Analysis on manifolds

Teo Banica

Department of Mathematics, University of Cergy-Pontoise, F-95000 Cergy-Pontoise, France. teo.banica@gmail.com

2010 *Mathematics Subject Classification.* 58C35

Key words and phrases. Smooth manifold, Integration

Abstract. This is an introduction to analysis on manifolds, with emphasis on integration techniques, and with all needed preliminaries included. We first review the standard calculus methods in $\mathbb{R}^N$, including the use of spherical coordinates, and with a look at calculus over the unit spheres $S^{N-1}_\mathbb{R}\subset\mathbb{R}^N$ too. Then we introduce and study the general smooth manifolds, and discuss calculus on them, notably with the Stokes formula, and with a look into Lie groups and symmetric spaces too. We then go into a similar study for the Riemannian manifolds, notably with various integration results. Finally, we provide an introduction to Lorentz manifolds, and our usual spacetime.

Preface

There are many interesting curves, surfaces and so on, in two or more dimensions, $X \subset \mathbb{R}^N$, and the study of functions on them, $f : X \to \mathbb{R}$, is a key problem. Typically you can think of $X \subset \mathbb{R}^N$ as being the space where your problem lives, coming from various algebraic or analytic constraints, and with the problem being that of minimizing or maximizing this or that function $f : X \to \mathbb{R}$, that you are interested in.

Mathematically, the spaces $X \subset \mathbb{R}^N$ having suitable regularity properties are called manifolds, and the study of functions $f : X \to \mathbb{R}$ is called analysis on manifolds.

The manifolds $X \subset \mathbb{R}^N$ can be of many types, for instance algebraic or analytic, depending on the exact types of constraints they come from. It is also possible to talk about abstract manifolds X, without reference to a surrounding space $\mathbb{R}^N$, by axiomatizing what you know about the manifolds $X \subset \mathbb{R}^N$ that you are interested in.

So, this was for the general idea, regarding analysis on manifolds, with this being basically the study of functions of type $f : X \to \mathbb{R}$. In practice, however, things endlessly ramify, depending on the precise type of question that you are interested in:

– As already mentioned above, the manifolds can be algebraic or analytic.

– Also, they can be real or complex, and even defined over a field F.

– In what regards their regularity, that can be C^0, C^1 and so on, up to C^∞.

– The abstract manifolds X can feature metrics, of various types, or not.

– Also, the manifolds X can have various extra algebraic features.

As you can see, many things going on here, and again, making your way through this jungle normally requires you to have some precise problems, that you are interested in. So that you can choose the type of manifolds that you need, and learn about them.

This book is a basic introduction to analysis on manifolds, taken concrete $X \subset \mathbb{R}^N$ or abstract X, usually assumed to be smooth, C^∞, with emphasis on integration, and with all needed preliminaries included. The book is organized in 4 parts, as follows:

I. We first review the standard calculus methods in $\mathbb{R}^N$, including the use of spherical coordinates, and with a look at calculus over the unit spheres $S^{N-1}_\mathbb{R} \subset \mathbb{R}^N$ too.

II. Then we introduce the smooth manifolds, and discuss calculus on them, notably with the Stokes formula, and with a look into Lie groups and symmetric spaces too.

III. We then go into a similar study for the Riemannian manifolds, by benefiting from the metric structure there, notably with various integration results.

IV. Finally, as an application of our various techniques, we provide an introduction to the Lorentz manifolds, and to our usual, relativistic spacetime.

This book is based on lecture notes from classes that I taught at Toulouse and Cergy, on differential geometry and related topics, and I would like to thank my students. Many thanks as well to my cats, for some help with the relativity theory computations.

Cergy, September 2025
Teo Banica

Contents

Part I

Space, spheres

CHAPTER 1

Vectors, calculus

1a. Linear maps

We will be interested in this book in studying the functions $f:X\to\mathbb R$ and $f:X\to\mathbb C$, with X being a manifold. And with such manifolds X being something similar to $\mathbb R^N$, or to the unit sphere $S^{N-1}_\mathbb R\subset\mathbb R^N$, or to the many curves and surfaces out there, inside $\mathbb R^N$. In short, the manifolds X will generalize all the spaces that we know, where we can do some analysis, and once defined and studied a bit, we will do analysis on them.

Which sounds very good, but we will do this in fact starting from Part II of this book. In the present Part I we will review our knowledge of the usual analysis, on $\mathbb R^N$, with all sorts of formulae that you might know or not, including those making use of spherical coordinates, and with a look at analysis over the unit spheres $S^{N-1}_\mathbb R\subset\mathbb R^N$ too. And in what regards general manifolds X, do not worry, we will discuss them afterwards.

Getting started now, we will be first interested in the functions $f:\mathbb R^N\to\mathbb R$ and $f:\mathbb R^N\to\mathbb C$. And regarding these, in view of $\mathbb C\simeq\mathbb R^2$, it makes sense to look at the general functions of type $f:\mathbb R^N\to\mathbb R^M$. So, here is the question to be solved:

Question 1.1. *What can we say about the functions $f:\mathbb R^N\to\mathbb R^M$, generalizing what we know from usual calculus, regarding the functions $f:\mathbb R\to\mathbb R$?*

In answer now, there are potentially many things that can be investigated, such as continuity, derivatives, higher derivatives, and integrals. However, leaving aside continuity, which looks like something quite standard, when thinking at derivatives we face right away a problem. Recall indeed the main formula from calculus, namely:

$$f(x+t)\simeq f(x)+f'(x)t$$

In our setting, $f:\mathbb R^N\to\mathbb R^M$, both quantities $f(x+t)$ and $f(x)$ are vectors in $\mathbb R^M$, and so we must have $f'(x)t\in\mathbb R^M$. But the variable t being a vector in $\mathbb R^N$, the derivative $f'(x)$ cannot be a number, and must be instead a map of the following type:

$$f'(x):\mathbb R^N\to\mathbb R^M$$

Nevermind. We will clarify this later, but from this perspective, the one-variable derivative appears as the linear map $t\to\lambda t$, with $\lambda=f'(x)$, and this suggests that in general, we can expect the derivative $f'(x):\mathbb R^N\to\mathbb R^M$ to be a linear map too.

Summarizing, we are led to the following reformulation of Question 1.1:

QUESTION 1.2. *The main idea of calculus is that the functions $f:\mathbb R\to\mathbb R$ are locally approximately linear, and in view of this, when looking for generalizations:*

(1) *What can we say about the linear maps $f:\mathbb R^N\to\mathbb R^M$?*
(2) *Then, what can we say about the arbitrary functions $f:\mathbb R^N\to\mathbb R^M$?*

Which sound good, now we have a serious plan, and time to develop it. Regarding the first question, about the linear maps $f:\mathbb R^N\to\mathbb R^M$, let us start with some basic linear algebra. At the beginning, we have the following result, that you surely know:

THEOREM 1.3. *The linear maps $f:\mathbb R^N\to\mathbb R^M$ are in correspondence with the matrices $A\in M_{M\times N}(\mathbb R)$, with the linear map associated to such a matrix being*

$$f(x)=Ax$$

and with the matrix associated to a linear map being $A_{ij}=<f(e_j),e_i>$.

PROOF. The first assertion is clear, because a linear map $f:\mathbb R^N\to\mathbb R^M$ must send a vector $x\in\mathbb R^N$ to a certain vector $f(x)\in\mathbb R^M$, all whose components are linear combinations of the components of x. Thus, we can write, for certain real numbers $A_{ij}\in\mathbb R$:

$$f\begin{pmatrix}x_1\\ \vdots\\ \vdots\\ x_N\end{pmatrix}=\begin{pmatrix}A_{11}x_1+\ldots+A_{1N}x_N\\ \vdots\\ \vdots\\ A_{M1}x_1+\ldots+A_{MN}x_N\end{pmatrix}$$

Now the parameters $A_{ij}\in\mathbb R$ can be regarded as being the entries of a certain matrix $A\in M_{M\times N}(\mathbb R)$, and with the usual convention for matrix multiplication, we have:

$$f(x)=Ax$$

Regarding the second assertion, with $f(x)=Ax$ as above, if we denote by $e_1,\ldots,e_N$ the standard basis of $\mathbb R^N$, then we have the following formula:

$$f(e_j)=\begin{pmatrix}A_{1j}\\ \vdots\\ \vdots\\ A_{Mj}\end{pmatrix}$$

But this gives the second formula, $<f(e_j),e_i>=A_{ij}$, as desired. $\square$

In order to reach now to sharper results, we will restrict the attention to the linear maps $f:\mathbb R^N\to\mathbb R^N$, which amounts in looking at the square matrices $A\in M_N(\mathbb R)$. And in what regards these latter matrices, we first have the following result:

PROPOSITION 1.4. *Given a matrix $A \in M_N(\mathbb{R})$, let us call $v \in \mathbb{R}^N$ an eigenvector, with corresponding eigenvalue λ, when A multiplies by λ in the direction of v:*

$$Av = \lambda v$$

Then, in case where $\mathbb{R}^N$ has a basis $v_1, \ldots, v_N$ formed by eigenvectors of A, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_N$, in this new basis A becomes diagonal, as follows:

$$A \sim \begin{pmatrix}\lambda_1&&\\&\ddots&\\&&\lambda_N\end{pmatrix}$$

Equivalently, if we denote by $D = diag(\lambda_1, \ldots, \lambda_N)$ the above diagonal matrix, and by $P = [v_1 \ldots v_N]$ the square matrix formed by the eigenvectors of A, we have:

$$A = PDP^{-1}$$

In this case we say that the matrix A is diagonalizable.

