica con un vettore assiale (in tre dimensioni) è una elementare espressione della dualità di Hodge, vedi Sez. 4.4.

di β . In conclusione, un moto rotatorio con spin ω_1 rispetto ad un sistema a sua volta in moto omorotatorio con spin ω_2 rispetto ad un secondo riferimento (che si può pensare come fisso) è un moto omorotatorio, rispetto al riferimento fisso, con spin $\omega = \omega_1 + \omega_2$. Sotto questo tipo di composizione, l'insieme dei moti omorotatori è dunque un gruppo abeliano (**gruppo dei moti omorotatori**), isomorfo al **gruppo** (additivo) **degli spin**.

Resta da considerare il caso di un generico moto rototraslatorio (anziché soltanto rotatorio), della terna mobile. Detta allora o l'origine della terna mobile, basta sostituire, nelle (15 ÷ 17)bis, $P - O$ con $P - o$, quindi $d_t P \equiv d_t(P-O)$ con $d_t P - d_t O$ e $d_t^2 P \equiv d_t^2(P-O)$ con $d_t^2 P - d_t^2 O$. Si ottiene così

$$(15ter) P = X_h E_h + o,$$

$$(16ter) d_t P = v_{rel} + \omega \times (P-o) + d_t o,$$

$$(17ter) d_t^2 P = a_{rel} + 2\omega \times v_{rel} + d_t \omega \times (P-o) + \omega \times (\omega \times (P-o)) + d_t^2 o.$$

Questo modo di procedere è un po' formale, ma la correttezza del risultato si può verificare in modo diretto come segue. Innanzitutto la (15ter) esprime semplicemente la $x_i e_i = X_h E_h$. La velocità relativa è $v_{rel} = d_t x_i e_i$. La velocità di trascinamento è $(^\circ) V_{tr} = \omega \times (P-o) + d_t o$. La (16ter) dice quindi che $d_t P = v_{rel} + V_{tr}$. Derivando quest'ultima, abbiamo $(^+) d_t^2 P = d_t^2 x_i e_i + d_t x_i d_t e_i + d_t V_{tr}$, in cui il 1° addendo è a_{rel} , e il 2° addendo è $\omega \times v_{rel}$ in forza della (6). Derivando la $(^\circ)$, abbiamo $(^*) d_t V_{tr} = \omega \times (d_t P - d_t o) + d_t \omega \times (P-o) + d_t^2 o$. Ma $d_t P - d_t o = v_{rel} + \omega \times (P-o)$; sostituendo questa nella $(^*)$ e sommando i tre addendi a 2° membro della $(^+)$ si ha la (17ter). Le (16ter) e (17ter) sono le espressioni della velocità e della accelerazione assoluta di P (come calcolabili in termini di v_{rel} e a_{rel} dalla terna $\{e\}$ in moto rototraslatorio ($d_t o \equiv V_o$, ω , $d_t V_o$, $d_t \omega$) dato) che si trovano usualmente nei manuali ove si tratta di cinematica relativa del punto. La maggiore semplicità *formale* delle equivalenti espressioni a 2° membro della (16) (con in più V_o), e rispettivamente della (17) (con in più $d_t V_o$), è evidente.⁷

Dobbiamo ancora fornire l'approfondimento di cui all'ultimo paragrafo della Sez. 2.1. Denotiamo con X_h le coordinate di P nella terna ortonormale $\{E\}$ di \mathfrak{S} , ora da pensare come generalmente non inerziale, e quindi mobile con legge arbitraria, e con x_i quelle rispetto alla terna ortonormale $\{e\}$ congruente alla $\{E\}$, in moto traslatorio uniforme rispetto a \mathfrak{S} . Per definizione, è:

$$(21) \quad X_h = \beta_{hi} x_i - b_h - C_h t,$$

ove β è una matrice ortonormale congruente e costante, e b, C , sono vettori costanti. Se scriviamo O [o] per l'origine della terna $\{E\}$ [$\{e\}$], evidentemente si ha $o - O = b + Ct$, $d_t o - d_t O = C$, e $d_t^2 o = d_t^2 O$. D'altra parte (adottando per brevità una notazione senza indici) risulta:

⁷ Le (16), (17) sono immediatamente generalizzabili al caso di uno spazio (euclideo) con un numero n arbitrario di dimensioni, bastando allo scopo che l'indice delle sommatorie sia inteso variare da 1 a n anziché da 1 a 3. Infine esse permetterebbero di proseguire banalmente la rappresentazione delle t-derivate di X_h di ordine superiore al secondo.

$$(22) \quad d_t^2 P = \alpha \cdot d_t^2 X + 2d_t \alpha \cdot d_t X + d_t^2 \alpha \cdot X + d_t^2 O,$$

dove α è la matrice ortonormale del moto di \mathfrak{S} (rispetto a un riferimento assoluto) e \cdot denota il prodotto standard righe \times colonne. Sostituendo la (21) nella (22) otteniamo:

$$(23) \quad d_t^2 P = \alpha \cdot \beta \cdot d_t^2 x + 2d_t \alpha \cdot (\beta \cdot d_t x - C) + d_t^2 \alpha \cdot (\beta \cdot x - (b + Ct)) + d_t^2 o.$$

Questa mostra anzitutto che l'aver tenuto conto di una rotazione β "una tantum" è superfluo, nel senso che ponendo $y =: \beta \cdot x$ la (23) si può riscrivere formalmente come se quella rotazione non ci fosse ($\beta \equiv$ identità), cioè come:

$$(23') \quad d_t^2 P = \alpha \cdot d_t^2 y + 2d_t \alpha \cdot d_t y + d_t^2 \alpha \cdot y + d_t^2 o - 2d_t \alpha \cdot C - d_t^2 \alpha \cdot (b + Ct).$$

Nel 2° membro della (23'), il 1° termine ha la *forma* di un'accelerazione relativa, il secondo di una accelerazione di Coriolis, la somma del 3° e del 4° quella di un'accelerazione di trascinamento (in \mathfrak{s} , con coordinate y); ma come c'era da attendersi, compaiono anche un 5° termine (che tiene conto della variazione della velocità relativa, da $d_t y$ a $d_t y - C$) e un 6° termine (che tiene conto della variazione della coordinata y , da y a $y - (b + Ct)$). Il passaggio da \mathfrak{S} a \mathfrak{s} induce dunque un cambiamento *in forma* dell'espressione dell'accelerazione assoluta, cambiamento che sparisce se $d_t \alpha = d_t^2 \alpha = 0$. Una condizione leggermente più debole (perché non richiede $d_t \alpha = 0$) che assicura l'invarianza *in forma* delle due espressioni è che (i) $d_t^2 \alpha = 0$, quindi che $d_t^2 e_i = 0 \forall i$, e così \mathfrak{S} abbia spin costante, e (ii) che questo spin costante sia parallelo a C , cioè che $d_t \alpha \cdot C = 0$. Banalmente, quelle espressioni sono poi invarianti in forma se il moto di \mathfrak{S} , e quindi anche quello di \mathfrak{s} , è traslatorio uniforme rispetto al riferimento assoluto.

APP. 2.B PROCEDURE DI SINCRONIZZAZIONE

Questa App. 2.B si riferisce alla Sez 2.2 e ne condivide i simboli. Supponiamo che l'orologio in B sia regolato su quello in A \neq B secondo il criterio ($^\circ$) della S.sez. 2.2.1, e quindi che $AB/(t_B - t_A) = c$. In forza dell'asserto ($^\circ$) della S.sez.2.2.1 applicato al segmento AB, risulta anche $2AB/(t'_A - t_A) = c$. Come abbiamo visto, da queste due condizioni segue che

$$(1) \quad t_B - t_A = t'_A - t_B \quad (\Leftrightarrow t_B = (t_A + t'_A)/2).$$

La (1) prova che i tempi di percorrenza del segnale da A a B e da B a A sono uguali, e dunque che

$$(2) \quad AB/(t_B - t_A) = c \Rightarrow AB/(t'_A - t_B) = c.$$

Quest'ultima assicura la proprietà di simmetria della relazione "B è regolato su A". Dovrebbe esser chiaro, tuttavia, che con questi risultati non è possibile dimostrare la transitività della relazione di sincronismo.

A questo scopo, torniamo all'asserto (°) riferendolo ad un triangolo non-degenere di vertici (1), (2), (3), e denotiamo con Δ_i la distanza tra il punto (i+1) e il punto (i) (leggi mod3), e con Δ il perimetro $\Delta_1+\Delta_2+\Delta_3$ del triangolo. Diciamo inoltre per brevità "positivo" il senso di percorrenza del perimetro (1)→(2)→(3)→(1), e "negativo" quello opposto. Siano t_1, t_2, t_3, t'_1 i tempi di partenza da (1), di passaggio per (2) e per (3), e di ritorno in (1), di un segnale che percorre in senso positivo il perimetro del triangolo, e $t^*_1, t^*_3, t^*_2, t'^*_1$ i corrispondenti tempi di un segnale che lo percorre in senso negativo. Supponiamo ora che (2) sia regolato su (1) secondo (°) di S.sez. 2.2.1, cioè che

$$(3) \quad t_2 - t_1 = \Delta_1/c;$$

similmente, supponiamo che (3) sia regolato su (1), cioè che $t^*_3 - t^*_1 = \Delta_3/c$. Per la simmetria, la seconda ipotesi equivale a che (1) sia regolato su (3), ossia che

$$(4) \quad t'_1 - t_3 = \Delta_3/c.$$

Infine in forza dell'asserto (°) deve essere

$$(5) \quad t'_1 - t_1 = \Delta/c.$$

Sommando la (3) e la (4) e tenendo conto della (5) otteniamo $t_2 - t_3 + \Delta/c = (\Delta_1+\Delta_2)/c$, ossia

$$(6) \quad t_3 - t_2 = \Delta_2/c;$$

ma quest'ultima afferma che (3) è regolato su (2) (ovvero (2) è regolato su (3)) secondo il criterio (°), cioè la transitività che volevamo dimostrare.

Operando le varie sincronizzazioni degli orologi nei vertici del triangolo con (°), e quindi scrivendo

$$(7_1) \quad t_2 = \Delta_1/c + t_1, \quad t_3 = \Delta_2/c + t_2, \quad t'_1 = \Delta_3/c + t_3$$

$$(7_2) \quad t^*_3 = \Delta_3/c + t^*_1, \quad t^*_2 = \Delta_2/c + t^*_3, \quad t'^*_1 = \Delta_1/c + t^*_2,$$

si provano immediatamente le uguaglianze

$$(8) \quad t_2+t^*_2 = t_3+t^*_3 = t_1+t'^*_1 = t'_1+t^*_1,$$

sulla cui base si ritrova la simmetria della relazione di sincronismo separatamente per i tre lati del triangolo, e in particolare la simmetria $t^*_3 - t^*_1 = t'_1 - t_3$ più sopra utilizzata. Questo significa tra l'altro che potremmo rinunciare ad utilizzare l'asserto (°) riferendolo ad un segmento e limitarci ad applicarlo ad un triangolo non-degenere. Abbiamo così completamente verificato la coerenza del criterio di sincronismo (°) di S.sez. 2.2.1 alla luce dell'asserto (°).

APP. 2.C COMPLEMENTI SULLE TRASFORMAZIONI DI LORENTZ

Si consideri la coppia $\langle X, U \rangle$ di \mathbb{R}^2 e la sua trasformazione lineare invertibile nella coppia $\langle x, u \rangle$ di \mathbb{R}^2 data da

$$(1) \quad x = aX + bU, \quad u = cX + dU,$$

dove (a, b, c, d) sono gli elementi della matrice di trasformazione A ($A_{11} = a$, $A_{12} = b$, $A_{21} = c$, $A_{22} = d$). A è cioè un automorfismo (lineare) di \mathbb{R}^2 . Come sappiamo, la trasformazione in oggetto (o la sua matrice) si dice ortonormale se lascia invariata la somma dei quadrati delle indeterminate, cioè se

$$(I) \quad X^2 + U^2 = x^2 + u^2,$$

e pseudortonormale se ne lascia invariata la differenza, cioè se

$$(II) \quad X^2 - U^2 = x^2 - u^2.$$

Esaminiamo separatamente i due casi.