PROOF. This is something which is clear, the idea being as follows:

(1) The first assertion is clear, because the matrix which multiplies each basis element v_i by a number λ_i is precisely the diagonal matrix $D = diag(\lambda_1, \ldots, \lambda_N)$.

(2) The second assertion follows from the first one, by changing the basis. We can prove this by a direct computation as well, because we have $Pe_i = v_i$, and so:

$$PDP^{-1}v_i = PDe_i = P\lambda_i e_i = \lambda_i Pe_i = \lambda_i v_i$$

Thus, the matrices A and PDP^{-1} coincide, as stated. □

In general, in order to study the diagonalization problem, the idea is that the eigenvectors can be grouped into linear spaces, called eigenspaces, as follows:

PROPOSITION 1.5. *Let $A \in M_N(\mathbb{R})$, and for any eigenvalue $\lambda \in \mathbb{R}$ define the corresponding eigenspace as being the vector space formed by the corresponding eigenvectors:*

$$E_\lambda = \left\{ v \in \mathbb{R}^N \Big| Av = \lambda v \right\}$$

These eigenspaces E_λ are then in a direct sum position, in the sense that given vectors $v_1 \in E_{\lambda_1}, \ldots, v_k \in E_{\lambda_k}$ corresponding to different eigenvalues $\lambda_1, \ldots, \lambda_k$, we have:

$$\sum_i c_i v_i = 0 \implies c_i = 0$$

In particular we have the following estimate, with sum over all the eigenvalues,

$$\sum_\lambda \dim(E_\lambda) \leq N$$

and our matrix is diagonalizable precisely when we have equality.

Proof. We prove the first assertion by recurrence on $k \in \mathbb{N}$. Assume by contradiction that we have a formula as follows, with the scalars $c_1, \ldots, c_k$ being not all zero:

$$c_1 v_1 + \ldots + c_k v_k = 0$$

By dividing by one of these scalars, we can assume that our formula is:

$$v_k = c_1 v_1 + \ldots + c_{k-1} v_{k-1}$$

Now let us apply A to this vector. We obtain in this way the following equality:

$$\lambda_k (c_1 v_1 + \ldots + c_{k-1} v_{k-1}) = c_1 \lambda_1 v_1 + \ldots + c_{k-1} \lambda_{k-1} v_{k-1}$$

On the other hand, we know by recurrence that the vectors $v_1, \ldots, v_{k-1}$ must be linearly independent. Thus, the coefficients must be equal, at left and at right:

$$\lambda_k c_1 = c_1 \lambda_1 \quad , \quad \ldots \quad , \quad \lambda_k c_{k-1} = c_{k-1} \lambda_{k-1}$$

Now since at least one of the numbers c_i must be nonzero, we obtain $\lambda_k = \lambda_i$, which is a contradiction. Thus, first assertion proved, and the second assertion follows it. □

In order to reach now to more advanced results, based on the above, we can use the characteristic polynomial, which appears in the following way:

Proposition 1.6. *Given a matrix $A \in M_N(\mathbb{R})$, consider its characteristic polynomial:*

$$P(x) = \det(A - x 1_N)$$

The eigenvalues of A are then the roots of P. Also, we have the inequality

$$\dim(E_\lambda) \leq m_\lambda$$

where m_λ is the multiplicity of λ, as root of P.

Proof. The first assertion follows from the following computation, using the fact that a linear map is bijective when the determinant of the associated matrix is nonzero:

$$\begin{aligned} \exists v, Av = \lambda v &\iff \exists v, (A - \lambda 1_N) v = 0 \\ &\iff \det(A - \lambda 1_N) = 0 \end{aligned}$$

Regarding now the second assertion, given an eigenvalue λ of our matrix A, consider the dimension $d_\lambda = \dim(E_\lambda)$ of the corresponding eigenspace. By changing the basis of $\mathbb{R}^N$, as for the eigenspace E_λ to be spanned by the first d_λ basis elements, our matrix becomes as follows, with B being a certain smaller matrix:

$$A \sim \begin{pmatrix} \lambda 1_{d_\lambda} & 0 \\ 0 & B \end{pmatrix}$$

We conclude that the characteristic polynomial of A is of the following form:

$$P_A = P_{\lambda 1_{d_\lambda}} P_B = (\lambda - x)^{d_\lambda} P_B$$

Thus the multiplicity m_λ of our eigenvalue λ, as a root of P, satisfies $m_\lambda \geq d_\lambda$, and this leads to the conclusion in the statement. □

It is convenient now to regard $A \in M_N(\mathbb{R})$ as a complex matrix, $A \in M_N(\mathbb{C})$, as for its characteristic polynomial P to have roots. We are led in this way to:

THEOREM 1.7. *Given a matrix $A \in M_N(\mathbb{C})$, consider its characteristic polynomial*

$$P(X) = \det(A - X1_N)$$

then factorize this polynomial, by computing the complex roots, with multiplicities,

$$P(X) = (-1)^N (X - \lambda_1)^{n_1} \ldots (X - \lambda_k)^{n_k}$$

and finally compute the corresponding eigenspaces, for each eigenvalue found:

$$E_i = \left\{ v \in \mathbb{C}^N \middle| Av = \lambda_i v \right\}$$

The dimensions of these eigenspaces satisfy then the following inequalities,

$$\dim(E_i) \leq n_i$$

and A is diagonalizable precisely when we have equality for any i.

PROOF. This follows by combining Propositions 1.5 and 1.6, or rather their complex version, whose proofs are identical to those in the real case. Indeed, by summing the inequalities $\dim(E_\lambda) \leq m_\lambda$ from Proposition 1.6, we obtain an inequality as follows:

$$\sum_\lambda \dim(E_\lambda) \leq \sum_\lambda m_\lambda \leq N$$

On the other hand, we know from Proposition 1.5 that our matrix is diagonalizable when we have global equality. Thus, we are led to the conclusion in the statement. □

In practice, diagonalizing a matrix remains something quite complicated. Let us record as well a useful algorithmic version of the above result, as follows:

THEOREM 1.8. *The square matrices $A \in M_N(\mathbb{C})$ can be diagonalized as follows:*

(1) *Compute the characteristic polynomial.*
(2) *Factorize the characteristic polynomial.*
(3) *Compute the eigenvectors, for each eigenvalue found.*
(4) *If there are no N eigenvectors, A is not diagonalizable.*
(5) *Otherwise, A is diagonalizable, $A = PDP^{-1}$.*

PROOF. This is an informal reformulation of Theorem 1.7, with (4) referring to the total number of linearly independent eigenvectors found in (3), and with $A = PDP^{-1}$ in (5) being the usual diagonalization formula, with P, D being as before. □

As an illustration for this, which is a must-know computation, we have:

PROPOSITION 1.9. *The rotation of angle $t \in \mathbb{R}$ in the plane diagonalizes as:*

$$\begin{pmatrix}\cos t & -\sin t\\ \sin t & \cos t\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1 & 1\\ i & -i\end{pmatrix}\begin{pmatrix}e^{-it} & 0\\ 0 & e^{it}\end{pmatrix}\begin{pmatrix}1 & -i\\ 1 & i\end{pmatrix}$$

Over the reals this is impossible, unless $t = 0, \pi$, where the rotation is diagonal.

PROOF. Observe first that, as indicated, unlike we are in the case $t = 0, \pi$, where our rotation is $\pm 1_2$, our rotation is a "true" rotation, having no eigenvectors in the plane. Fortunately the complex numbers come to the rescue, via the following computation:

$$\begin{pmatrix}\cos t & -\sin t\\ \sin t & \cos t\end{pmatrix}\begin{pmatrix}1\\ i\end{pmatrix} = \begin{pmatrix}\cos t - i\sin t\\ i\cos t + \sin t\end{pmatrix} = e^{-it}\begin{pmatrix}1\\ i\end{pmatrix}$$

We have as well a second complex eigenvector, coming from:

$$\begin{pmatrix}\cos t & -\sin t\\ \sin t & \cos t\end{pmatrix}\begin{pmatrix}1\\ -i\end{pmatrix} = \begin{pmatrix}\cos t + i\sin t\\ -i\cos t + \sin t\end{pmatrix} = e^{it}\begin{pmatrix}1\\ -i\end{pmatrix}$$

Thus, we are led to the conclusion in the statement. □

As another basic illustration, in N dimensions, we have the following result:

PROPOSITION 1.10. *The all-one matrix diagonalizes as follows,*

$$\begin{pmatrix}1 & \ldots & 1\\ \vdots & & \vdots\\ 1 & \ldots & 1\end{pmatrix} = \frac{1}{N}F_N\begin{pmatrix}N & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0\end{pmatrix}F_N^*$$

with $F_N = (w^{ij})_{ij}$ with $w = e^{2\pi i/N}$ being the Fourier matrix.