Nel caso (I), devono essere soddisfatti i tre vincoli

$$(2) \quad a^2 + c^2 = 1,$$

$$(3) \quad b^2 + d^2 = 1,$$

$$(4) \quad ab + cd = 0;$$

quindi ci si aspetta che i quattro coefficienti (a, b, c, d) contengano un parametro libero. Si dimostra facilmente che vi sono esattamente *due* tipi “irriducibili” di soluzioni (a, b, c, d) del sistema (2 ÷ 4), dati da

$$(5_1) \quad a = d = \cos \varphi, \quad b = -c = \sin \varphi,$$

e rispettivamente da

$$(5_2) \quad a = -d = \cos \varphi, \quad b = c = \sin \varphi,$$

ove $\varphi \in \mathbb{R} \bmod{2\pi}$, secondo quanto era già stato anticipato in Sez. 1.6. La (5_1) , che è una rotazione (propria) di angolo φ , ha determinante = +1; mentre la (5_2) , che è la stessa rotazione seguita o preceduta dall’inversione dell’asse u , o rotazione impropria, ha determinante = -1. Inoltre, per la (5_1) esiste un unico valore di φ , che è $\varphi = 0$, per cui A diventa l’identità $x = X$, $u = U$, mentre per la (5_2) non ne esiste alcuno. La irriducibilità delle due trasformazioni consiste appunto nel fatto che l’una non può farsi coincidere con l’altra variandovi opportunamente il valore del parametro φ (ciò si esprime anche dicendo che le due trasformazioni sono “sconnesse”).

Nel caso (II), devono invece essere soddisfatti i tre vincoli

$$(6) \quad a^2 - c^2 = 1,$$

$$(7) \quad d^2 - b^2 = 1,$$

$$(8) \quad ab - cd = 0,$$

e ancora si prevede che (a,b,c,d) contengano un parametro libero. Si dimostra facilmente che esistono esattamente *quattro* tipi irriducibili di soluzioni delle (6 ÷ 8), date da

$$(9_1) \quad a = d = \text{Ch}\theta, \quad b = c = \text{Sh}\theta,$$

$$(9_2) \quad a = -d = \text{Ch}\theta, \quad -b = c = \text{Sh}\theta,$$

$$(9_3) \quad a = -d = -\text{Ch}\theta, \quad -b = c = -\text{Sh}\theta,$$

$$(9_4) \quad a = d = -\text{Ch}\theta, \quad b = c = -\text{Sh}\theta,$$

ove $\theta \in \mathbb{R}$. Le $(9_1 \div 9_4)$ sono **rotazioni immaginarie** (o **pseudorotazioni**) perché, come sappiamo, sostituendovi θ con $i\theta$, U con iU e u con iu , si riducono formalmente a rotazioni (di tipo (5₁), la (9₁) e la (9₄)) o a rotazioni improprie (di tipo (5₂), la (9₂) e la (9₃)). Inoltre il determinante delle (9₁) e (9₄) vale +1, e quello delle (9₂) e (9₃) vale -1.⁸

Denotando qui con c la celerità della luce per non confonderla con il coefficiente c , identifichiamo U e u con le coordinate temporali cT e rispettivamente ct , e X e x con le coordinate spaziali nella stessa direzione. Le (9₁) e (9₃) si dicono **equicrone**⁹, perché conservano il verso del tempo T o t (essendo comunque $\text{Ch}\theta \geq 1$), e le (9₂) e (9₄) **anticrone**; inoltre, le (9₁) e (9₂) sono (spazialmente) equiverse (perché conservano il verso dell'asse spaziale X o x), e le (9₃) e (9₄) antiverse. Infine soltanto la trasformazione (9₁) può ridursi all'identità, e ciò per il valore (unico) 0 di θ . Sempre dicendo \mathfrak{S} il riferimento con coordinate maiuscole e \mathfrak{s} quello con coordinate minuscole, l'evento E che ha in \mathfrak{S} coordinate $(X = 1, U = 0)$ ha in \mathfrak{s} coordinate $(x = a, u = c)$, e l'evento E' che ha in \mathfrak{S} coordinate $(X = 0, U = 1)$ ha in \mathfrak{s} coordinate $(x = b, u = d)$. È allora chiaro che, possibilmente cambiando il verso dell'asse spaziale di \mathfrak{S} (o di quello di \mathfrak{s}) [il verso di avanzamento degli orologi di \mathfrak{S} (o di quelli di \mathfrak{s})], si potrà sempre fare in modo che E abbia coordinate spaziali [che E' abbia coordinate temporali] dello stesso segno nei due riferimenti. Questo significa che l'imporre $a > 0$ e $d > 0$ non comporta sostanziali limitazioni di generalità; e dunque la sola trasformazione (1) che conserva la *differenza* dei quadrati delle coordinate alla quale possiamo limitare il nostro interesse è la (9₁). Per quanto abbiamo osservato più sopra, la selezione del tipo (9₁) si sarebbe anche ottenuta richiedendo la riducibilità della trasformazione alla identità per $\theta \rightarrow 0$.

La richiesta della linearità e della conservazione delle differenze dei quadrati delle coordinate porta dunque al tipo (9₁), e il solo problema che rimane è quello di esprimere θ in

⁸ Due conseguenze immediate delle (6 ÷ 8) sono $a^2 = d^2 (\geq 1)$ e $b^2 = c^2$. Se gli orologi sono equicroni (vedi nota seguente) e gli assi spaziali sono equiversi allora $a = d$ (quindi anche $b = c$), e le (9₂, 9₃) sono escluse.

⁹ Equicroni [anticroni], riferito a *due* orologi, significa "orientati nello stesso senso" ["orientati in senso opposto"], ed è da distinguere da ortocrono [antiortocrono], riferito ad *un* orologio, che significa "orientato verso il futuro" ["orientato verso il passato"].

termini della velocità Y (valore e segno) di \mathfrak{s} rispetto ad \mathfrak{S} . Questo è banale, essendo chiaramente $\text{Th}\theta = -Y/c$; per cui, tenuto conto delle usuali ed univoche espressioni di $\text{Ch}\theta$ e $\text{Sh}\theta$ in funzione di $\text{Th}\theta$, si ottengono le (2.2.1, 2).

Le trasformazioni di Lorentz speciali (2.2.1, 2) si potrebbero anche ricavare partendo da altre premesse. Ad esempio, si assuma che $x = \kappa(Y)(X - Yt)$ e $X = \kappa'(Y)(x + Yt)$, dove κ e κ' sono funzioni da determinare di Y , alle quali si richiede soltanto, secondo la (3), di essere uguali a 1 per $Y \rightarrow 0$. Se allora si suppone che $x^2 - c^2t^2 = 0 \Rightarrow X^2 - c^2T^2 = 0$ (o viceversa) si trova facilmente che $\kappa(Y) = \kappa'(Y) = (1 - Y^2/c^2)^{-1/2}$, e le (2.2.1, 2) seguono con passaggi banali. L'ultima implicazione postulata è fisicamente significativa, perché afferma che un'onda luminosa emessa dall'origine di \mathfrak{s} al tempo $t = 0$, che si propaga nei due sensi con celerità c secondo \mathfrak{s} , è vista esattamente allo stesso modo da \mathfrak{S} . Alternative come quella appena illustrata ricevono talvolta attenzione nella letteratura didattica.

La generalizzazione delle (1) di presente interesse è

$$(10_1) \quad x_i = B_{ik}X_k + B_{in}U,$$

$$(10_2) \quad u = B_{nk}X_k + B_{nn}U,$$

da presupporre similmente invertibili ($\det\{B_{ik}\}_{i,k=1,\dots,n} \neq 0$), dove gli indici greci variano da 1 a $n-1$, e la loro ripetizione significa somma. A queste corrisponde, nel caso (II), la

$$(IIbis) \quad x_i x_i - u^2 = X_\eta X_\eta - U^2.$$

Deve dunque aversi, identicamente rispetto alle X_i, U ,

$$(11) \quad x_i x_i - u^2 = (B_{ik}X_k + B_{in}U)(B_{i\eta}X_\eta + B_{in}U) - (B_{nk}X_k + B_{nn}U)(B_{n\eta}X_\eta + B_{nn}U) = X_i X_i - U^2,$$

e questo comporta le

$$(12_1) \quad B_{ik}B_{i\eta} - B_{nk}B_{n\eta} = \delta_{k\eta},$$

$$(12_2) \quad B_{in}B_{in} - B_{nn}B_{nn} = -1,$$

$$(12_3) \quad B_{ik}B_{in} - B_{nk}B_{nn} = 0.$$

Il numero dei vincoli (12) tra gli n^2 coefficienti B è $n(n-1)/2 + 1 + (n-1) = n(n+1)/2$, cioè 3 per $n = 2$, 6 per $n = 3$, 10 per $n = 4$, ecc. Questo significa che restano liberi $4 - 3 = 1$ parametri per $n = 2$ (come abbiamo appena visto), $9 - 6 = 3$ parametri per $n = 3$, $16 - 10 = 6$ parametri per $n = 4$, ecc., cioè i numeri che già conosciamo e ci attendevamo. (In particolare per $n = 4$, i 6 parametri sono le tre componenti della velocità relativa tra i due riferimenti e i tre angoli dell'orientamento relativo una tantum.)

I vincoli (12) si possono unificare nella forma compatta

$$(13) \quad B_{ih}B_{ik}\varepsilon(i) = \delta_{hk}\varepsilon(h),$$

dove gli indici latini variano da 1 a n , la loro ripetizione significa somma (ma ovviamente non si somma rispetto a (h) a 2° membro), e $\varepsilon(i = 1, \dots, n-1) = 1$, $\varepsilon(n) = -1$. Matrici che soddisfano le (13) si dicono **lorentziane generalizzate** (generalizzate in quanto n è genericamente $\neq 4$), o anche **pseudortonormali**, e $\varepsilon(h)\delta_{hk}$ **matrice pseudounitaria**. L'analoga generalizzazione del caso (I) è banale e si ottiene dalle (13) ignorandovi i coefficienti ε . Naturalmente i vincoli sono ancora $n(n+1)/2$ e si ha lo stesso numero di parametri liberi, che tuttavia hanno significato diverso da quello del caso (II). Ad esempio per $n = 3$, nel caso (II) i 3 parametri liberi sono le due componenti della velocità relativa e l'angolo dell'orientamento una tantum tra i due riferimenti a 2 dimensioni spaziali, mentre nel caso (I) essi sono i tre angoli dell'orientamento tra i due riferimenti a 3 dimensioni spaziali. Le $(n \times n)$ -matrici che soddisfano le (13) [le (13) con $\varepsilon \equiv 1$, $B_{ih}B_{ik} = \delta_{hk}$] formano un gruppo sotto moltiplicazione (**gruppo lorentziano** (generalizzato)) [**gruppo delle rotazioni** (generalizzate)].

Si noti ancora che, se \mathbf{B} è una qualsiasi $(n \times n)$ -matrice che soddisfa alle (13) in cui le $\varepsilon(i)$ sono numeri diversi da zero e per il resto arbitrari, allora $\det^2\{\mathbf{B}\} = 1$. La dimostrazione è quasi immediata: il 1° membro delle (13) può scriversi ${}^tB_{hi}\delta_{ij}\varepsilon(i)B_{jk}$ (${}^t \equiv$ trasposizione) ed è dunque il prodotto, nell'ordine, delle tre matrici ${}^tB_{hi}$, $\delta_{ij}\varepsilon(i)$ e B_{jk} . Prendendo il determinante degli elementi $({}_{hk})$ nei due membri delle (13) si ha $\det\{{}^t\mathbf{B}\}\prod_{i=1}^n \varepsilon(i)\det\{\mathbf{B}\} = \prod_{i=1}^n \varepsilon(i)$; ma $\det\{{}^t\mathbf{B}\} = \det\{\mathbf{B}\}$, e per ipotesi $\prod_{i=1}^n \varepsilon(i) \neq 0$, da cui scende immediatamente la tesi. Inoltre nelle (13) nella forma ${}^tB_{hi}\delta_{ij}\varepsilon(i)B_{jk} = \delta_{hk}\varepsilon(h)$ le matrici \mathbf{B} e ${}^t\mathbf{B}$ hanno un ruolo simmetrico, per cui si presume che accanto alle (13) debbano valere le simili

$$(13\text{bis}) \quad B_{hi}B_{ki}\varepsilon(i) = \delta_{hk}\varepsilon(h),$$

in cui a 1° membro le somme sono fatte rispetto agli indici di colonna, e non di riga come nelle (13). In realtà le (13) e le (13bis) sono equivalenti: mutatis mutandis, la dimostrazione è strettamente analoga a quella data nell'ultima nota della Sez. 1.2 nel caso delle matrici ortonormali.