PROOF. The all-one matrix being N times the projection on the all-one vector, the diagonal form is the one in the statement. In order to find now the explicit diagonalization formula, with passage matrix and its inverse, we must solve the following equation:

$$x_1 + \ldots + x_N = 0$$

And this is not an easy task, if we want a nice basis for the space of solutions. Fortunately, the complex numbers come to the rescue, via the following formula:

$$\sum_{k=0}^{N-1} w^{ks} = N\delta_{N|s}$$

But this leads, after some thinking, to the conclusion in the statement. □

Getting back now to general theory, given a complex matrix $A \in M_N(\mathbb{C})$, let us construct its adjoint matrix $A^* \in M_N(\mathbb{C})$ by the following formula:

$$(A^*)_{ij} = \overline{A_{ji}}$$

With this convention, we have the following advanced result:

THEOREM 1.11. *The following happen, regarding the matrices $A \in M_N(\mathbb{C})$:*
1. *The self-adjoint matrices, $A = A^*$, are diagonalizable.*
2. *The unitary matrices, $A^* = A^{-1}$, are diagonalizable too.*
3. *In fact, the normal matrices, $AA^* = A^*A$, are diagonalizable.*

Moreover, the commuting families of normal matrices are jointly diagonalizable.

PROOF. This is something more advanced, the idea being as follows:

(1) This generalizes the fact, that you probably know from basic linear algebra, that any symmetric matrix $A \in M_N(\mathbb{R})$ is diagonalizable. And with the proof being similar.

(2) This is certainly something complex number specific, for instance in view of Proposition 1.9. However, the proof is quite routine, coming as a variation of (1).

(3) This is a joint generalization of (1) and (2), the matrices appearing there being normal, and for the proof here, we refer to any advanced linear algebra book.

(4) As for the last assertion, again we refer here to any advanced linear algebra book. Alternatively, you can consult any introductory operator theory book. □

As a last general linear algebra result, that you should know too, we have:

THEOREM 1.12. *Given a matrix $A \in M_N(\mathbb{C})$, the following happen:*
1. *A^*A being positive, we can extract its square root $|A| = \sqrt{A^*A}$.*
2. *When A is invertible, we have $A = U|A|$, with U being a unitary.*
3. *In general, we still have $A = U|A|$, with U being a partial isometry.*

PROOF. Again, this is something more advanced, the idea being as follows:

(1) The matrix A^*A being self-adjoint, and with positive eigenvalues, with this coming from $< A^*Ax, x >= ||Ax||^2$, our claim follows from Theorem 1.11 (1), by diagonalizing this matrix, and then taking the square roots of all eigenvalues.

(2) According to our definition of the modulus, namely $|A| = \sqrt{A^*A}$, we have:

$$< |A|x, |A|y >=< x, |A|^2y >=< x, A^*Ay >=< Ax, Ay >$$

We conclude that the following linear map is well-defined, and isometric:

$$U : Im|A| \to Im(A) \quad , \quad |A|x \to Ax$$

But with A assumed to be invertible, U is a unitary, and $A = U|A|$, as desired.

(3) Getting back to the linear isometric map U constructed in (2), we can extend this map into a partial isometry $U : \mathbb{C}^N \to \mathbb{C}^N$, in a straightforward way, by setting:

$$Ux = 0 \quad , \quad \forall x \in Im|A|^\perp$$

And the point is that, with this convention, we have again $A = U|A|$, as desired. □

Back now to Theorem 1.7 and basics, at the level of examples of diagonalizable matrices, we have the following result, providing us with the "generic" examples:

THEOREM 1.13. *For a matrix $A \in M_N(\mathbb{C})$ the following conditions are equivalent,*

(1) *The eigenvalues are different, $\lambda_i \neq \lambda_j$,*
(2) *The characteristic polynomial P has simple roots,*
(3) *The characteristic polynomial satisfies $(P, P') = 1$,*
(4) *The resultant of P, P' is nonzero, $R(P, P') \neq 0$,*
(5) *The discriminant of P is nonzero, $\Delta(P) \neq 0$,*

and in this case, the matrix is diagonalizable.

PROOF. The last assertion holds indeed, due to Theorem 1.7. As for the equivalences in the statement, these are all standard, coming from the theory of the resultant R and discriminant Δ, that you can find if needed in any advanced linear algebra book. □

As already mentioned above, one can prove that the matrices having distinct eigenvalues are "generic", so that Theorem 1.13 basically captures the whole situation. We have in fact the following collection of density results, which are quite advanced:

THEOREM 1.14. *The following happen, inside $M_N(\mathbb{C})$:*

(1) *The invertible matrices are dense.*
(2) *The matrices having distinct eigenvalues are dense.*
(3) *The diagonalizable matrices are dense.*

PROOF. These are quite advanced results, which can be proved as follows:

(1) This is clear, intuitively speaking, because the invertible matrices are given by the condition $\det A \neq 0$. Thus, the set formed by these matrices appears as the complement of the hypersurface $\det A = 0$, and so must be dense inside $M_N(\mathbb{C})$, as claimed.

(2) Here we can use a similar argument, this time by saying that the set formed by the matrices having distinct eigenvalues appears as the complement of the hypersurface given by $\Delta(P_A) = 0$, and so must be dense inside $M_N(\mathbb{C})$, as claimed.

(3) This follows from (2), via the fact that the matrices having distinct eigenvalues are diagonalizable, that we know from Theorem 1.13. There are of course some other proofs as well, for instance by putting the matrix in Jordan form. □

As an application of the above results, and of our methods in general, we have:

PROPOSITION 1.15. *The following happen:*

(1) *We have $P_{AB} = P_{BA}$, for any two matrices $A, B \in M_N(\mathbb{C})$.*
(2) *AB, BA have the same eigenvalues, with the same multiplicities.*
(3) *If A has eigenvalues $\lambda_1, \ldots, \lambda_N$, then $f(A)$ has eigenvalues $f(\lambda_1), \ldots, f(\lambda_N)$.*

PROOF. These results can be deduced by using Theorem 1.14, as follows:

(1) It follows from definitions that the characteristic polynomial is invariant under conjugation, $P_C = P_{ACA^{-1}}$. Now observe that, when A is invertible, we have:

$$AB = A(BA)A^{-1}$$

Thus, the result holds when A is invertible, and by density, we have it everywhere.

(2) This is a reformulation of (1), which is quite hard to prove with bare hands, coming from the fact that P encodes the eigenvalues, with multiplicities.

(3) This is something more advanced, clear for the diagonal or diagonalizable matrices, and then for all matrices, provided that f has suitable regularity properties. □

Let us go back now to the main problem raised by the diagonalization procedure, namely the computation of the roots of characteristic polynomials. We have here:

THEOREM 1.16. *The complex eigenvalues of a matrix $A \in M_N(\mathbb{C})$, counted with multiplicities, have the following properties:*

(1) *Their sum is the trace.*
(2) *Their product is the determinant.*

PROOF. Consider indeed the characteristic polynomial P of the matrix:

$$\begin{aligned} P(X) &= \det(A - X1_N) \\ &= (-1)^N X^N + (-1)^{N-1} Tr(A) X^{N-1} + \ldots + \det(A) \end{aligned}$$

We can factorize this polynomial, by using its N complex roots, and we obtain:

$$\begin{aligned} P(X) &= (-1)^N (X - \lambda_1) \ldots (X - \lambda_N) \\ &= (-1)^N X^N + (-1)^{N-1} \left(\sum_i \lambda_i \right) X^{N-1} + \ldots + \prod_i \lambda_i \end{aligned}$$

Thus, we are led to the conclusion in the statement. □

Regarding now the intermediate terms, we have here:

THEOREM 1.17. *Assume that $A \in M_N(\mathbb{C})$ has eigenvalues $\lambda_1, \ldots, \lambda_N \in \mathbb{C}$, counted with multiplicities. The basic symmetric functions of these eigenvalues, namely*

$$c_k = \sum_{i_1 < \ldots < i_k} \lambda_{i_1} \ldots \lambda_{i_k}$$

are then given by the fact that the characteristic polynomial of the matrix is

$$P(X) = (-1)^N \sum_{k=0}^{N} (-1)^k c_k X^k$$

and all symmetric functions of the eigenvalues are polynomials in these coefficients c_k.

Proof. These results, which are all very standard, can be proved indeed by doing some algebraic study, that we will leave here as an instructive exercise. □

And with this, good news, done with the linear algebra that we should now. We have now a good understanding of the linear maps $f:\mathbb R^N\to\mathbb R^M$, and we will make a heavy use of this material, when investigating the arbitrary functions $f:\mathbb R^N\to\mathbb R^M$.

1b. Continuity basics

Let us turn now to the arbitrary functions $f:\mathbb R^N\to\mathbb R^M$. As a first job here, we would like to talk about the continuity of such functions, and extend to this setting the basic results from one-variable calculus. For this purpose, it is most convenient to use the advanced approach to continuity, using open and closed sets. Let us start with:

Proposition 1.18. *We can talk about open and closed sets in $\mathbb R^N$, in the obvious way, exactly as we do it in $\mathbb R$, and the following happen:*
1. *Open balls are open, closed balls are closed.*
2. *Union of open sets is open, intersection of closed sets is closed.*
3. *Finite intersection of open sets is open, finite union of closed sets is closed.*
4. *The open sets are exactly the complements of closed sets.*

Proof. This is obviously something quite informal, the idea being as follows:

(1) This is something which is obvious.

(2) Again, something clear, and a good exercise for you.

(3) Another good exercise, and think at some related counterexamples too.

(4) Exercise too, provided that you didn't use this as definition, for open or closed. □

Getting now to the functions, we have the following result about them:

Theorem 1.19. *Given a function $f:\mathbb R^N\to\mathbb R^M$, the following are equivalent:*
1. *f is continuous.*
2. *If O is open, then $f^{-1}(O)$ is open.*
3. *If C is closed, then $f^{-1}(C)$ is closed.*

Proof. This is indeed something very standard, with (1) $\iff$ (2) coming from definitions, and with (2) $\iff$ (3) coming from Proposition 1.18 (4). □

Regarding now the compact and connected sets, their basic theory is as follows:

Proposition 1.20. *We can talk about compact and connected sets in $\mathbb R^N$, in the obvious way, exactly as we do it in $\mathbb R$, and the following happen:*
1. *Compact is the same as being closed and bounded.*
2. *Convex $\implies$ path connected $\implies$ connected.*

PROOF. This is something more subtle, the idea being as follows:

(1) Let us call indeed $K \subset \mathbb{R}^N$ compact when any open cover $K \subset \cup_i O_i$ has a finite subcover. It is clear then, by using suitable covers, exactly as in the 1-dimensional case, that K must be closed, and bounded as well. As for the converse, this follows again as in the 1-dimensional case, with the main ingredient here, which again can be proved by using a suitable cover, being the fact that the unit cube is indeed compact.

(2) Let us call indeed $E \subset \mathbb{R}^N$ connected when it is not possible to split it into two parts, that is, when it is not possible to have $E \subset A \sqcup B$, with A, B open. It is then clear that if E is path connected, then it must be connected, because when assuming $E \subset A \sqcup B$ with A, B open, we cannot have any path from $a \in A$ to $b \in B$. □

Getting now to the functions, we have the following result about them:

THEOREM 1.21. *Assuming that $f : \mathbb{R}^N \to \mathbb{R}^M$ is continuous:*
(1) *If K is compact, then $f(K)$ is compact.*
(2) *If E is connected, then $f(E)$ is connected.*

PROOF. This is indeed something very standard, with (1) coming from the definition of compactness, and with (2) coming from the definition of connectedness. □

Getting now to what we really wanted to say about continuous functions, intermediate value theorem, this is (2) above, so let us have this highlighted, as follows:

THEOREM 1.22 (Intermediate values). *Assuming that a function*

$$f : X \to \mathbb{R}^M$$

with $X \subset \mathbb{R}^N$ is continuous, if the domain X is connected, so is its image $f(X)$.

PROOF. We have already stated this in Theorem 1.21 (2), but let us see now how the detailed proof goes as well. Assume by contradiction that $f(X)$ is not connected, which in practice means that we can find two disjoint open sets A, B such that:

$$f(X) \subset A \sqcup B$$

By taking inverse images, we obtain from this a disjoint union as follows:

$$X \subset f^{-1}(A) \sqcup f^{-1}(B)$$

Now since inverse image of an open set is open, wich this being something which is clear from definitions, both the above sets $f^{-1}(A)$ and $f^{-1}(B)$ are open. Thus we have managed to split X into two parts, contradicting its connectivity, as desired. □

At a more advanced level now, we have the following key result:

THEOREM 1.23 (Heine, Cantor). *Any continuous real function*

$$f : X \to \mathbb{R}$$

with $X \subset \mathbb{R}^N$ compact is automatically uniformly continuous.

PROOF. This is something quite standard, the idea being as follows:

(1) Given $\varepsilon > 0$, for any $x \in X$ we know that we have a $\delta_x > 0$ such that:

$$||x - y|| < \delta_x \implies |f(x) - f(y)| < \frac{\varepsilon}{2}$$

So, consider the following open balls, centered at the various points $x \in X$:

$$U_x = B_x\left(\frac{\delta_x}{2}\right)$$

These open balls then obviously cover X, in the sense that we have:

$$X \subset \bigcup_{x \in X} U_x$$

(2) But, we know that X is compact. So, consider a finite subcover of this cover:

$$X \subset \bigcup_i U_{x_i}$$

With this done, consider the following number, which is strictly positive:

$$\delta = \min_i \frac{\delta_{x_i}}{2}$$

Now assume $||x - y|| < \delta$, and pick i such that $x \in U_{x_i}$. By the triangle inequality we have then $||x_i - y|| < \delta_{x_i}$, which shows that we have $y \in U_{x_i}$ as well. But by applying now f, this gives as desired $|f(x) - f(y)| < \varepsilon$, again via the triangle inequality. □

So long for the continuity basics. We will be back with more, when needed.

1c. First derivatives

Getting now to more advanced analysis, let us discuss the differentiability in several variables. At order 1, the situation is quite simple, as follows:

THEOREM 1.24. *The derivative of a function $f : \mathbb{R}^N \to \mathbb{R}^M$, making the formula*

$$f(x + t) \simeq f(x) + f'(x)t$$

work, must be the matrix of partial derivatives at x, namely

$$f'(x) = \left(\frac{df_i}{dx_j}(x)\right)_{ij} \in M_{M\times N}(\mathbb{R})$$

acting on the vectors $t \in \mathbb{R}^N$ by usual multiplication.

PROOF. As a first observation, the formula in the statement makes sense indeed, as an equality, or rather approximation, of vectors in $\mathbb{R}^M$, as follows:

$$f\begin{pmatrix}x_1+t_1\\ \vdots\\ x_N+t_N\end{pmatrix}\simeq f\begin{pmatrix}x_1\\ \vdots\\ x_N\end{pmatrix}+\begin{pmatrix}\frac{df_1}{dx_1}(x)&\ldots&\frac{df_1}{dx_N}(x)\\ \vdots&&\vdots\\ \frac{df_M}{dx_1}(x)&\ldots&\frac{df_M}{dx_N}(x)\end{pmatrix}\begin{pmatrix}t_1\\ \vdots\\ t_N\end{pmatrix}$$

In order to prove now this formula, we can proceed by recurrence, as follows:

(1) First of all, at $N=M=1$ what we have is a usual 1-variable function $f:\mathbb{R}\to\mathbb{R}$, and the formula in the statement is something that we know well, namely:

$$f(x+t)\simeq f(x)+f'(x)t$$

(2) Let us discuss now the case $N=2,M=1$. Here what we have is a function $f:\mathbb{R}^2\to\mathbb{R}$, and by using twice the basic approximation result from (1), we obtain:

$$\begin{aligned}f\begin{pmatrix}x_1+t_1\\ x_2+t_2\end{pmatrix}&\simeq f\begin{pmatrix}x_1+t_1\\ x_2\end{pmatrix}+\frac{df}{dx_2}(x)t_2\\ &\simeq f\begin{pmatrix}x_1\\ x_2\end{pmatrix}+\frac{df}{dx_1}(x)t_1+\frac{df}{dx_2}(x)t_2\\ &= f\begin{pmatrix}x_1\\ x_2\end{pmatrix}+\begin{pmatrix}\frac{df}{dx_1}(x)&\frac{df}{dx_2}(x)\end{pmatrix}\begin{pmatrix}t_1\\ t_2\end{pmatrix}\end{aligned}$$

(3) More generally, we can deal in this way with the case $N\in\mathbb{N},M=1$, by recurrence. But this gives the result in the general case $N,M\in\mathbb{N}$ too. Indeed, let us write:

$$f=\begin{pmatrix}f_1\\ \vdots\\ f_M\end{pmatrix}$$

We can apply our result to each of the components $f_i:\mathbb{R}^N\to\mathbb{R}$, and we get:

$$f_i\begin{pmatrix}x_1+t_1\\ \vdots\\ x_N+t_N\end{pmatrix}\simeq f_i\begin{pmatrix}x_1\\ \vdots\\ x_N\end{pmatrix}+\begin{pmatrix}\frac{df_i}{dx_1}(x)&\ldots&\frac{df_i}{dx_N}(x)\end{pmatrix}\begin{pmatrix}t_1\\ \vdots\\ t_N\end{pmatrix}$$

But this is precisely what we want, at the level of the global map $f:\mathbb{R}^N\to\mathbb{R}^M$. $\square$

As a technical complement to the above result, we have:

THEOREM 1.25. *For a function $f:X\to\mathbb{R}^M$, with $X\subset\mathbb{R}^N$, the following conditions are equivalent, and in this case we say that f is continuously differentiable:*

(1) *f is differentiable, and the map $x\to f'(x)$ is continuous.*
(2) *f has partial derivatives, which are continuous with respect to $x\in X$.*

If these conditions are satisfied, $f'(x)$ is the matrix fomed by the partial derivatives at x.

Proof. We already know, from Theorem 1.24, that the last assertion holds. Regarding now the proof of the equivalence, this goes as follows:

(1) $\Longrightarrow$ (2) Assuming that f is differentiable, we know from Theorem 1.24 that $f'(x)$ is the matrix fomed by the partial derivatives at x. Thus, for any $x, y \in X$:

$$\frac{df_i}{dx_j}(x) - \frac{df_i}{dx_j}(y) = f'(x)_{ij} - f'(y)_{ij}$$

By applying now the absolute value, we obtain from this the following estimate:

$$\begin{aligned}\left|\frac{df_i}{dx_j}(x) - \frac{df_i}{dx_j}(y)\right| &= |f'(x)_{ij} - f'(y)_{ij}| \\ &= |(f'(x) - f'(y))_{ij}| \\ &\leq ||f'(x) - f'(y)||\end{aligned}$$

But this gives the result, because if the map $x \to f'(x)$ is assumed to be continuous, then the partial derivatives follow to be continuous with respect to $x \in X$.