Alla rappresentazione (10) delle trasformazioni di Lorentz generalizzate (n arbitrario ≥ 2) – quindi soddisfacenti alla (IIbis) e alle (13,13bis) – se ne può sostituire un'altra equivalente e spesso più comoda. Precisamente, lasciando invariate le coordinate x_t, X_k , si ponga:

$$(14_1) \quad u = v/i, \quad U = V/i,$$

$$(14_2) \quad B_{tk} = C_{tk}, \quad B_{tn} = -C_{tn}/i, \quad B_{nt} = C_{nt}/i, \quad B_{nn} = C_{nn},$$

dove i è l'unità immaginaria. Sostituendo queste posizioni nelle (10₁) abbiamo

$$(15_1) \quad x_t = C_{tk}X_k + C_{tn}V,$$

$$(15_2) \quad v = C_{nk}X_k + C_{nn}V.$$

Le (15) sono formalmente identiche alle (10) (avendo sostituito in queste ultime (u,U) con (v,V) e B con C). D'altra parte, sostituendo (u,U) con (v,V) in (IIbis) abbiamo $x_i x_i + v^2 = X_k X_k + V^2$, e questo ci fa prevedere che le (13, 13bis) diventino

$$(16) \quad C_{ih} C_{ik} = \delta_{hk},$$

e rispettivamente

$$(16bis) \quad C_{hi} C_{ki} = \delta_{hk}.$$

La verifica diretta non presenta difficoltà, ed è lasciata al lettore. La matrice C risulta dunque “formalmente” ortonormale. In conseguenza delle (16) (o delle equivalenti (16bis)) si vede che $\det^2\{C\} = \det^2\{B\} = 1$; non solo, ma le (14) implicano che $\det\{C\} = \det\{B\}$. Secondo le (16), l'inversa della matrice C è la trasposta ${}^t C$, e viceversa l'inversa di ${}^t C$ è C . Ovviamente il passaggio dalle (10) alle equivalenti (15) è comodo perché evita “somme alla Lorentz” nonché modifiche delle regole standard del calcolo tensoriale e della stessa algebra matriciale: se gli spazi originali (x_i, u) e (X_k, U) erano lorentziani, i nuovi spazi delle (x_i, v) e (X_k, V) sono (formalmente) euclidei. Come abbiamo visto, quel passaggio lascia le trasformazioni di Lorentz formalmente immodificate, ma modifica la natura delle matrici di trasformazione: se ad es. B era reale e simmetrica, C non lo è più, ma ha una particolare struttura, in parte reale-simmetrica (due indici greci, o due indici n) e in parte immaginaria-antisimmetrica (un solo indice greco). Inoltre la trasformazione da B a C è unicamente invertibile. Supposta B reale, C potrà dirsi “associata a B per passaggio a n -ma coordinata immaginaria” (e B “associata a C per passaggio a n -ma coordinata reale”). Abbiamo già incontrato questo concetto nel caso standard relativistico ($n = 4$), in cui B è reale e simmetrica.

APP. 2.D LE FORMULE DI TRASFORMAZIONE DI LORENTZ PARALLELA

In questa App. Spec 2.D abbiamo anzitutto elencato alcune importanti formule di trasformazione relative al passaggio dal riferimento inerziale \mathfrak{S} (maiuscolo, come le grandezze ad esso riferite) al riferimento inerziale \mathfrak{s} (minuscolo, come le grandezze ad esso riferite), *parallelo ad* \mathfrak{S} e in moto rispetto ad esso con velocità vettoriale (costante) Υ . Le notazioni (del tipo (minuscolo, maiuscolo)) sono quelle usate nel testo principale.

§1. Trasformazione parallela di Poincaré (Lorentz-affine) delle coordinate, a sette parametri Υ (tre parametri), x_0 (tre parametri) e t_0 (un parametro). Con $g =: (1 - \Upsilon^2/c^2)^{-1/2} \equiv g(|\Upsilon|)$, si ha:

$$(1a) \quad x - x_0 = X + \Upsilon [\Upsilon \cdot X (g-1)/\Upsilon^2 - Tg],$$

$$(1b) \quad t - t_0 = g(T - \Upsilon \cdot X/c^2).$$

Le (1a) sono 3-vettoriali e si possono proiettare sui versori coordinati comuni ai due riferimenti \mathfrak{S} e \mathfrak{s} . Facendo $x_0 = t_0 = 0$, supponendo il vettore Υ parallelo al primo versore coordinato (con componente che continuiamo a scrivere come Υ), e infine proiettando x sui tre versori coordinati comuni supposti mutuamente ortogonali, le (1) si riducono immediatamente alle trasformazioni di Lorentz speciali (2.2.1, 2). In quanto segue, ci riferiremo sempre a trasformazioni parallele delle coordinate. I quattro parametri x_0, t_0 non intervengono nelle trasformazioni successive, che sono tutte a tre parametri Υ . §

§2. Trasformazione delle velocità (di un punto cinematico). Posto al solito $\alpha = \alpha(\Upsilon \cdot V) =: (1 - \Upsilon \cdot V/c^2)^{-1}$, esse sono:

$$(2) \quad v = \alpha \{ g^{-1}V + [(1 - g^{-1})\Upsilon \cdot V/\Upsilon^2 - 1]\Upsilon \}.$$

Se il vettore Υ è parallelo all'asse x (con componente che continuiamo a scrivere Υ), dalla (2) riotteniamo subito la (2.2.1, 4). Della (2) possiamo calcolare il componente parallelo a Υ (diciamo v_{\parallel}) e quello normale a Υ (diciamo v_{\perp}). Poiché $(\)_{\parallel} = (\) \cdot \Upsilon \Upsilon / \Upsilon^2$, e $(\)_{\perp} = (\) - (\)_{\parallel} = (\) \cdot (\mathbf{1} - \Upsilon \Upsilon / \Upsilon^2)$, si trova subito:

$$(2_{\parallel}) \quad v_{\parallel} = \alpha(V_{\parallel} - \Upsilon),$$

e rispettivamente

$$(2_{\perp}) \quad v_{\perp} = \alpha g^{-1}V_{\perp}.$$

Se poi Υ e V sono perpendicolari, la (2) si riduce a $v = g^{-1}V - \Upsilon$. Con ogni evidenza, le (2, 2 $_{\parallel}$, 2 $_{\perp}$) si invertono scambiando tra loro v con V e sostituendo Υ con $-\Upsilon$. Sempre partendo dalla (2), si trova che, posto $\gamma =: (1 - v^2/c^2)^{-1/2}$ e $\Gamma =: (1 - V^2/c^2)^{-1/2}$,

$$(3) \quad \alpha\gamma = g\Gamma,$$

e quindi, per quanto appena detto, anche

$$(3\text{bis}) \quad \alpha^+ \Gamma = g\gamma,$$

dove $\alpha^+ = \alpha^+(\Upsilon \cdot v) =: (1 + \Upsilon \cdot v/c^2)^{-1}$. Combinando le (3, 3bis), si vede che $\alpha\alpha^+ = g^2$, ovvero che $(1 - V \cdot \Upsilon/c^2)(1 + v \cdot \Upsilon/c^2) = 1 - \Upsilon^2/c^2$. Questa è una identità alla luce della (2 $_{\parallel}$) o della sua inversa, e ciò può considerarsi una verifica della (3).

Si vede facilmente che $|V| = c \Leftrightarrow |v| = c$, come deve essere in questo caso limite. Se $|V| = |v| = c$, i due vettori v e V risultano legati dalla $v = \alpha[g^{-1}V \cdot (\mathbf{1} - \Upsilon \Upsilon / \Upsilon^2) - \Upsilon \cdot (\mathbf{1} - V \Upsilon / \Upsilon^2)]$.¹⁰ Si verifica subito che se in questa si pone $V = c\Upsilon/|\Upsilon|$, si trova $v = V$.

La (2) può essere interpretata come la legge che in relatività speciale lega la velocità assoluta V_a (di un punto) alla sua velocità relativa v_r e alla velocità di trascinamento V_t : basta porvi

¹⁰ In effetti, ponendo in questa $V = \Omega c$ con Ω unitario, e facendone il quadrato, si trova $v^2 = c^2$. Se poi nella stessa si pone $V = c\Upsilon/|\Upsilon|$, si ha $v = V$.

$V = V_a$, $v = v_r$ e $Y = V_t$. Nel caso di una trasformazione composta dal passaggio da \mathfrak{S} a \mathfrak{s} (in moto con velocità Y rispetto a \mathfrak{S}) seguito dal passaggio da \mathfrak{s} a \mathfrak{s}' (in moto con velocità Y' rispetto a \mathfrak{s}), per la velocità Y_+ di \mathfrak{s}' rispetto a \mathfrak{S} si trova l'equazione (°) $[g^{-1}\mathbf{I} + (1-g^{-1})YY'/Y^2 + Y'Y/c^2] \cdot Y_+ = Y + Y'$, che è equivalente all'espressione di Y_+ data nel testo, S.sez. 2.2.1. Infatti la matrice a fattore di Y_+ , nella (°), è non singolare (il suo determinante vale $g^{-2}(1+Y \cdot Y'/c^2)$), e la menzionata espressione di Y_+ , si verifica mediante qualche passaggio, soddisfa alla (°). Si noti infine che la (°) si riduce alla solita $Y_+ = (Y + Y')/(1 + YY'/c^2)$ nel caso di trasformazioni di Lorentz speciali. §

§3. Trasformazione delle accelerazioni:

$$(4) \quad a = g^{-1}\alpha^2 \{ [g^{-1}A + (1-g^{-1})YY \cdot A/Y^2] + \alpha \{ g^{-1}V + [(1-g^{-1})Y \cdot V/Y^2 - 1]Y \} Y \cdot A/c^2 \}.$$

Mediante manipolazioni banali, la (4) può essere riscritta nella forma

$$(4bis) \quad a = g^{-2}\alpha^2 A \cdot (\mathbf{I} - YY/Y^2) + g^{-2}\alpha^3 V \cdot (\mathbf{I} - YY/Y^2) Y \cdot A/c^2 + g^{-3}\alpha^3 A \cdot YY/Y^2;$$

se allora il vettore Y è parallelo all'asse x (con componente Y), la (4bis) permette di riottenere quasi immediatamente le formule (2.2.1, 7). Infatti in tal caso la componente (x) del vettore $\pi \cdot (\mathbf{I} - YY/Y^2)$, dove π è un vettore generico, è zero, e si è subito ridotti alla prima delle (2.2.1, 7). Invece per la componente (y) [(z)] dello stesso vettore si ha che la componente (y) [(z)] dell'ultimo addendo della (4bis) è manifestamente zero, per cui si è ridotti alla seconda [alla terza] delle (2.2.1, 7). Se poi $\mathfrak{S} \equiv \mathfrak{S}^0$ è riferimento di quiete ($V = V^0 = 0$), si ha più semplicemente (scrivendo A come A^0),

$$(4ter) \quad a = g^{-2}[A^0 \cdot (\mathbf{I} - YY/Y^2) + g^{-1}A^0 \cdot YY/Y^2].$$

Partendo dalla (4) o dalla (4bis) (o anche dalle (2.2.1, 7)), si controlla immediatamente, infine, che $a = A$ nel limite galileiano $c \rightarrow \infty$. §

§4. Trasformazione delle forze:

$$(5) \quad f = \alpha [g^{-1}F + (1-g^{-1})YY \cdot F/Y^2 - F \cdot VY/c^2];$$

qui V è la velocità (rispetto ad \mathfrak{S}) del punto sul quale si esercita la forza. Nel limite galileiano, $f = F$. Ancora, se $\mathfrak{S} \equiv \mathfrak{S}^0$ è il riferimento di quiete per quel punto, allora $V \equiv V^0 = 0$, $\alpha = \alpha^0 = 1$, e siamo ridotti a (scrivendo F come F^0)

$$(5bis) \quad f = g^{-1}F^0 + (1-g^{-1})YY \cdot F^0/Y^2,$$

mentre $f = F^0$ nel limite galileiano.

Come è anticipato nel testo (vedi (2.2.2, 4)), la formula di trasformazione (5bis) è legata alla possibilità che la forza sia di origine elettromagnetica, e quindi che essa si trasformi secondo la (2.2.2, 4); ma questa tesi va ora giustificata alla luce delle formule di trasformazione (2.5.1, 12a) e (2.5.1, 12b) e della formula di Lorentz (2.5.1, 8). La prova consiste in una semplice verifica. Senza rinunciare alla generalità, possiamo assumere $q = 1$ nella (2.5.1, 8). Facendo in questa, come si

deve, $v = -Y$, abbiamo $f = e - Y \times h/c$; inoltre è evidentemente $F^0 = E$. Orientando gli assi (x) e (X) delle due terne cartesiane parallele come Y , e utilizzando le (2.5.1, 12a,12b), si ha innanzitutto $e_x = E_X$, $f_x = e_x$, $F^0_X = E_X$, quindi $f_x = F^0_X$. Questa prova la correttezza della componente lungo l'asse (x \equiv X) della (5bis). Inoltre $e_y = g(E_Y - |Y|H_Z/c)$, $-(Y \times h/c)_y = g(|Y|H_Z/c - Y^2 E_Y/c^2)$; quindi $f_y = e_y - (Y \times h/c)_y = gE_Y(1 - Y^2/c^2) = g^{-1}E_Y = g^{-1}F^0_Y$. Anche la componente lungo l'asse (y \equiv Y) della (5bis) è quindi corretta. In modo strettamente analogo si trova $f_z = g^{-1}F^0_Z$, e la (5bis) è così dimostrata. Per l'ulteriore generalizzazione (5) della formula di trasformazione per le forze, vedi App. Spec. 2.F. §

§5. Trasformazioni delle velocità proprie.

Poiché u (velocità propria in \mathfrak{s}) = γv e U (velocità propria in \mathfrak{S}) = ΓV , si trova:

$$(6) \quad u = \gamma v = \Gamma g \alpha^{-1} v = U + (g-1)U \cdot Y Y / Y^2 - g \Gamma Y.$$

La seconda uguaglianza (6) è dovuta alla (3), e la terza alla (2). Nel limite galileiano, $u \equiv v = U - Y$. §

§6. **Trasformazioni delle accelerazioni proprie.** Con $w =: d_\tau u$, $W =: d_T U$, si ha:

$$(7) \quad w = W + (g-1)W \cdot Y Y / Y^2 - g Y d_\tau \Gamma;$$

(la (7) si ottiene derivando la (6) rispetto a τ , e si riduce ad una forma lineare in W esprimendo $d_\tau \Gamma$ come $\Gamma U \cdot W / (c^2 - U^2)$). Ancora, nel limite galileiano $w = W$: infatti in tale limite $g \equiv 1$, e l'ultimo termine a 2° membro della (7) diventa zero, perché $\Gamma \equiv 1$. §

§7. Trasformazione delle forze di Minkowski.

È formalmente identica a quella delle accelerazioni proprie, come si desume immediatamente dalle (2.2.2, 7). §

Le cinque leggi di trasformazione parallela (2, 4, 5, 6, 7) non danno certo una particolare impressione di unità né di semplicità formale. Eppure esse hanno un'unica e semplice genesi, e la loro apparente complessità deriva dall'averne proposto la versione *3-vettoriale* (e non quella più naturale *4-vettoriale*), e nell'aver introdotto i vettori di velocità ed accelerazione come derivate, prime e seconde, rispetto a t (nel riferimento \mathfrak{s}) e rispetto a T (nel riferimento \mathfrak{S}) – e non rispetto a τ (ma quest'ultima osservazione non si applica alle (6, 7) – delle coordinate cartesiane ortogonali di x e rispettivamente di X . Unità e semplicità formale si recuperano partendo dalle (1) scritte in linguaggio vettoriale 4-dim. Diamo per ovvie, o per note, le regole dell'algebra lineare (somma, moltiplicazione per un reale) dei 4-vettori di uno spazio lorentziano 4-dim(ensionale); ed altrettanto valga per la regola leibniziana che dà il differenziale del prodotto di un reale per un 4-vettore.¹¹ Se

¹¹ Cioè, se \mathbf{a} è il 4-vettore e r è il reale, $d(r\mathbf{a}) = r d\mathbf{a} + dr\mathbf{a}$. Naturalmente le stesse regole valgono in un generico spazio pseudoeuclideo n -dim, e infine si estendono analogamente ai 2-tensori, 3-tensori, .. di questo.

introduciamo i 4-vettori $\mathbf{x} =: (x, icT)$, $\mathbf{X} =: (X, icT)$ (passaggio a coordinata römèriana immaginaria, cfr. 2.C), le (1) possono scriversi nella forma $\mathbf{x} - \mathbf{x}_0 = \mathbf{C} \circ \mathbf{X}^t$, dove il soprascritto (t) significa al solito “trasposto” (quindi \mathbf{X}^t è la colonna trasposta della riga \mathbf{X}), e dove \mathbf{C} è la seguente 4×4-matrice unitaria ($\det \mathbf{C} = 1$):

$$(8) \quad \mathbf{C} = \begin{matrix} 1+(g-1)\Upsilon_X\Upsilon_X/\Upsilon^2 & (g-1)\Upsilon_X\Upsilon_Y/\Upsilon^2 & (g-1)\Upsilon_X\Upsilon_Z/\Upsilon^2 & ig\Upsilon_X/c \\ (g-1)\Upsilon_Y\Upsilon_X/\Upsilon^2 & 1+(g-1)\Upsilon_Y\Upsilon_Y/\Upsilon^2 & (g-1)\Upsilon_Y\Upsilon_Z/\Upsilon^2 & ig\Upsilon_Y/c \\ (g-1)\Upsilon_Z\Upsilon_X/\Upsilon^2 & (g-1)\Upsilon_Z\Upsilon_Y/\Upsilon^2 & 1+(g-1)\Upsilon_Z\Upsilon_Z/\Upsilon^2 & ig\Upsilon_Z/c \\ -ig\Upsilon_X/c & -ig\Upsilon_Y/c & -ig\Upsilon_Z/c & g \end{matrix} .$$

È chiaro che gli elementi C_{ik} ($1 \leq i \leq 4$ indice di riga, $1 \leq k \leq 4$ indice di colonna) devono coincidere con i C_{ik} che abbiamo introdotto con le (2.C, 15) (in cui si faccia $n = 4$); quindi essi devono soddisfare i vincoli (2.C, 16) e (2.C, 16bis). Questo può essere direttamente verificato senza difficoltà, e con poca fatica nel caso di trasformazioni di Lorentz speciali.¹²

Pensiamo ora $\mathbf{x} = \mathbf{x}(\tau)$ e $\mathbf{X} = \mathbf{X}(\tau)$ come 4-traiettorie, in \mathfrak{s} e rispettivamente in \mathfrak{S} , di un punto cinematico con tempo proprio τ ; poiché Υ è per ipotesi costante, \mathbf{C} è anch’essa costante (cioè non dipende da τ), e si ha subito (\circ è il prodotto 4-dim):

$$(9) \quad \mathbf{v} =: d_\tau \mathbf{x} = \mathbf{C} \circ d_\tau \mathbf{X}^t \equiv \mathbf{C} \circ \mathbf{V}^t.$$

Tenendo conto delle $d_\tau t = \gamma$, $d_\tau T = \Gamma$ e della (3) e sviluppando i calcoli, si ritrovano allora le (2). Se in particolare $\Upsilon_Y = \Upsilon_Z = 0$ (trasformazioni di Lorentz speciali), si ha

$$(10_1) \quad v_x = \gamma^{-1} \Gamma g (V_X - \Upsilon_X) = \alpha (V_X - \Upsilon_X),$$

$$(10_2) \quad v_y = \gamma^{-1} \Gamma V_Y = \alpha g^{-1} V_Y,$$

$$(10_3) \quad v_z = \gamma^{-1} \Gamma V_Z = \alpha g^{-1} V_Z,$$

$$(10_4) \quad dT/dt = \alpha g^{-1}.$$

Le (10₁, 10₂, 10₃) coincidono con le (2) speciali, o (2.2.1, 4). Quanto alla (10₄), essa equivale alla t-derivata della (1b). Poiché $\mathbf{v} = \gamma(\mathbf{v}, ic)$, $\mathbf{V} = \Gamma(\mathbf{V}, ic)$, le (10₁, 10₂, 10₃) possono scriversi come

$$(11) \quad \mathbf{v} = \gamma^{-1} [\mathbf{C} \circ \mathbf{V}^t]_{sp},$$

dove con $[]_{sp}$ si intende la parte spaziale del 4-vettore entro le $[]$.

¹² La struttura della matrice \mathbf{C} implica che, se si fossero usate coordinate römèriane reali, la corrispondente matrice \mathbf{B} (vedi 2.C) sarebbe stata reale e simmetrica, avendo gli elementi della quarta riga o colonna pari a $(-g\Upsilon/c, g)$. *Le due opzioni sono del tutto equivalenti e intercambiabili*: a seconda dei casi, l’una o l’altra può apparire più naturale, o più comoda, o semplicemente più sicura. Il lettore può farsi un’idea concreta di come andrebbero le cose con l’opzione “coordinate römèriane reali” verificando che l’ultima parte del testo a partire da sotto le (10) fino alla fine di questa appendice 2.D dovrebbe essere modificata osservando le seguenti semplici regole: 1) sostituire la matrice \mathbf{C} con la \mathbf{B} ; 2) abolire l’unità immaginaria ovunque la si incontra; 3) sostituire al “prodotto interno formalmente euclideo” il “prodotto interno lorentziano”. È un esercizio un po’ noioso, ma può essere istruttivo.

I vantaggi – essenzialmente concettuali – di questo modo di procedere diventano anche più evidenti con il calcolo delle accelerazioni (4), o (2.2.1, 7). Basta prendere la derivata rispetto a τ della (9), ottenendo:

$$(12) \quad \mathbf{a} =: d_\tau \mathbf{v} = \mathbf{C} \circ d_\tau \mathbf{V}^t = \mathbf{C} \circ \mathbf{A}^t,$$

con $\mathbf{A} =: d_\tau \mathbf{V}$. Poiché $\mathbf{a} = \gamma d_t(\gamma(\mathbf{v}, ic)) = \gamma(a\gamma + v d_t \gamma, ic d_t \gamma)$, e similmente sostituendo ovunque le minuscole con le maiuscole, $\mathbf{A} = \Gamma d_\tau(\Gamma(\mathbf{V}, ic)) = \Gamma(A\Gamma + V d_\tau \Gamma, ic d_\tau \Gamma)$, si conclude che

$$(13) \quad \gamma(a\gamma + v d_t \gamma) = [\mathbf{C} \circ \Gamma(A\Gamma + V d_\tau \Gamma, ic d_\tau \Gamma)^t]_{Sp}.$$

Armandosi di un po' di pazienza, si verifica che queste (13) equivalgono alle (4), o (2.2.1, 7).

Quanto alle (5), esse sono una conseguenza diretta della legge dinamica (2.2.2, 6). Precisamente, si introducono 4-forze $\mathbf{f} = \gamma(f, id_t e/c)$, $\mathbf{F} = \Gamma(F, id_\tau E/c)$ (ove e , E , sono le energie) e le si uguagliano a $m\mathbf{a}$ e rispettivamente a $m\mathbf{A}$. Poiché $\mathbf{a} = \mathbf{C} \circ \mathbf{A}^t$, si ha subito $\mathbf{f} = \mathbf{C} \circ \mathbf{F}^t$. Secondo la (2.2.2, 6), risulta $\mathbf{f} = m\alpha g^{-1}[\mathbf{C} \circ id_\tau \mathbf{V}^t]_{Sp} = \alpha g^{-1}[\mathbf{C} \circ (F, imcd_\tau \Gamma)^t]_{Sp}$. Questa equivale alla (5): basta infatti tener conto della $d_\tau \Gamma = \Gamma^3 \mathbf{V} \cdot \mathbf{A} / c^2$ per avere $md_\tau \Gamma = \mathbf{F} \cdot \mathbf{V} / c^2$, ciò che conduce immediatamente alla tesi. In conclusione, anche le (5) si ricavano agevolmente entro la procedura generale considerata. Infine le (6) e (7) non abbisognano di verifiche dello stesso tipo, perché sono già di per sé trasformazioni tra derivate rispetto a τ . Si noti che se si fa il prodotto interno tra \mathbf{f} e \mathbf{v} si trova $\mathbf{f} \cdot \mathbf{v} = \gamma^2(\mathbf{f} \cdot \mathbf{v} - d_t e) \equiv 0$ alla luce della (2.2.2, 8). Lo stesso vale ovviamente per $\mathbf{F} \cdot \mathbf{V}$, perché il prodotto interno è invariante. Detto diversamente, la 4-forza e la 4-velocità propria sono pseudortogonali.¹³

Diventa naturale, a questo punto, introdurre una “4-QdM relativistica” come $\mathbf{p} =: m\mathbf{v} = (p, i\gamma mc)$, avendo posto al solito $\mathbf{p} =: \gamma m\mathbf{v}$. Si ha

$$(14) \quad \mathbf{p} = \mathbf{C} \circ \mathbf{P}^t,$$

con $\mathbf{P} =: m\mathbf{V}$. Secondo la (2.2.2, 10) l'equazione dinamica relativistica può trascriversi come $d_t \mathbf{p} = \mathbf{f}$. L'espressione di \mathbf{p} può modificarsi introducendo l'energia $e = \gamma mc^2 \equiv \mu c^2$ del punto (dove μ è la sua massa di moto), e diventa $\mathbf{p} = (p, ie/c)$. Il componente spaziale della (14) è $[\mathbf{p}]_{Sp} = [\mathbf{C} \circ \mathbf{P}^t]_{Sp}$, e coincide con le (2.2.2, 9) per trasformazioni di Lorentz speciali. Nello stesso caso, la quarta componente della (14) è $e = g(E - \Upsilon P_x)$, in accordo con la quarta (2.2.2, 9). Più in generale, la legge di trasformazione parallela per il 4-vettore \mathbf{p} si scrive

$$(15_1) \quad \mathbf{p} = \mathbf{P} + \Upsilon[\Upsilon \cdot \mathbf{P}(g - 1)/\Upsilon^2 - E g/c^2]$$

$$(15_2) \quad e = g(E - \Upsilon \cdot \mathbf{P}).$$

¹³ Questo fatto diventa evidente per suo conto se si considera che equivale alla $\mathbf{a} \cdot \mathbf{u} = 0$, e che $\mathbf{u} \cdot \mathbf{u} = \gamma^2(v^2 - c^2) \equiv -c^2$. Prendendo la d_τ di questa identità si ha subito $0 = d_\tau(\mathbf{u} \cdot \mathbf{u}) = 2\mathbf{a} \cdot \mathbf{u}$.