(2) $\Longrightarrow$ (1) This is something more technical. For simplicity, let us assume $M = 1$, the proof in general being similar. Given $x \in X$ and $\varepsilon > 0$, let us pick $r > 0$ such that the ball $B = B_x(r)$ belongs to X, and such that the following happens, over B:

$$\left|\frac{df}{dx_j}(x) - \frac{df}{dx_j}(y)\right| < \frac{\varepsilon}{N}$$

Our claim is that, with this choice made, we have the following estimate, for any $t \in \mathbb{R}^N$ satisfying $||t|| < r$, with A being the vector of partial derivatives at x:

$$|f(x+t) - f(x) - At| \leq \varepsilon ||t||$$

In order to prove this claim, the idea will be that of suitably applying the mean value theorem, over the N directions of $\mathbb{R}^N$. Indeed, consider the following vectors:

$$t^{(k)} = \begin{pmatrix} t_1 \\ \vdots \\ t_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

In terms of these vectors, we have the following formula:

$$f(x+t) - f(x) = \sum_{j=1}^N f(x + t^{(j)}) - f(x + t^{(j-1)})$$

Also, the mean value theorem gives a formula as follows, with $s_j \in [0,1]$:

$$f(x+t^{(j)}) - f(x+t^{(j-1)}) = \frac{df}{dx_j}(x + s_j t^{(j)} + (1-s_j)t^{(j-1)}) \cdot t_j$$

But, according to our assumption on $r > 0$ from the beginning, the derivative on the right differs from $\frac{df}{dx_j}(x)$ by something which is smaller than ε/N:

$$\left|\frac{df}{dx_j}(x + s_j t^{(j)} + (1-s_j)t^{(j-1)}) - \frac{df}{dx_j}(x)\right| < \frac{\varepsilon}{N}$$

Now by putting everything together, we obtain the following estimate:

$$\begin{aligned}
|f(x+t) - f(x) - At| &= \left|\sum_{j=1}^N f(x+t^{(j)}) - f(x+t^{(j-1)}) - \frac{df}{dx_j}(x) \cdot t_j\right| \\
&\leq \sum_{j=1}^N \left|f(x+t^{(j)}) - f(x+t^{(j-1)}) - \frac{df}{dx_j}(x) \cdot t_j\right| \\
&= \sum_{j=1}^N \left|\frac{df}{dx_j}(x + s_j t^{(j)} + (1-s_j)t^{(j-1)}) \cdot t_j - \frac{df}{dx_j}(x) \cdot t_j\right| \\
&= \sum_{j=1}^N \left|\frac{df}{dx_j}(x + s_j t^{(j)} + (1-s_j)t^{(j-1)}) - \frac{df}{dx_j}(x)\right| \cdot |t_j| \\
&\leq \sum_{j=1}^N \frac{\varepsilon}{N} \cdot |t_j| \\
&\leq \varepsilon||t||
\end{aligned}$$

Thus we have proved our claim, and this gives the result. □

Moving on, with this done, our task will be that of extending to several variables our basic results from one-variable calculus. As a standard result here, we have:

THEOREM 1.26. *We have the chain derivative formula*

$$(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$$

as an equality of matrices.

PROOF. This is something standard in one variable, and in several variables the proof is similar, by using the abstract notion of derivative coming from Theorem 1.24. To be more precise, consider a composition of functions, as follows:

$$f : \mathbb{R}^N \to \mathbb{R}^M \quad , \quad g : \mathbb{R}^K \to \mathbb{R}^N \quad , \quad f \circ g : \mathbb{R}^K \to \mathbb{R}^M$$

According to Theorem 1.24, the derivatives of these functions are certain linear maps, corresponding to certain rectangular matrices, as follows:

$$f'(g(x)) \in M_{M\times N}(\mathbb{R}) \quad , \quad g'(x) \in M_{N\times K}(\mathbb{R}) \qquad (f\circ g)'(x) \in M_{M\times K}(\mathbb{R})$$

Thus, our formula makes sense indeed. As for proof, this comes from:

$$\begin{aligned}(f\circ g)(x+t) &= f(g(x+t))\\ &\simeq f(g(x)+g'(x)t)\\ &\simeq f(g(x))+f'(g(x))g'(x)t\end{aligned}$$

Thus, we are led to the conclusion in the statement. □

As a standard application of the above chain rule differentiation result, generalizing some basic things that we know from one-variable calculus, we have:

THEOREM 1.27. *Assuming that $f : X \to \mathbb{R}^M$ is differentiable, with $X \subset \mathbb{R}^N$ being convex, we have the estimate*

$$||f(x)-f(y)|| \leq M||x-y||$$

for any $x, y \in X$, where the quantity on the right is given by:

$$M = \sup_{x\in X} ||f'(x)||$$

Moreover, this estimate can be sharp, for instance for the linear functions.

PROOF. This is something quite tricky, which in several variables cannot be proved with bare hands. However, we can get it by using our chain derivative formula. Consider indeed the path $\gamma : [0,1] \to \mathbb{R}^M$ given by the following formula:

$$\gamma(t) = tx + (1-t)y$$

Now let us set $g(t) = f(\gamma(t))$. We have then, according to the chain rule formula:

$$\begin{aligned}g'(t) &= f'(\gamma(t))\gamma'(t)\\ &= f'(\gamma(t))(x-y)\end{aligned}$$

But this gives the following estimate, with $M > 0$ being as in the statement:

$$\begin{aligned}|g'(t)| &\leq ||f'(\gamma(t))||\cdot||x-y||\\ &\leq M||x-y||\end{aligned}$$

Now by using one-variable results that we know, we obtain from this:

$$||g(1)-g(0)|| \leq ||M||\cdot||x-y||$$

But since we have $g(1) = f(x), g(0) = f(y)$, this gives the formula in the statement. Finally, the last assertion is clear. □

As a conclusion to this, we have extended to the case of vector functions most of what we know about the one-variable functions, at the general level, of basic calculus.

1d. Second derivatives

Moving on, we can talk as well about higher derivatives, simply by performing the operation of taking derivatives recursively. As a first result here, we have:

THEOREM 1.28. *The double derivatives of a function $f:\mathbb{R}^2\to\mathbb{R}$ satisfy*

$$\frac{d^2f}{dxdy}=\frac{d^2f}{dydx}$$

called Clairaut formula.

PROOF. This is something very standard, the idea being as follows:

(1) Before pulling out a formal proof, as an intuitive justification for our formula, let us consider a product of power functions, $f(z)=x^py^q$. We have then:

$$\frac{d^2f}{dxdy}=\frac{d}{dx}\left(\frac{dx^py^q}{dy}\right)=\frac{d}{dx}\left(qx^py^{q-1}\right)=pqx^{p-1}y^{q-1}$$

$$\frac{d^2f}{dydx}=\frac{d}{dy}\left(\frac{dx^py^q}{dx}\right)=\frac{d}{dy}\left(px^{p-1}y^q\right)=pqx^{p-1}y^{q-1}$$

Next, let us consider a linear combination of power functions, $f(z)=\sum_{pq}c_{pq}x^py^q$, which can be finite or not. We have then, by using the above computation:

$$\frac{d^2f}{dxdy}=\frac{d^2f}{dydx}=\sum_{pq}c_{pq}pqx^{p-1}y^{q-1}$$

Thus, we can see that our commutation formula for derivatives holds indeed, and this due to the fact that the functions in x and y commute. Of course, this does not prove our formula, in general. But exercise for you, to have this idea fully working.

(2) Getting now to more standard techniques, given a point in the plane, $z=(a,b)$, consider the following functions, depending on $h,k\in\mathbb{R}$ small:

$$u(h,k)=f(a+h,b+k)-f(a+h,b)$$

$$v(h,k)=f(a+h,b+k)-f(a,b+k)$$

$$w(h,k)=f(a+h,b+k)-f(a+h,b)-f(a,b+k)+f(a,b)$$

By the mean value theorem, for $h,k\neq 0$ we can find $\alpha,\beta\in\mathbb{R}$ such that:

$$\begin{aligned}
w(h,k)&=&&u(h,k)-u(0,k)\\
&=&&h\cdot\frac{d}{dx}\,u(\alpha h,k)\\
&=&&h\left(\frac{d}{dx}\,f(a+\alpha h,b+k)-\frac{d}{dx}\,f(a+\alpha h,b)\right)\\
&=&&hk\cdot\frac{d}{dy}\cdot\frac{d}{dx}\,f(a+\alpha h,b+\beta k)
\end{aligned}$$

Similarly, again for $h,k\neq 0$, we can find $\gamma,\delta\in\mathbb R$ such that:

$$\begin{aligned} w(h,k) &= v(h,k)-v(h,0)\\ &= k\cdot\frac{d}{dy}\,v(h,\delta k)\\ &= k\left(\frac{d}{dy}\,f(a+h,b+\delta k)-\frac{d}{dy}\,f(a,b+\delta k)\right)\\ &= hk\cdot\frac{d}{dx}\cdot\frac{d}{dy}\,f(a+\gamma h,b+\delta k)\end{aligned}$$

Now by dividing everything by $hk\neq 0$, we conclude from this that the following equality holds, with the numbers $\alpha,\beta,\gamma,\delta\in\mathbb R$ being found as above:

$$\frac{d}{dy}\cdot\frac{d}{dx}\,f(a+\alpha h,b+\beta k)=\frac{d}{dx}\cdot\frac{d}{dy}\,f(a+\gamma h,b+\delta k)$$

But with $h,k\to 0$ we get from this the Clairaut formula, at $z=(a,b)$, as desired. □

In arbitrary dimensions now, we have the following result:

Theorem 1.29. *Given $f:\mathbb R^N\to\mathbb R$, we can talk about its higher derivatives,*

$$\frac{d^kf}{dx_{i_1}\ldots dx_{i_k}}=\frac{d}{dx_{i_1}}\cdots\frac{d}{dx_{i_k}}(f)$$

provided that these derivatives exist indeed. Moreover, due to the Clairaut formula,

$$\frac{d^2f}{dx_idx_j}=\frac{d^2f}{dx_jdx_i}$$

the order in which these higher derivatives are computed is irrelevant.