Come è naturale, con le sostituzioni $(p,P) \leftrightarrow (x,X)$ e $(e,E)/c^2 \leftrightarrow (t,T)$ le (15) diventano formalmente identiche alle trasformazioni parallele *omogenee* ($x_0 = t_0 = 0$) (1). Si noti anche che i rapporti p/x e $e/(c^2t)$ sono dimensionalmente omogenei, la loro comune dimensione essendo quella di una massa divisa per un tempo.

App. 2.E INDUZIONE DI LEGGI DI CONSERVAZIONE IN MECCANICA RELATIVISTICA SPECIALE

Esattamente come la legge dinamica newtoniana (2.1.2, 1) non esaurisce la fondazione della relativa meccanica del punto – che va completata con il principio di sovrapposizione degli effetti dinamici (v. S.sez. 2.1.3) e con quello di azione e reazione (2.1.3, 3, 3bis) –, la legge dinamica (2.2.2, 6) non esaurisce la fondazione della meccanica relativistica speciale del punto: occorrono convenienti analoghi principi di sovrapposizione e di azione e reazione.

Il principio relativistico di sovrapposizione è (induttivamente) ovvio: se su un punto materiale si esercitano più 4-forze (\equiv forze 4-dim) f , l'effetto sul suo moto è quello che eserciterebbe la 4-forza *somma* delle singole 4-forze, definita “per componenti” secondo le regole standard dell'algebra lineare dei 4-vettori di uno spazio 4-dim (e in generale degli n -vettori di uno spazio n -dim).

L'induzione del principio relativistico di azione e reazione è un po' più delicata. Di fatto, la versione newtoniana di questo principio presentata nella S.sez. 2.1.2 si riferiva per semplicità al caso di *due* soli punti materiali; ma essa si estende (induttivamente) senza difficoltà ad un insieme \mathcal{A} di un numero finito arbitrario di punti soggetti soltanto a forze interne, cioè a forze esercitate su ogni punto di \mathcal{A} dagli altri e soli punti di \mathcal{A} stesso. Se f_a è la forza interna agente sul punto a -mo di \mathcal{A} , alla (2.1.2, 3) si sostituisce la $\sum_{a \in \mathcal{A}} f_a = 0$; e alla (2.1.2, 3bis), o piuttosto alla equivalente (2.1.2, 3ter), la $\sum_{a \in \mathcal{A}} f_a \times (P_a - O) = 0$, dove P_a è la posizione del punto a -mo e O è un polo arbitrario, ad esempio l'origine delle coordinate.

Per passare all'analogo principio relativistico, conviene analizzare cosa succede al moto di un singolo punto quando non soltanto la sua 4-accelerazione, ma anche il “momento rispetto all'origine” di questa, sono uguali a zero. La risposta è ovvia nel primo caso: $d_\tau v = 0$, cioè $v = \gamma(v, ic)$ si conserva. Nel secondo caso, occorre innanzitutto chiarire cosa si intende per momento (rispetto all'origine) di un 4-vettore. A differenza di quanto succede in uno spazio 3-dim, in cui il numero delle componenti di un vettore uguaglia quello delle componenti indipendenti di un 2-tensore antisimmetrico (3, e rispettivamente $3(3-1)/2$), in uno spazio ($n \neq 3$)-dim vi è un divario tra

questi due numeri, che sono n e rispettivamente $n(n-1)/2$ (ad esempio per $n = 4$, 4 e rispettivamente 6). Inoltre nello spazio euclideo 3-dim il prodotto vettore di due vettori c e d (nell'ordine) è un vettore $c \times d$ di componenti $(c \times d)_i = c_{i+1}d_{i+2} - c_{i+2}d_{i+1}$ (da leggere mod3, e per $i = 1,2,3$). Passando ad uno spazio 4-dim e a due suoi generici 4-vettori c e d , otteniamo per analogia il *di-4-vettore* antisimmetrico di componenti $(i_k) c_i d_k - c_k d_i$, con $i,k = 1, \dots, 4$ (le componenti indipendenti di questo oggetto sono sei). Se ora identifichiamo c con la 4-accelerazione a del punto, e d con la sua 4-posizione x , otteniamo, scrivendo come è usuale ${}_{[i k]}$ in luogo di $(i_k - k_i)/2$, $a_{[i} x_{k]} = d_\tau v_{[i} x_{k]} = d_\tau(v_{[i} x_{k]}) - v_{[i} d_\tau x_{k]} \equiv d_\tau(v_{[i} x_{k]})$, perché $d_\tau x = v$. Per definizione, le $2c_{[i} x_{k]}$ sono le componenti (i_k) del momento (rispetto all'origine) di c . Se dunque il momento di a rispetto all'origine è zero (cioè sono nulle le $a_{[i} x_{k]}$ per ogni $i,k = 1, \dots, 4$), allora il corrispondente momento di v (un di-4-vettore antisimmetrico) si conserva. Conviene snellire le notazioni scrivendo u per il di-4-vettore antisimmetrico di componenti (i_k) uguali a $2v_{[i} x_{k]}$. Allora il precedente asserto si può ripetere dicendo che “ u si conserva”.

Ciò premesso, sia \mathcal{A} un insieme di punti materiali soggetti soltanto a 4-forze interne (vedi la corrispondente precedente definizione nel caso euclideo-newtoniano). Se f_a è la 4-forza interna agente sul punto a -mo, il principio relativistico di azione e reazione afferma che

$$(1) \quad \sum_{a \in \mathcal{A}} (f_a)_i = 0,$$

$$(2) \quad \sum_{a \in \mathcal{A}} (f_a)_{[i} (x_a)_{k]} = 0$$

per ogni $i,k = 1, \dots, 4$. Rimarchiamo che le (1,2) sono *indotte* (e non *dedotte*).

In forza della $m_a a_a = f_a$, valida per ogni a di \mathcal{A} , non sono dunque la 4-accelerazione a_a del singolo punto e il suo momento (rispetto all'origine) ad essere zero – come più sopra supposto per meglio spiegarci – ma piuttosto le loro combinazioni lineari con pesi m_a per $a \in \mathcal{A}$, secondo le

$$(3) \quad \sum_{a \in \mathcal{A}} m_a (a_a)_i = 0,$$

$$(4) \quad \sum_{a \in \mathcal{A}} m_a (a_a)_{[i} (x_a)_{k]} = 0,$$

sempre per ogni $i,k = 1, \dots, 4$. Concludiamo che la (3) equivale alla

$$(5) \quad d_\tau(\sum_{a \in \mathcal{A}} m_a v_a) = 0,$$

e la (4) alla

$$(6) \quad d_\tau(\sum_{a \in \mathcal{A}} m_a u_a) = 0.$$

Le (5,6) traducono i due fondamentali principi relativistici di conservazione per un insieme “isolato” \mathcal{A} di punti: la (5), quello di conservazione della somma delle 4-QdM dei punti di \mathcal{A} , o

brevemente della “4-QdM di \mathcal{A} ”; e la (6), quello della somma dei loro momenti, o brevemente del “momento della 4-QdM di \mathcal{A} ”.¹⁴

Le stesse (5,6) trovano immediata applicazione in **dinamica relativistica impulsiva**, se le interazioni tra i punti di \mathcal{A} consistono in *collisioni*. In questo caso le derivate rispetto a τ diventano idealmente distribuzioni di Dirac, ed integrando le (5,6) attraverso dette collisioni si ottengono le

$$(7) \quad \sum_{a \in \mathcal{A}} m_a v_a = 0,$$

$$(8) \quad \sum_{a \in \mathcal{A}} m_a u_a = 0,$$

dove $\sum =: \sum^+ - \sum^-$ e i soprascritti $(^+, -)$ significano “per τ (dopo le, prima delle) collisioni”.

Naturalmente le definizioni e i risultati che precedono hanno il loro giusto senso se le masse di quiete m_a sono, come devono essere nel modello macroscopico al quale ci siamo sempre limitati, invarianti di moto. Se questa condizione viene meno, cioè se l’energia intrinseca di massa $m_a c^2$ è significativamente alterata attraverso le collisioni – come avviene di norma nel mondo microscopico – allora le equazioni di conservazione cambiano completamente forma e significato. Ad esempio in luogo delle quattro (7) si hanno le tre $\sum_{a \in \mathcal{A}} \mu_a v_a = 0$ più la $\sum_{a \in \mathcal{A}} \mu_a = 0$, dove μ_a è la massa di moto γm_a del punto a -mo. Tuttavia la considerazione di tale tipo di eventi esula dal nostro programma generale, giusto il titolo di questo libro.

APP. 2.F ANCORA SULLA RELAZIONE TRA MASSA DI QUIETE E MASSA DI MOTO (SECONDO LEWIS E TOLMAN)

Ripercorriamo brevemente la procedura *induttiva* che nella S.sez. 2.2.2 ha portato alla fondamentale legge dinamica relativistica (2.2.2, 6). Il primo passo è quello della legge dinamica (2.2.2, 1) nel riferimento di quiete istantanea \mathfrak{S} . Un secondo passo è quello della uguaglianza tra m (massa di P in \mathfrak{s}) e M (massa di P in \mathfrak{S}). Il terzo passo è quello della struttura della funzione $\aleph(m, a, f)$, assunta del tipo $m\Lambda(a) - f$, con Λ lineare in a . Segue la trasformazione tra f e F secondo la (2.2.2, 4bis) che abbiamo giustificato in (2.D, 4), e quindi la $\Lambda(a) = (g^3 a_x, g a_y, g a_z)$. La (2.2.2, 6) segue infine con un calcolo elementare. Secondo la nuova notazione introdotta nella S.sez. 2.2.2 ($m^0 \equiv$ massa di quiete, $m \equiv$ massa di moto), la (2.2.2, 6) si riscrive come $d_t(mv) = f$ (vedi (2.2.2, 6bis)), in cui

¹⁴ Questa espressione può essere fraintesa, ma ogni equivoco dovrebbe escludersi finché \mathcal{A} consta di più di un punto materiale; infatti in tal caso la posizione della “4-QdM di \mathcal{A} ” non è definita, e quindi non avrebbe senso calcolarne il momento. Invece la stessa espressione è ben definita e corretta se \mathcal{A} comprende un unico punto.

$$(1) \quad m = \gamma m^0$$

e γ è al solito $(1-v^2/c^2)^{-1/2}$.

La relazione (1) è un'altra "icona" della relatività speciale, ed è naturale chiedersi se essa non possa essere indotta mediante una procedura che *non* richieda, come la precedente appena ricordata, il riferimento alla teoria elettromagnetica. Lewis e Tolman¹⁵ hanno mostrato come questo sia possibile *restando in ambito strettamente meccanico*, ma naturalmente facendo uso di altre premesse. Essi partono dalla equazione dinamica (2.2.2, 6bis), $d_t[m(|v|)v] = f$, in cui tuttavia la funzione $m = m(|v|)$ si suppone ancora indeterminata. Più precisamente, di essa ci si limita a richiedere che sia definita in $(0,c)$, sia ivi continua e strettamente monotona, e che il suo limite per $|v| \rightarrow 0+$ esista finito. Denoteremo con $m(0)$ tale limite.