Proof. There are several things going on here, the idea being as follows:

(1) First of all, we can talk about the quantities in the statement, with the remark however that at each step of our recursion, the corresponding partial derivative can exist of not. We will say in what follows that our function is k times differentiable if the quantities in the statement exist at any $l\leq k$, and smooth, if this works with $k=\infty$.

(2) Regarding now the second assertion, this is something more tricky. Let us first recall from Theorem 1.28 that the second derivatives of a twice differentiable function of two variables $f:\mathbb R^2\to\mathbb R$ are subject to the Clairaut formula, namely:

$$\frac{d^2f}{dxdy}=\frac{d^2f}{dydx}$$

(3) But this result clearly extends to our function $f:\mathbb R^N\to\mathbb R$, simply by ignoring the unneeded variables, so we have the Clairaut formula in general, also called Schwarz

formula, which is the one in the statement, namely:

$$\frac{d^2f}{dx_idx_j}=\frac{d^2f}{dx_jdx_i}$$

(4) Now observe that this tells us that the order in which the higher derivatives are computed is irrelevant. That is, we can permute the order of our partial derivative computations, and a standard way of doing this is by differentiating first with respect to x_1, as many times as needed, then with respect to x_2, and so on. Thus, the collection of partial derivatives can be written, in a more convenient form, as follows:

$$\frac{d^kf}{dx_1^{k_1}\ldots dx_N^{k_N}}=\frac{d^{k_1}}{dx_1^{k_1}}\cdots\frac{d^{k_N}}{dx_N^{k_N}}(f)$$

(5) To be more precise, here $k\in\mathbb N$ is as usual the global order of our derivatives, the exponents $k_1,\ldots,k_N\in\mathbb N$ are subject to the condition $k_1+\ldots+k_N=k$, and the operations on the right are the familiar one-variable higher derivative operations.

(6) This being said, for certain tricky questions it is more convenient not to order the indices, or rather to order them according to what order best fits our computation, so what we have in the statement is the good formula, and (4-5) are mere remarks.

(7) And with the remark too that for trivial questions, what we have in the statement is the good formula, simply because there are less indices to be written, when compared to what we have to write when using the ordering procedure in (4-5) above. □

All this is very nice, and as an illustration, let us work out in detail the case $k=2$. Here things are quite special, and we can formulate the following definition:

Definition 1.30. *Given a twice differentiable function $f:\mathbb R^N\to\mathbb R$, we set*

$$f''(x)=\left(\frac{d^2f}{dx_idx_j}\right)_{ij}$$

which is a symmetric matrix, called Hessian matrix of f at the point $x\in\mathbb R^N$.

To be more precise, we know that when $f:\mathbb R^N\to\mathbb R$ is twice differentiable, its order $k=2$ partial derivatives are the numbers in the statement. Now since these numbers naturally form a $N\times N$ matrix, the temptation is high to call this matrix $f''(x)$, and so we will do. And finally, we know from Clairaut that this matrix is symmetric:

$$f''(x)_{ij}=f''(x)_{ji}$$

Observe that at $N=1$ this is compatible with the usual definition of the second derivative f'', because in this case, the 1×1 matrix from Definition 1.30 is:

$$f''(x)=(f''(x))\in M_{1\times1}(\mathbb R)$$

As a word of warning, however, never use Definition 1.30 for functions $f:\mathbb R^N\to\mathbb R^M$, where the second derivative can only be something more complicated. Also, never attempt either to do something similar at $k = 3$ or higher, for functions $f : \mathbb{R}^N \to \mathbb{R}$ with $N > 1$, because again, that beast has too many indices, for being a true, honest matrix.

Back now to our usual business, approximation, we have the following result:

THEOREM 1.31. *Given a twice differentiable function $f : \mathbb{R}^N \to \mathbb{R}$, we have*

$$f(x+t) \simeq f(x) + f'(x)t + \frac{< f''(x)t, t >}{2}$$

where $f''(x) \in M_N(\mathbb{R})$ stands as usual for the Hessian matrix.

PROOF. This is something more tricky, the idea being as follows:

(1) As a first observation, at $N = 1$ the Hessian matrix as constructed in Definition 1.30 is the 1×1 matrix having as entry the second derivative $f''(x)$, and the formula in the statement is something that we know well from basic calculus, namely:

$$f(x+t) \simeq f(x) + f'(x)t + \frac{f''(x)t^2}{2}$$

(2) In general now, this is in fact something which does not need a new proof, because it follows from the one-variable formula above, applied to the restriction of f to the following segment in $\mathbb{R}^N$, which can be regarded as being a one-variable interval:

$$I = [x, x+t]$$

To be more precise, let $y \in \mathbb{R}^N$, and consider the following function, with $r \in \mathbb{R}$:

$$g(r) = f(x + ry)$$

We know from (1) that the Taylor formula for g, at the point $r = 0$, reads:

$$g(r) \simeq g(0) + g'(0)r + \frac{g''(0)r^2}{2}$$

And our claim is that, with $t = ry$, this is precisely the formula in the statement.

(3) So, let us see if our claim is correct. By using the chain rule, we have the following formula, with on the right, as usual, a row vector multiplied by a column vector:

$$g'(r) = f'(x + ry) \cdot y$$

By using again the chain rule, we can compute the second derivative as well:

$$\begin{aligned}
g''(r) &= (f'(x+ry)\cdot y)' \\
&= \left(\sum_i \frac{df}{dx_i}(x+ry)\cdot y_i\right)' \\
&= \sum_i\sum_j \frac{d^2f}{dx_idx_j}(x+ry)\cdot\frac{d(x+ry)_j}{dr}\cdot y_i \\
&= \sum_i\sum_j \frac{d^2f}{dx_idx_j}(x+ry)\cdot y_iy_j \\
&= < f''(x+ry)y, y >
\end{aligned}$$

(4) Time now to conclude. We know that we have $g(r) = f(x+ry)$, and according to our various computations above, we have the following formulae:

$$g(0) = f(x) \quad , \quad g'(0) = f'(x) \quad , \quad g''(0) =< f''(x)y, y >$$

Buit with this data in hand, the usual Taylor formula for our one variable function g, at order 2, at the point $r = 0$, takes the following form, with $t = ry$:

$$\begin{aligned}
f(x+ry) &\simeq f(x) + f'(x)ry + \frac{< f''(x)y, y > r^2}{2} \\
&= f(x) + f'(x)t + \frac{< f''(x)t, t >}{2}
\end{aligned}$$

Thus, we have obtained the formula in the statement.

(5) Finally, for completness, let us record as well a more numeric formulation of what we found. According to our usual rules for matrix calculus, what we found is:

$$f(x+t) \simeq f(x) + \sum_{i=1}^N \frac{df}{dx_i}\,t_i + \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N \frac{d^2f}{dx_idx_j}\,t_it_j$$

Observe that, since the Hessian matrix $f''(x)$ is symmetric, most of the terms on the right will appear in pairs, making it clear what the $1/2$ is there for, namely avoiding redundancies. However, this is only true for the off-diagonal terms, so instead of further messing up our numeric formula above, we will just leave it like this. □

As in the one variable case, the Taylor formula is useful for computing the local extrema of the function. Indeed, let us first look at the order 1 formula, namely:

$$f(x+t) \simeq f(x) + f'(x)t$$

It is clear then, exactly as in the one-variable case, that in order to have a local extremum, we must have $f'(x) = 0$. Next, assuming that this holds, let us look at the order 2 Taylor formula, which in the case $f'(x)=0$ takes the following form:

$$f(x+t)\simeq f(x)+\frac{<f''(x)t,t>}{2}$$

We conclude from this, again as in the one-variable case, that when $f''(x)>0$, with this meaning that the symmetric matrix $f''(x)\in M_N(\mathbb{R})$ must be strictly positive, we have a local minimum, and that when $f''(x)<0$, we have a local maximum.

At higher order now, things become more complicated, as follows:

THEOREM 1.32. *Given an order k differentiable function $f:\mathbb{R}^N\to\mathbb{R}$, we have*

$$f(x+t)\simeq f(x)+f'(x)t+\frac{<f''(x)t,t>}{2}+\ldots$$

and this helps in identifying the local extrema, when $f'(x)=0$ and $f''(x)=0$.

PROOF. The study here is very similar to that at $k=2$, from the proof of Theorem 1.31, with everything coming from the usual Taylor formula, applied on:

$$I=[x,x+t]$$

Thus, it is pretty much clear that we are led to the conclusion in the statement. We will leave some study here as an instructive exercise. □

1e. Exercises

This was a standard multivariable calculus chapter, and as exercises, we have:

EXERCISE 1.33. *Clarify the details, in the diagonalization of the all-one matrix.*

EXERCISE 1.34. *Learn the spectral theorem, in its various forms mentioned above.*

EXERCISE 1.35. *Learn about the resultant and discriminant of polynomials.*

EXERCISE 1.36. *Learn more about the positive matrices, and their properties.*

EXERCISE 1.37. *Clarify what we said, in relation with open and closed sets.*

EXERCISE 1.38. *Clarify also what we said about compact and connected sets.*

EXERCISE 1.39. *Learn some other formulations of the chain rule formula.*

EXERCISE 1.40. *Work out the multivariable Taylor formula at arbitrary order.*

As bonus exercise, read more linear algebra, as much as you can. All good learning.

CHAPTER 2

Integration theory

2a. Measure theory

With the derivatives of the functions $f:\mathbb R^N\to\mathbb R^M$ understood, time now to discuss the integrals. Obviously, the integral of a function $f:\mathbb R^N\to\mathbb R^M$ can only be the vector of $\mathbb R^M$ formed by the integrals of its components $f_i:\mathbb R^N\to\mathbb R$, so in order to construct the integral, we can assume $M=1$. Thus, we are led to the following question:

QUESTION 2.1. *How to integrate the functions $f:\mathbb R^N\to\mathbb R$,*

$$f\to\int_{\mathbb R^N}f(z)dz$$

in analogy with what we know about integrating functions $f:\mathbb R\to\mathbb R$?

In answer, and taking $N=2$ for simplifying, I bet that your answer would be that we can define the multivariable integral by iterating, as follows:

$$\int_{\mathbb R^2}f(z)dz=\int_{\mathbb R}\int_{\mathbb R}f(x,y)dxdy$$

Which looks fine, at a first glance, but there is in fact a bug, with this. Indeed, assuming so, we would have by symmetry the following formula too:

$$\int_{\mathbb R^2}f(z)dz=\int_{\mathbb R}\int_{\mathbb R}f(x,y)dydx$$

Thus, and forgetting now about what we wanted to do, we can see that our method is based on the Fubini formula, stating that we must have:

$$\int_{\mathbb R}\int_{\mathbb R}f(x,y)dxdy=\int_{\mathbb R}\int_{\mathbb R}f(x,y)dydx$$

But, you might already know from calculus that Fubini does not always work, with the counterexamples being not very difficult to construct, as follows:

THEOREM 2.2. *The Fubini formula, namely*

$$\int_{\mathbb R}\int_{\mathbb R}f(x,y)dxdy=\int_{\mathbb R}\int_{\mathbb R}f(x,y)dydx$$

can fail, for certain suitably chosen functions.

Proof. We have indeed the following computation:

$$\begin{aligned}\int_0^1\int_0^1\frac{y^2-x^2}{(x^2+y^2)^2}\,dxdy &= \int_0^1\left[\frac{x}{x^2+y^2}\right]_0^1 dy\\ &= \int_0^1\frac{1}{1+y^2}\,dy\\ &= \frac{\pi}{4}\end{aligned}$$

On the other hand, by using this, and symmetry, we have as well:

$$\begin{aligned}\int_0^1\int_0^1\frac{y^2-x^2}{(x^2+y^2)^2}\,dydx &= \int_0^1\int_0^1\frac{x^2-y^2}{(x^2+y^2)^2}\,dxdy\\ &= -\int_0^1\int_0^1\frac{y^2-x^2}{(x^2+y^2)^2}\,dxdy\\ &= -\frac{\pi}{4}\end{aligned}$$

Thus Fubini can fail for certain functions, as said in the statement. □

Summarizing, our method does not work. Which can be regarded as good news for us mathematicians, time now to get into some action, show our skills, and fix all this. Let us start with the following abstract definition, which is something very general:

Definition 2.3. *An abstract measured space is a set X, given with a set of subsets $M\subset P(X)$, called measurable sets, which form an algebra, in the sense that:*

(1) $\emptyset, X\in M$.
(2) $E\in M\implies E^c\in M$.
(3) M *is stable under countable unions and intersections.*

Obviously, this is something quite abstract. If the last axiom, (3), is only satisfied for the finite unions and intersections, we say that $M\subset P(X)$ is a finite algebra. Getting now to the concrete examples of abstract measured spaces, we have:

Definition 2.4. *Any metric space X is automatically an abstract measured space, with the algebra of measurable sets being*

$$B=\bar{O}$$

that is, the smallest algebra containing the open sets, called Borel algebra of X.

Observe that the Borel sets include all open sets, all closed sets, as well as all countable unions of closed sets, and all countable intersections of open sets. As an example here, in the case $X=\mathbb{R}$, with its usual topology, all kinds of intervals are Borel sets:

$$(a,b),\ [a,b],\ (a,b],\ [a,b)\in B$$

Indeed, the first interval is open, and the second one is closed, so these are certainly Borel sets. As for the third and fourth intervals, these appear as countable unions of closed intervals, or as countable intersections of open intervals, so they are Borel too.

Back to theory, we can talk about measurable functions, as follows:

DEFINITION 2.5. *Given a measured space X, and a metric space Y, a function*

$$f : X \to Y$$

is called measurable when it satisfies the following condition:

$$U \in O \implies f^{-1}(U) \in M$$

When X comes with a measure, we also call such functions integrable.

We can talk as well about measures on measurable spaces, as follows:

DEFINITION 2.6. *Given a measured space (X, M), a measure on it is a function*

$$\mu : M \to [0, \infty]$$

which is countably additive, in the sense that we have

$$\mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$$

for any countable family of disjoint measurable sets $E_i \in M$.

Time now to put everything together. We have the following result, which is stated a bit informally, with some of the details being left to you, as an instructive exercise:

THEOREM 2.7. *We can integrate the measurable functions $f : X \to \mathbb{R}_+$ by setting*

$$\int_X f(x) d\mu(x) = \sup_{0 \le \varphi \le f} \int_X \varphi(x) d\mu(x)$$

with sup over measurable step functions, then extend this by linearity.

PROOF. This is something very standard, the idea being as follows:

(1) We can certainly integrate the step functions $\varphi : X \to \mathbb{R}_+$, by writing each such function as a linear combination of characteristic functions, as follows:

$$\varphi = \sum_i \lambda_i \chi_{E_i}$$

Indeed, with this formula in hand, we can integrate our function φ, as follows:

$$\int_X \varphi(x)\, d\mu(x) = \sum_i \lambda_i \mu(E_i)$$

The integral of step functions constructed in this way has then all the linearity and positivity properties that you might expect, and behaves well with respect to limits.

(2) Next, consider an arbitrary measurable function $f : X \to [0, \infty]$. It is routine to see, from definitions, that we can write this function as an increasing limit, as follows, with $0 \leq \varphi_1 \leq \varphi_2 \leq \ldots \leq f$, and with each φ_i being a measurable step function:

$$f(x) = \lim_{n \to \infty} \varphi_n(x)$$

But this suggests to define the integral of f by the formula in the statement, namely:

$$\int_X f(x) d\mu(x) = \sup_{0 \leq \varphi \leq f} \int_X \varphi(x) d\mu(x)$$

Indeed, we can see that the integral constructed in this way has all the linearity and positivity properties that you might expect, and behaves well with respect to limits. □

Good news, eventually, the above is all that we need, as abstract preliminaries. We can now formulate the key result in general measure theory, as follows:

THEOREM 2.8 (Riesz). *Any positive functional $I : C_c(X) \to \mathbb{R}$ comes by integrating with respect to a certain measure on X,*

$$I(f) = \int_X f(x) d\mu(x)$$

which has the following additional properties,

(1) $\mu(K) < \infty$*, for any $K \subset X$ compact.*
(2) $\mu(E) = \inf\{\mu(U) | E \subset U \text{ open}\}$*, for any E measurable.*
(3) $\mu(E) = \sup\{\mu(K) | K \subset E \text{ compact}\}$*, for any E open, or of finite measure.*

and which is unique with these properties, modulo the null sets.

PROOF. This is something quite long and tricky, the idea being as follows:

(1) Let us first prove the uniqueness. Assuming that our measure μ produces I, and has the various properties in the statement, it is clear that μ is uniquely determined by its values on the compact sets $K \subset X$. Thus, we must show that given two measures μ_1, μ_2 as in the statement, and a compact set $K \subset X$, we have:

$$\mu_1(K) = \mu_2(K)$$

For this purpose, let us pick $\varepsilon > 0$. By using the properties (1,3) in the statement, for the measure μ_2, we can find $K \subset U$ open such that:

$$\mu_2(U) < \mu_2(K) + \varepsilon$$

Next, using $K \subset U$, pick a continuous, compactly supported function f, such that:

$$\chi_K \leq f \leq \chi_U$$

But, with this choice of f, we have the following computation:

$$\begin{aligned}\mu_1(K) &\leq \int_X f(x)d\mu_1(x)\\ &= \int_X f(x)d\mu_2(x)\\ &\leq \mu_2(U)\\ &< \mu_2(K)+\varepsilon\end{aligned}$$

Thus $\mu_1(K) \leq \mu_2(K)$, and by interchanging μ_1, μ_2 we have the reverse inequality as well, so we obtain, as desired, $\mu_1(K) = \mu_2(K)$, proving the uniqueness statement.