Quanto segue illustra come la funzione m possa essere determinata. La nuova premessa induttiva consiste nel richiedere la conservazione della QdM totale di un sistema *isolato* costituito da due punti materiali identici P_1 e P_2 (quindi in ultima analisi nell'uso del principio di azione e reazione). Questi punti, le cui masse di quiete possono essere assunte unitarie senza limitazione di generalità, sono visti dai riferimenti inerziali \mathfrak{S} e \mathfrak{s} , il secondo al solito in moto con velocità Υ (di modulo in $(0,c)$) rispetto al primo. P_1 e P_2 si suppongono collidere in un certo istante: in questo consiste la loro interazione. Denotando con V_1 e V_2 [con v_1 e v_2] le loro velocità rispetto a \mathfrak{S} [rispetto a \mathfrak{s}] *prima* della collisione, e con gli stessi simboli affetti da apici le loro velocità *dopo* la collisione, l'assunta conservazione della QdM totale, ad es. vista da \mathfrak{S} , richiede che sia

$$(2) \quad m(|V_1|)V_1 + m(|V_2|)V_2 = m(|V_1'|)V_1' + m(|V_2'|)V_2'.$$

(La condizione di conservazione deve valere anche in \mathfrak{s} , e si ottiene sostituendo nella (2) le velocità maiuscole con le minuscole; ma sarà qui sufficiente limitarci alla (2).) In particolare, si supponga adesso che $v_2 = -V_1$ (in forza della (2.D, 2) deve allora anche essere $v_1 = -V_2$); e inoltre, che $V_1' = -\alpha V_1$ con $\alpha > 0$ (deve allora anche essere $v_2' = -\alpha v_2$). Infine si supponga che $\Upsilon \cdot V_1 = 0$ (quindi che $\Upsilon \cdot v_2 = 0$, $\Upsilon \cdot V_1' = 0$, $\Upsilon \cdot v_2' = 0$). Sotto l'insieme di queste condizioni, ancora la (2.D, 2) implica che

$$(3) \quad V_2 = \Upsilon - V_1 g^{-1},$$

ove g^{-1} è al solito $(1-\Upsilon^2/c^2)^{1/2}$, e quindi che

$$(4) \quad V_2^2 = \Upsilon^2 + V_1^2 g^{-2};$$

e similmente, che

$$(3') \quad V_2' = \Upsilon - V_1' g^{-1},$$

e quindi che

¹⁵ G. Lewis, R. Tolman, *Phil. Mag.* **18**, 510 (1909).

$$(4') \quad V_2'^2 = Y^2 + V_1'^2 g^{-2}.$$

Sostituendo le (3,3') e le loro conseguenze (4,4') nella (2) si ottiene

$$(5) \quad m(|V_1|)V_1 + m([Y^2 + V_1^2 g^{-2}]^{1/2})(Y - V_1 g^{-1}) = m(|V_1'|)V_1' + m([Y^2 + V_1'^2 g^{-2}]^{1/2})(Y - V_1' g^{-1}),$$

in cui compaiono soltanto le velocità di P_1 rispetto a \mathfrak{S} . Si può senz'altro supporre $V_1 \neq 0$, perché in caso contrario la (5) si riduce ad una identità. Dalla (5) si trae subito, moltiplicandola scalarmente per Y e dividendola per $Y^2 (\neq 0)$:

$$(6) \quad m([Y^2 + V_1^2 g^{-2}]^{1/2}) = m([Y^2 + V_1'^2 g^{-2}]^{1/2}).$$

Per l'assunto carattere monotono di m , la (6) implica che siano uguali gli argomenti tra [], cioè che $V_1^2 = V_1'^2$, ossia che $\alpha = 1$ e $V_1 = -V_1'$: le velocità di P_1 prima e dopo la collisione sono uguali ed opposte. Sotto questa condizione, e sotto la $V_1 \neq 0$, la (5) si riduce alla equazione funzionale in m

$$(7) \quad gm(|V_1|) = m([Y^2 + V_1^2 g^{-2}]^{1/2}).$$

Prendendo il limite per $|V_1| \rightarrow 0+$ di questa, per l'assunta continuità di m abbiamo

$$(8) \quad gm(0) = m(|Y|),$$

che è valida per ogni $|Y|$ in $(0,c)$ e che si riduce ad una identità per $|Y| \rightarrow 0+$. (Nella (8), $g = (1 - Y^2/c^2)^{-1/2}$ e $m(0) = m^0$.) La funzione m è così completamente identificata, e soddisfa le restrizioni assunte a priori. La (8) coincide inoltre con la (1), che è così giustificata senza ricorrere alla teoria elettromagnetica, come si voleva.

Avendo determinato la funzione $m(|v|) = m^0 \gamma(|v|)$, sono completamente determinate la QdM $p = mv$ e l'energia totale $e = mc^2$; quindi, un 4-vettore che risulta trasformarsi, nel solito passaggio dal sistema inerziale \mathfrak{S} al sistema \mathfrak{s} , come la 4-coordinata (x,t) , ovvero secondo le (2.D, 15_{1,2}). D'altra parte l'equazione dinamica relativistica $d_t p = f$ è stata ora assunta come punto di partenza, e con essa la conseguenza $d_t e = f \cdot v$. Poiché a questo punto le formule di trasformazione per p ed e sono note, possiamo ricavare le formule di trasformazione per le forze, che risultano coincidere con le (2.D, 5) (le verifiche sono lasciate al lettore). L'intera dinamica relativistica speciale del punto è così (ri)costruita su basi strettamente meccaniche, senza ricorrere alla teoria elettromagnetica.

APP. 2.G DESCRIZIONE LAGRANGIANA E DESCRIZIONE EULERIANA DEL MOTO DI UN CONTINUO

È parso opportuno dedicare la presente appendice 2.G ad una questione che interviene frequentemente nella fisica matematica dei mezzi continui. Diremo di usare una **descrizione storica** o **lagrangiana** del moto di un continuo ($m \geq 1$)-dimensionale se la posizione P al tempo t del suo generico elemento – elemento che diremo per brevità “ π ” – è data come funzione di t e di m

parametri reali che identificano stabilmente π . In particolare tali parametri possono farsi coincidere con la posizione P_0 di π ad un tempo prefissato t_0 (ad esempio $t_0 = 0$). In questo caso la posizione P di π al tempo t intorno a t_0 è nota come funzione (che si supporrà abbastanza regolare) di t e di P_0 , diciamo

$$(1) \quad P = f(t, P_0)$$

in un aperto $U = \Delta t \times \Omega_0$, con $\Delta t \equiv$ intervallo (aperto) contenente t_0 e $\Omega_0 \equiv$ aperto connesso di \mathbb{R}^m contenente P_0 . Per definizione, la funzione f è soggetta alla condizione $f(t_0, P_0) = P_0$. Supponendo la f (continuamente) derivabile rispetto a t in U , la velocità di π è ovviamente data da

$$(2) \quad v = v(t, P_0) = \partial f / \partial t(t, P_0);$$

quindi, supponendola due volte derivabile, l'accelerazione di π è data da

$$(3) \quad a = a(t, P_0) = \partial^2 f / \partial t^2(t, P_0),$$

... e così via, per ogni $(t, P_0) \in U$.

Evidentemente la descrizione lagrangiana del moto di un continuo è la più naturale possibile dal punto di vista concettuale; ma non è quella più usata in pratica nelle applicazioni. Una descrizione alternativa e spesso più comoda dello stesso moto è infatti quella cosiddetta **di campo**¹⁶ o **euleriana**. Essa semplicemente presuppone che la (1) sia invertibile rispetto a P_0 in U , presentandosi quindi nella forma equivalente

$$(1bis) \quad P_0 = P_0(t, P) = g(t, P),$$

con $g(P_0, t_0) = P_0$.

Per definizione si ha dunque $f(t, g(t, P)) \equiv P$ identicamente per ogni $t \in \Delta t$ e ogni $P \in \Omega_t \equiv f(t, \Omega_0)$; e in particolare, $f(t_0, g(t_0, P_0)) \equiv P_0$ per ogni $P_0 \in \Omega_0$.

Il moto del continuo in oggetto è facilmente descrivibile in termini della funzione g , supposta di congrua CdC in $\Delta t \times \Omega_t$. Introduciamo innanzitutto il **campo di spostamenti**

$$(4) \quad S = S(t, P) =: P - P_0(t, P) = P - g(t, P),$$

come funzione di t e della posizione *attuale* P dell'elemento π di posizione P_0 al tempo t_0 . Se g è di classe C^1 , la velocità di π (o piuttosto il campo di velocità $v = v(t, P)$ degli elementi di posizione attuale P) è la derivata *sostanziale* (o *convettiva*) rispetto a t , d/dt , di S , cioè

$$(5) \quad v = v(t, P) = dS/dt(t, P) = \partial S / \partial t(t, P) + \partial f / \partial t(t, g(t, P)) \cdot \partial S / \partial P(t, P),$$

dove il (\cdot) ha il solito significato di prodotto scalare in dimensione m . Per verificarlo, cominciamo con l'osservare che $\partial S / \partial t = -\partial g / \partial t$ e $\partial S / \partial P = \mathbf{1}_m - \partial g / \partial P$, dove $\mathbf{1}_m$ è la $(m \times m)$ -diade unitaria. Quindi

$$(6) \quad v = -\partial g / \partial t + \partial f / \partial t \cdot (\mathbf{1}_m - \partial g / \partial P).$$

¹⁶ Il riferimento è essenzialmente al "campo di spostamenti" di cui alla successiva (4).

D'altra parte derivando sostanzialmente rispetto a t la (1bis) (e tenendo conto del fatto che P_0 è indipendente da t) abbiamo

$$(7) \quad 0 = \partial g / \partial t + \partial f / \partial t \cdot \partial g / \partial P;$$

sostituita nella (6), questa dà immediatamente $v = \partial f / \partial t$, qed. In modo analogo ma più laborioso si procede per il campo delle accelerazioni, avendo supposto la g di CdC 2. Si ha:

$$(8) \quad a = d^2 S / dt^2 = \partial^2 S / \partial t^2 + 2 \partial f / \partial t \cdot \partial^2 S / \partial P \partial t + \partial^2 f / \partial t^2 \cdot \partial S / \partial P + \partial f / \partial t \partial f / \partial t : \partial^2 S / \partial P \partial P,$$

dove la (7) ha permesso di eliminare un contributo in $\partial^2 f / \partial P_0 \partial t$. Qui $\partial^2 S / \partial t^2 = - \partial^2 g / \partial t^2$, $\partial^2 S / \partial P \partial t = - \partial^2 g / \partial P \partial t$ e $\partial^2 S / \partial P \partial P = - \partial^2 g / \partial P \partial P$. Non occorre sostituire materialmente queste posizioni nella (8) per convincersi che, confrontando il risultato con la t -derivata sostanziale della (7), e in forza della $\partial S / \partial P = \mathbf{1}_m - \partial g / \partial P$, si è ridotti alla $a = \partial^2 f / \partial t^2$, qed. (Il calcolo esplicito può comunque essere fatto, e porta alla conclusione anzidetta.) Gli stessi meccanismi, in cui si tenga conto delle equazioni di ordine più basso (come della (7) nel precedente caso della derivata sostanziale di S di ordine 2) assicurano risultati analoghi per le derivate sostanziali di S di ordine > 2 .

Mediante la (4) è facile ricavare l'espressione dello spostamento S (supposto di congrua CdC) in un intorno infinitesimo di P . Precisamente, se il vettore $Q - P$ è trattato come quantità piccola del 1° ordine, trascurando di evidenziare la dipendenza da t abbiamo

$$(9) \quad S(Q) = S(P) + \partial S / \partial P(P) \cdot (Q - P).$$

Per $m = 3$, scomponendo l'omografia $\partial S / \partial P$ in un componente simmetrico $(S_{ij} + S_{ji})/2$ (con ${}_j =: \partial / \partial x^j$) e in un componente antisimmetrico $(S_{ij} - S_{ji})/2$, e tenendo conto delle proprietà del simbolo totalmente antisimmetrico (triplo) Υ_{ijk} dello spazio euclideo 3-dim, si trova senza difficoltà:

$$(10) \quad S(Q) = S(P) + (1/2) \text{rot} S(P) \times (Q - P) + \text{dil} S(P) \cdot (Q - P),$$

dove $\text{dil} S$ è l'omografia simmetrica (detta anche **dilatazione**) corrispondente al 2-tensore simmetrico $(S_{ij} + S_{ji})/2$, o **2-tensore di deformazione** relativo al campo di spostamenti S . Così la (10) mostra che il campo di spostamenti S in un intorno del 1° ordine di P si decompone nella somma di uno spostamento rototraslatorio di traslazione $S(P)$ e angolo $(1/2) \text{rot} S(P)$, e in una dilatazione $\text{dil} S(P) \cdot (Q - P)$. Dividendo la (10) per il tempo infinitesimo δt durante il quale si verifica lo spostamento infinitesimo S , abbiamo:

$$(11) \quad v(Q) = v(P) + (1/2) \text{rot} v(P) \times (Q - P) + \text{dil} v(P) \cdot (Q - P),$$

dove $v = S / \delta t$, e $\text{dil} v$ è l'omografia simmetrica (detta anche **velocità di dilatazione**) corrispondente al 2-tensore simmetrico $(v_{ij} + v_{ji})/2$ (**tensore di velocità di deformazione** relativo al campo di velocità $v(P)$). Secondo la (11), il moto in un intorno del 1° ordine del punto P si decompone nella somma di un moto traslatorio di velocità $v(P)$, di un moto rotatorio intorno a P e di spin $(1/2) \text{rot} v(P)$, e di un moto dilatatorio $\text{dil} v(P) \cdot (Q - P)$.

Nella cinematica dei continui, e in particolare (come abbiamo visto nel Cap. 2) dei continui *rigidi*, di norma si adotta una descrizione euleriana del moto. Uno spostamento rigido è lineare affine, e si può verificare direttamente che in tal caso $\text{dil}S \equiv 0$ e $\text{rot}S$ non dipende da P ; la (10) si riduce dunque a

$$(10\text{bis}) S(Q) = S(P) + (1/2)\text{rot}S \times (Q-P).$$

Per la linearità, inoltre, questa è corretta per $Q - P$ non necessariamente infinitesimo. Passando alla (11), abbiamo $\text{dil}v \equiv 0$ e $\text{rot}v$ indipendente da P , cioè

$$(11\text{bis}) v(Q) = v(P) + (1/2)\text{rot}v \times (Q-P),$$

valida per $Q - P$ non necessariamente infinitesimo. Se questa si confronta con la (2.A, 16ter) in cui si faccia $v_{\text{rel}} = 0$, si vede che $\text{rot}v/2 = \omega$.

App. 2.H ELEMENTI DI DINAMICA CLASSICA E RELATIVISTICA DEI MEZZI MATERIALI CONTINUI

Abbiamo introdotto la legge dinamica newtoniana (2.1.2, 1) e relativistica (2.2.2, 6), che qui riscriviamo per comodità del lettore

$$(1_1) \quad m^0 d_t v(t) = f^0(x, t, v);$$

$$(1_2) \quad d_t x = v,$$

e rispettivamente

$$(2_1) \quad d_t(mv)(t) = f(x, t, v);$$

$$(2_2) \quad d_t x = v,$$

nonché le loro conseguenze “di potenza” (2.1.2, 1')

$$(1_3) \quad m^0 d_t v^2(t)/2 = v \cdot f^0(x, t, v)$$

e rispettivamente (2.2.2, 8)

$$(2_3) \quad d_t(mc^2)(t) = v \cdot f(x, t, v),$$

riferendoci ad un punto materiale, o al più, ad un numero finito di tali punti. Nelle precedenti, m^0 è la massa di quiete del punto ed m quella relativistica, legate tra loro, ricordiamo, dalla $m = \gamma m^0$, con $\gamma = \gamma(v) = (1 - v^2/c^2)^{-1/2}$; e similmente f^0 è la forza di quiete agente sul punto, ed f quella relativistica, legate tra loro dalla $f = \gamma^{-1} f^0 + \gamma v \cdot f^0 (1 - \gamma^{-1})/v^2$, o se si preferisce, dalle $f_{\parallel} = f^0_{\parallel}$, $f_{\perp} = \gamma^{-1} f^0_{\perp}$ (con $\parallel \equiv$ componente parallelo, $\perp \equiv$ componente perpendicolare).

Una parte cospicua dell'interesse teoretico va tuttavia alla dinamica dei mezzi materiali *continui*, cioè dei mezzi materiali che occupano una intera regione aperta dello spazio euclideo. Fondandoci sulle (1, 2), vogliamo indurre le leggi dinamiche che governano – a meno delle solite

condizioni accessorie – il moto di questi mezzi. Partiremo allo scopo da un sistema costituito da *molti* punti materiali di *piccola* massa, tanti da potere da una parte trattare il numero di punti contenuti in un *piccolo* volume come una variabile continua, e dall'altra tale volume e la sua massa come infinitesimi.¹⁷

Ragionando dapprima in termini di meccanica newtoniana, e denotando con (x,t) le usuali variabili indipendenti euclidee-newtoniane, distingueremo mediante il soprascritto $(^0)$ le grandezze “classiche” (cioè, nel limite $c \rightarrow \infty$) dalle corrispondenti grandezze relativistiche. (Questo appesantisce un po' la notazione, ma vale la pena pagarne il prezzo.) Va qui peraltro osservato che mentre la differenza concettuale tra grandezza classica $(^0)$ e corrispondente grandezza “di quiete istantanea” $(^0)$ è ovvia e importante, in molti casi essa sparisce. Ad esempio, e come si può facilmente verificare, $m^0 = m^0$, $(d_t m)^0 = (d_t m)^0 (= 0)$, ecc.

Detto dunque $\tau^0 \neq 0$ il piccolo volume in questione, denoteremo con $n^0 \tau^0$ il numero dei punti in τ^0 , che si conserva durante il moto. n^0 (\equiv numero dei punti nell'unità di volume) si dice **densità numerica**. Sia n^0 che τ^0 vanno pensati come campi euleriani (v. 2.G) abbastanza regolari, $n^0 = n^0(x,t)$, $\tau^0 = \tau^0(x,t)$; ed il fatto che $n^0 \tau^0$ si conserva è tradotto dalla $(^+)$ $d_t(n^0 \tau^0) = 0$, dove $d_t =: \partial_t + u \cdot \nabla_x$ è la t -derivata sostanziale e $u = u(x,t)$ è il campo euleriano di velocità, anch'esso supposto abbastanza regolare. Scrivendo per semplicità ∇_x come ∇ , risulta $d_t \tau^0 = \tau^0 \nabla \cdot u$ (esercizio); quindi la $(^+)$ diventa $0 = n^0 \nabla \cdot u + \partial_t n^0 + u \cdot \nabla n^0 = \nabla \cdot (u n^0) + \partial_t n^0$. Per il dato campo $u(x,t)$, l'operatore differenzialparziale (lineare ma in generale non leibniziano) $C_u(*) =: \nabla \cdot (u *) + \partial_t(*) \equiv (*) \nabla \cdot u + d_t(*)$ (dove $*$ indica il campo operando, possibilmente vettoriale o tensoriale) si dice **di conservazione nel continuo** (o semplicemente **di conservazione**). $C_u n^0 = 0$ equivale dunque alla $(^+)$, attestando che l'insieme dei punti $n^0 \tau^0$ si conserva durante il moto. Se i punti hanno *comune* massa m^0 , il prodotto $\mu^0 =: n^0 m^0$ è la **densità di massa** del mezzo (\equiv massa nell'unità di volume). Poiché m^0 è indipendente da t , $C_u n^0 = 0$ equivale a $C_u \mu^0 = 0$; questa esprime dunque la conservazione della generica porzione di massa durante il moto. Denotiamo ora con $\varphi^0 = n^0 f^0 =: \varphi^0(x,t,u)$ la densità di forza (\equiv forza agente sull'unità di volume) classica: dalla (1_1) induciamo la **legge dinamica newtoniana per un continuo materiale** (sottintendendo la $C_u \mu^0 = 0$):

$$(3) \quad \mu^0 d_t u = \varphi^0(u),$$

dove $d_t u$ è l'accelerazione euleriana del mezzo, e abbiamo trascurato di esplicitare le variabili indipendenti. Mostriamo adesso che sotto la $C_u \mu^0 = 0$ il 1° membro della (3) uguaglia $\nabla \cdot (\mu^0 u u) + \partial_t (\mu^0 u) \equiv C_u (\mu^0 u)$. Sviluppando, risulta infatti $C_u (\mu^0 u) = u \nabla \cdot (\mu^0 u) + \mu^0 u \cdot \nabla u + u \partial_t \mu^0 +$

¹⁷ Questo stesso punto di vista è stato adottato (senza denunciarlo esplicitamente, in modo intuitivo) nella Sez. 2.5.

+ $\mu^0 \partial_t u = u C_u \mu^0 + \mu^0 d_t u \equiv \mu^0 d_t u$.¹⁸ In conclusione, il modello newtoniano della dinamica di un continuo materiale è caratterizzato dalle quattro equazioni differenziali parziali (sempre trascurando di esplicitare le variabili indipendenti x, t)

$$(4_1) \quad C_u(\mu^0 u) = \phi^0(u) (= n^0 f^0(u))$$

$$(4_2) \quad C_u \mu^0 = 0,$$

nelle altrettante incognite (u, μ^0) . $\mu^0 u = (\mu^0 u)(x, t)$ è la **densità di quantità di moto** (QdM) del continuo materiale. Sotto le opportune condizioni accessorie (iniziale e al contorno), le (4) governano il moto newtoniano del continuo materiale considerato.¹⁹

La legge scalare “di potenza” segue dalla (4₁) moltiplicandola scalarmente per u ed è:

$$(4_3) \quad u \cdot C_u(\mu^0 u) = \phi^0(u) =: u \cdot \phi^0(u).$$

Mostriamo che sotto la $C_u \mu^0 = 0$ il 1° membro della (4₃) uguaglia $C_u k^0$, dove $k^0 =: \mu^0 u^2 / 2 \equiv$ **densità di energia cinetica** del mezzo. Abbiamo infatti $\nabla \cdot (\mu^0 u u^2) = u^2 \nabla \cdot (u \mu^0) + \mu^0 u \cdot \nabla u^2$, e similmente $\partial_t (\mu^0 u^2) = u^2 \partial_t \mu^0 + \mu^0 \partial_t u^2$; da cui, sommando e dividendo per 2, $C_u(\mu^0 u^2 / 2) \equiv C_u k^0 = (u^2 C_u \mu^0 + \mu^0 d_t u^2) / 2 = \mu^0 u \cdot d_t u = u \cdot C_u(\mu^0 u)$. In conclusione potremo scrivere la (4₃) nella forma

$$(4_3') \quad C_u k^0 = \phi^0.$$

Continuando a denotare con (x, t) le variabili indipendenti, passiamo ora ad occuparci di moto relativistico. La legge del moto relativistico del punto è la (2₁); moltiplicando questa per n , densità numerica relativistica, uguale a γn^0 in forza della contrazione di Lorentz, induciamo la legge dinamica relativistica per un continuo materiale

$$(5) \quad n d_t(\mu u) = \phi(u)$$

dove $\phi = \phi(x, t, u)$ ($= n f$) è la densità di forza relativistica agente sul mezzo, e abbiamo al solito trascurato di esplicitare le variabili indipendenti. Mostriamo che sotto la $C_u n = 0$ ²⁰, il 1° membro di questa è uguale a $\nabla \cdot (\mu u u) + \partial_t(\mu u) \equiv C_u(\mu u)$, con $\mu =: n m$, densità di massa relativistica, uguale a $\gamma^2 \mu^0$. Infatti $n d_t(\mu u) = d_t(\mu u) - \mu u d_t n = d_t(\mu u) - \mu u C_u n + \mu u \nabla \cdot u = \nabla \cdot (\mu u u) + \partial_t(\mu u) = C_u(\mu u)$. In conclusione la **legge dinamica relativistica per un continuo materiale** è (sempre trascurando di esplicitare le variabili indipendenti)

$$(6_1) \quad C_u(\mu u) = \phi(u).$$

¹⁸ In base ad una nota identità vettoriale, il termine $u \cdot \nabla u$ viene spesso scritto come $\nabla u^2 / 2 + (\nabla \times u) \times u$.

¹⁹ La (3), o la (4₁) ad essa equivalente sotto la $C_u n^0 = 0$, sono state dedotte per un modello di continuo costituito da una *singola* specie di punti, ma valgono anche per continui costituiti da *più* specie di punti, purché un *singolo* campo di velocità descriva il moto macroscopico.

²⁰ Per il suo significato, la $d_t \tau = \tau \nabla \cdot v$ vale anche in ambito relativistico. Tenuto allora conto della $n \tau = \gamma n^0 \gamma^{-1} \tau^0 = n^0 \tau^0$, si ha $\tau C_u n = \tau d_t n + n d_t \tau = d_t(n \tau) = d_t(n^0 \tau^0) = 0$, quindi $C_u n = 0$. In altre parole, dal punto di vista delle trasformazioni $C_u n$ è uno scalare 4-dim. Usando i 4-vettori $u^i = \gamma(u^i, ic)$ e $n^0 u^i = n(u^i, ic)$, risulta infatti: $C_u n = \nabla \cdot (n u) + \partial_t n = \partial_t(n u^i) + (1/c) \partial_t(nc) = \partial_t(n^0 u^i)$.

Confrontata con la corrispondente legge classica (4₁), essa appare formalmente identica a quest'ultima con la sola cancellazione dei soprascritti (⁰). Alla (6₁) si deve aggiungere una quarta equazione scalare (perché le incognite sono sempre quattro, μ e u), diciamo la

$$(6_2) \quad C_u \mu = \mu \gamma^2 d_t u \cdot u / c^2,$$

con $\gamma^2 = (1 - u^2/c^2)^{-1}$. Infatti $C_u \mu = d_t \mu + \mu \nabla \cdot u = n d_t m + m(C_u n - n \nabla \cdot u) + \mu \nabla \cdot u = n d_t m$ per la $C_u n = 0$; e d'altra parte

$$(7) \quad d_t m = m^0 d_t \gamma = m \gamma^2 u \cdot d_t u / c^2,$$

da cui la (6₂).

La legge scalare “di potenza” che consegue dalla (7₁) si ottiene al solito moltiplicandola scalarmente per u :

$$(6_3) \quad u \cdot C_u(\mu u) = u \cdot \varphi(u) =: \phi(u),$$

ancora formalmente identica alla (4₃) del caso classico con la sola cancellazione dei (⁰). Si noti che $u \cdot f^0 = u \cdot f$ (vedi App. Spec. 2.D, 4), per cui $\phi = u \cdot f n = u \cdot f^0 n^0 \gamma = \phi^0 \gamma$. Mostriamo adesso che, sotto la solita condizione di conservazione del numero, $C_u n = 0$, il 1° membro della (6₃) è uguale a $C_u(\mu c^2)$. Infatti in tal caso si ha da una parte (⁺) $C_u \mu = n d_t m$, e dall'altra $C_u(\mu u) = n d_t(mu)$. Moltiplicando scalarmente la seconda di queste per u si ha $u \cdot C_u(\mu u) = n(u^2 d_t m + \mu u \cdot d_t u) = n c^2 d_t m$, ove l'ultima uguaglianza scende dalla (7). In forza della (⁺) è poi $u \cdot C_u(\mu u) = c^2 C_u \mu = C_u(\mu c^2)$. In definitiva potremo scrivere la (6₃) nella forma

$$(8) \quad C_u(\mu c^2) = \phi(u),$$

dove sotto C_u compare la **densità di energia relativistica** μc^2 .

Il sistema delle equazioni (6₁, 8) (equivalente al sistema delle (6₁, 6₂)) può ricevere, con certi vantaggi formali e sostanziali, una formulazione quadridimensionale. Facendo al solito variare gli indici latini su 1, ..., 4 e quelli greci su 1, ..., 3, e usando coordinate römeriane immaginarie, la 4-velocità propria (v. 2.D, 8) è $u^i =: c dx^i / \sqrt{(-ds^2)} = \gamma(u^i, ic)$. L'ultima uguaglianza, che si verifica immediatamente, evidenzia che u^i è un 4-vettore. Si ricordi ancora che lo scalare 4-dim $u_i u^i$ (con $u_i = \gamma(u_i, ic)$) vale $\gamma^2(u^2 - c^2) = -c^2$. (Le posizioni in alto e in basso degli indici ripetuti servono qui soltanto a facilitare la lettura delle formule, perché si opera in uno spazio formalmente euclideo.) Mediante u^i , introduciamo il 2-tensore simmetrico $t^{ik} =: \mu^0 u^i u^k$. Qui μ^0 (densità di massa di quiete \equiv densità di massa classica) è ovviamente un invariante scalare 4-dim, e questo ci assicura che t^{ik} è un 2-tensore 4-dim. Allora $t^{i\kappa} = \mu^0 \gamma^2 u^i u^\kappa = \mu u^i u^\kappa$, $t^{i4} = \mu^0 u^i u^4 = i \mu^0 \gamma^2 u^i c = i \mu u^i c$, $t^{4i} = t^{i4}$, e $t^{44} = \mu^0 u^4 u^4 = -\mu^0 \gamma^2 c^2 = -\mu c^2$. Si consideri ora la divergenza $\partial_k t^{ik}$ (quindi con $x^1 = x$, $x^2 = y$, $x^3 = z$, $x^4 = ict$ e $g_{ik} = \delta_{ik}$). Abbiamo, per $i = \iota$, $\partial_k t^{ik} = \partial_\kappa(\mu u^\iota u^\kappa) + (1/c) \partial_\iota(\mu u^\iota c) = \varphi^\iota$ secondo la (6₁); e per $i = 4$, $\partial_k t^{4k} = (i/c)[\partial_\kappa(\mu c^2 u^\kappa) + \partial_\iota(\mu c^2)] = i\phi/c$ secondo la (8). Basta dunque definire un quadrivettore

f^i uguale a φ^i per $i = 1$ e a $i\phi/c$ per $i = 4$, cioè $f^i =: (\varphi^i, i\phi/c)$, per riassumere le leggi relativistico-speciali del moto di un continuo materiale nella forma

$$(9) \quad \partial_k t^{ik} = f^i.$$

Il 2-tensore simmetrico $t_{(2)}$ si dice **energetico dinamico** (relativistico), ed il vettore f **densità di quadriforza** (relativistica) ²¹.

Tradizionalmente, le quattro componenti indipendenti del tensore t^{ik} per i e/o k uguale a 4 (tenuto conto della sua simmetria), si trascrivono in una notazione diversa. Precisamente, si pone $icg^i =: t^{i4}$, quindi $g^i = \mu u^i$, $is^i/c =: t^{4i}$ ($= t^{i4}$), quindi $s^i = c^2 g^i$, e $-h =: t^{44}$. In questa notazione, le (9) assumono la forma:

$$(10_1) \quad \partial_k t^{ik} + \partial_t g^i = \varphi^i$$

per $i = 1, 2, 3$, e

$$(10_2) \quad \partial_k s^k + \partial_t h = \phi$$

per $i = 4$. Le dieci grandezze indipendenti t^{ik} (sei), g^i (tre) e h (una) sono ovviamente tutte reali, e altrettanto lo sono le quattro φ^i, ϕ .

L'interpretazione fisica delle (10) è abbastanza ovvia. La parte spaziale t^{ik} del tensore energetico $\mu u^i u^k = g^i u^k = g^k u^i$ è un flusso di QdM (di dimensioni $MLT^{-1}L^{-2}T^{-1} = \text{forza/area}$), g^i è una densità di QdM (dimensioni: $ML^{-2}T^{-1}$), $s^k = \mu c^2 u^k = hu^k$ è un flusso di energia (dimensioni: MT^{-3}), h è una densità di energia (dimensioni: $ML^{-1}T^{-2}$, come quelle di t^{ik}), φ^i è una densità di forza ($ML^{-2}T^{-2}$), e ϕ è una densità di potenza ($ML^{-1}T^{-3}$). Quindi la (10₁) è un bilancio, per unità di volume e di tempo, di QdM; e similmente la (10₂) è un bilancio, sempre per unità di volume e di tempo, di energia. Insieme con altri sistemi di equazioni con la stessa struttura (o possibilmente più generale, cioè con un tensore non necessariamente simmetrico in luogo di $t_{(2)}$), le (9) o le (10) sono le equazioni fondamentali della dinamica relativistica dei mezzi continui.

A conclusione di questa appendice, segnaliamo una speciale conseguenza delle (10₁). Moltiplicandole per x^λ abbiamo $(\partial_k t^{ik} + \partial_t g^i)x^\lambda = \partial_k(t^{ik}x^\lambda) - t^{i\lambda} + \partial_t(g^i x^\lambda) = \varphi^i x^\lambda$; sottraendo da questa la sua analoga con lo scambio di i con λ e tenendo conto della simmetria di $t_{(2)}$, seguono le

$$(11) \quad \partial_k(t^{ik}x^\lambda - t^{k\lambda}x^i) + \partial_t(g^i x^\lambda - g^\lambda x^i) = \varphi^i x^\lambda - \varphi^\lambda x^i.$$

È naturale introdurre i 2-tensori 3-dim antisimmetrici $a^{i\lambda} =: g^i x^\lambda - g^\lambda x^i$ e $b^{i\lambda} =: \varphi^i x^\lambda - \varphi^\lambda x^i$. Poiché $t^{ik}x^\lambda - t^{k\lambda}x^i = g^i u^k x^\lambda - g^\lambda u^k x^i \equiv a^{i\lambda} u^k$ concludiamo che

$$(12) \quad \partial_k(a^{i\lambda} u^k) + \partial_t(a^{i\lambda}) = C_u a^{i\lambda} = b^{i\lambda}.$$

²¹ Che $(\varphi^k, i\phi/c)$ sia una densità di quadriforza è confermato calcolando l'invariante $u_k f^k = \gamma(u_k \varphi^k - c\phi/c) = 0$; u e f sono formalmente ortogonali, come devono essere.

Questo è un bilancio, per unità di volume e di tempo, di momento rispetto all'origine (che spesso si dice momento angolare) di QdM, e vale in quanto la parte spaziale di $t_{(2)}$ è simmetrica. Tenuto conto della $d_t\tau = \tau\partial_\kappa u^\kappa$, la (12) si può riscrivere nella forma

$$(13) \quad \tau^{-1}d_t(\tau a^{i\lambda}) = b^{i\lambda}.$$

App. 2. I I PARADOSSI DELLE VELOCITÀ SUPERLUMINALI IN RELATIVITÀ SPECIALE

Le trasformazioni speciali di Lorentz (2.2.1, 2) (o se si preferisce, (2.2.1, 6) o (2.2.1, 6bis)) presuppongono che $\Upsilon^2/c^2 = \text{Th}^2\theta$ sia < 1 . Se infatti questa condizione venisse meno, il fattore di Lorentz $g = (1 - \Upsilon^2/c^2)^{-1/2}$ diventerebbe immaginario (se $\Upsilon^2/c^2 > 1$), oppure le (2.2.1, 2) non potrebbero più scriversi (se $\Upsilon^2/c^2 = 1$) perché g diventerebbe infinito. In entrambi i casi le trasformazioni di Lorentz, e con esse l'intera teoria relativistica speciale, degenererebbero. Questo indica che la velocità relativa tra due riferimenti inerziali deve comunque essere supposta minore di c in valore assoluto.

La predetta condizione $\Upsilon^2/c^2 < 1$ implica anche che un punto materiale non possa muoversi con velocità di modulo $\geq c$ rispetto ad un qualunque riferimento inerziale \mathfrak{S} . Se infatti questo fosse possibile, una rete rigida fatta di tali punti traslanti uniformemente con quella velocità rispetto a \mathfrak{S} , e pensati come orologi normali, costituirebbe a tutti gli effetti un riferimento inerziale “paradosso” nel senso specificato più sopra.

Esaminiamo ora le implicazioni di una velocità v (che per semplicità supporremo costante) “superluminale” (cioè per cui $v^2 > c^2$) non più di un punto materiale, ma di un segnale di qualunque natura che si sposti con velocità v rispetto all'osservatore inerziale \mathfrak{S} , da un punto-istante \mathcal{P} ad un punto-istante \mathcal{Q} . Senza limitazioni di generalità, potremo supporre che tali punti-istanti siano entrambi sull'asse (x), e che le coordinate di \mathcal{Q} differiscano da quelle di \mathcal{P} per $\Delta x > 0$, $\Delta t > 0$. Questo implica che $v = \Delta x/\Delta t$ sia positiva. Nel riferimento inerziale \mathfrak{S}' in moto uniforme rispetto a \mathfrak{S} lungo l'asse (x) con velocità Υ positiva ($e < c$), alla luce delle trasformazioni di Lorentz avremo in particolare

$$(1) \quad \Delta t' = g(\Upsilon/c)(\Delta t - \Upsilon\Delta x/c^2).$$

Se v^2 fosse $> c^2$, potremmo comunque scegliere Υ in modo che $c^2/v^2 < \Upsilon/v$. Poiché $\Delta x = v\Delta t$, in base alla (1) avremmo

$$(2) \quad \Delta t' = g(\Upsilon/c)\Delta t(1 - \Upsilon v/c^2),$$

in cui $\gamma v/c^2 = (\gamma/v)(v^2/c^2)$ sarebbe > 1 ; e dunque, $\Delta t' < 0$. Vale a dire, secondo \mathcal{S}' il segnale arriverebbe in Q *prima* del suo invio da \mathcal{P} .

Si deve concludere che, in relatività speciale, l'esistenza di segnali superluminali implicherebbe quella di riferimenti inerziali, come \mathcal{S} e \mathcal{S}' , in cui il rapporto di causa ed effetto viene rovesciato: in \mathcal{S}' , l'“effetto secondo \mathcal{S}' ” (arrivo del segnale in Q) *precede* la “causa secondo \mathcal{S}' ” (invio del segnale da \mathcal{P}). Se allora si vuole che il principio di causalità (“l'effetto segue la causa”) sia uniformemente osservato da tutti i riferimenti inerziali, è necessario bandire l'esistenza di segnali superluminali di qualunque tipo. Insomma, la consistenza della relatività speciale con il principio di causalità presuppone la *non-esistenza* dei segnali superluminali, secondo quanto l'esperienza (macroscopica) conferma.