(2) With this done, let us get now to the main part, and discuss the construction of the algebra M, that we will choose to be saturated with respect to the null sets, and of the measure μ. What we have as data is a positive functional, as follows:

$$I : C_c(X) \to \mathbb{R}$$

We must first measure the Borel sets $E \subset X$. And here, to start with, in order to measure the open sets $U \subset X$, the formula is straightforward, namely:

$$\mu(U) = \sup\left\{I(f)\Big|f \leq \chi_U\right\}$$

Now observe that with this definition in hand, for the open sets, we have:

$$U_1 \subset U_2 \implies \mu(U_1) \leq \mu(U_2)$$

We deduce that the following happens, for the open sets:

$$\mu(E) = \inf\left\{\mu(U)\Big|E \subset U \text{ open}\right\}$$

But this can serve as the definition of μ, for all the subsets $E \subset X$.

(3) Regarding now the measurable sets, let us first consider the class N of subsets $E \subset X$ which are of finite measure, $\mu(E) < \infty$, and satisfy the following condition:

$$\mu(E) = \sup\left\{\mu(K)\Big|K \subset E \text{ compact}\right\}$$

Then, we can define the class of measurable sets $M \subset P(X)$ as follows:

$$M = \left\{E \subset X\Big|E \cap K \in N, \forall K \subset X \text{ compact}\right\}$$

Summarizing, done with definitions, and it remains now to prove that M is an algebra, and that μ is a measure on it, along with the other things claimed in the statement.

(4) In order to prove that μ is indeed a measure, our first claim is that we have the following inequality, for any two open sets $U_1, U_2 \subset X$:

$$\mu(U_1 \cup U_2) \leq \mu(U_1) + \mu(U_2)$$

In order to prove this claim, choose an arbitrary function $g : X \to [0,1]$, supported on $U_1 \cup U_2$. We can then find a certain partition of unity, $f_1 + f_2 = 1$ on the support of g, and it follows that we have the following equality:

$$g = f_1 g + f_2 g$$

Now by applying our integration functional I, we obtain from this:

$$\begin{aligned} I(g) &= I(f_1 g + f_2 g) \\ &= I(f_1 g) + I(f_2 g) \\ &\leq \mu(U_1) + \mu(U_2) \end{aligned}$$

Now since this holds for any $g : X \to [0,1]$ as above, this proves our claim.

(5) Our second claim is that we have in fact, more generally, the following inequality, valid for any family of subsets $E_1, E_2, E_3, \ldots$ of our space X:

$$\mu\left(\bigcup_{i=1}^\infty E_i\right) \leq \sum_{i=1}^\infty \mu(E_i)$$

Indeed, this inequality trivially holds if $\mu(E_i) = \infty$ for some i, so we can assume $\mu(E_i) < \infty$ for any i. Now pick $\varepsilon > 0$, and choose open sets $U_i \supset E_i$ such that:

$$\mu(U_i) < \mu(E_i) + \frac{\varepsilon}{2^i}$$

Consider also the union of these open sets U_i, chosen as above:

$$U = \bigcup_{i=1}^\infty U_i$$

Now pick an arbitrary function $f : X \to [0,1]$, supported on this open set U. Since f has compact support, we can find a certain integer $n \in \mathbb{N}$ such that:

$$supp(f) \subset U_1 \cup \ldots \cup U_n$$

Now by using what we found in (4), recursively, we obtain:

$$\begin{aligned} I(f) &\leq \mu(U_1 \cup \ldots \cup U_n) \\ &\leq \mu(U_1) + \ldots + \mu(U_n) \\ &\leq \sum_{i=1}^n \mu(E_i) + \varepsilon \end{aligned}$$

Since this holds for any $f : X \to [0,1]$ as above, we deduce that we have:

$$\mu\left(\bigcup_{i=1}^\infty E_i\right) \leq \mu(U) \leq \sum_{i=1}^\infty \mu(E_i) + \varepsilon$$

Now since the number $\varepsilon > 0$ was arbitrary, this proves our claim.

(6) Our next claim, which proves in the assertion (1) in the theorem, is that any compact set $K\subset X$ is measurable, and that we have:

$$\mu(K)<\infty$$

In order to prove this claim, consider an arbitrary continuous function $f:X\to[0,1]$ satisfying $f(x)=1$ on K, then pick $\alpha<1$, and consider the following open set:

$$U_\alpha=\left\{x\in X\Big|f(x)>\alpha\right\}$$

We have then $K\subset U_\alpha$, and so $\alpha g\leq f$, whenever a continuous function $g:X\to[0,1]$ is supported by U_α. By using this, we obtain the following estimate:

$$\begin{aligned}\mu(K)&\leq\mu(U_\alpha)\\&=\sup\left\{I(g)\Big|g:X\to[0,1],\ supp(g)\subset U_\alpha\right\}\\&\leq\alpha^{-1}I(f)\end{aligned}$$

Now with $\alpha\to1$, we obtain from this the following estimate:

$$\mu(K)\leq I(f)$$

Thus $\mu(K)<\infty$, as claimed, and the fact that K is measurable is clear too.

(7) Our next claim, which improves what we found in (6), is that for any compact set $K\subset X$ we have in fact the following estimate:

$$\mu(K)=\inf\left\{I(f)\Big|f:X\to[0,1],\ f(x)=1\text{ on }K\right\}$$

In order to prove this, let us go back to the proof of (6). We know from there that for any continuous function $f:X\to[0,1]$ satisfying $f(x)=1$ on K, as above, we have:

$$\mu(K)\leq I(f)$$

Now pick $\varepsilon>0$, and choose an open set $U\supset K$ satisfying:

$$\mu(U)<\mu(K)+\varepsilon$$

We can then find a compactly supported function f such that:

$$\chi_K\leq f\leq\chi_U$$

But this gives the following estimate, using the above inequality $\mu(K)\leq I(f)$:

$$\begin{aligned}I(f)&\leq\mu(V)\\&<\mu(K)+\varepsilon\\&\leq I(f)+\varepsilon\end{aligned}$$

Now since the number $\varepsilon>0$ was arbitrary, this proves our claim.

(8) Our claim now is that any open set $U\subset X$ satisfies the following condition, that we used in (3) in order to define the measurable sets:

$$\mu(U)=\sup\left\{\mu(K)\Big|K\subset U\text{ compact}\right\}$$

In order to prove this, pick an arbitrary number $\alpha<\mu(U)$. We can then find a continuous function $f:X\to[0,1]$ supported on U, such that:

$$I(f)>\alpha$$

Consider now the compact set $K=supp(f)$. Then for any open set V containing K we have $I(f)\leq\mu(V)$, and we deduce from this that we have:

$$I(f)\leq\mu(K)$$

In other words, given $\alpha<\mu(U)$, we have found a compact set K such that:

$$\mu(K)\geq\alpha$$

Now since our number $\alpha<\mu(U)$ was arbitrary, this proves our claim.

(9) Getting now towards the proof of the additivity of our measure μ, let us first prove that for any two disjoint compact sets $K_1,K_2\subset X$, we have:

$$\mu(K_1\cup K_2)=\mu(K_1)+\mu(K_2)$$

For this purpose, pick an arbitrary $\varepsilon>0$. We can then find a compactly supported continuous function $f:X\to[0,1]$ satisfying:

$$f(x)=\begin{cases}1&\text{on }K_1\\0&\text{on }K_2\end{cases}$$

On the other hand, by using (7) we can find a compactly supported continuous function $g:X\to[0,1]$ satisfying $g(x)=1$ on $K_1\cup K_2$, such that:

$$I(g)\leq\mu(K_1\cup K_2)+\varepsilon$$

By using this, and the linearity of I, we have the following estimate:

$$\begin{aligned}\mu(K_1)+\mu(K_2)&\leq I(fg)+I(g-fg)\\&=I(g)\\&<\mu(K_1\cup K_2)+\varepsilon\end{aligned}$$

Now since our number $\varepsilon>0$ was arbitrary, this proves our claim, via (5).

(10) We are now in position of making a key verification, that of the additivity property of our measure μ. To be more precise, our claim is that we have the following equality, valid for any family of pairwise disjoint subsets $E_1,E_2,E_3,\ldots$ of our space X:

$$\mu\left(\bigcup_{i=1}^\infty E_i\right)=\sum_{i=1}^\infty\mu(E_i)$$

In order to prove this, consider the union on the left, namely:

$$E = \bigcup_{i=1}^\infty E_i$$

In the case $\mu(E) = \infty$ our claim is proved by (5). So, assume $\mu(E) < \infty$, and pick an arbitrary number $\varepsilon > 0$. We can then find compact sets $K_i \subset E_i$ such that:

$$\mu(K_i) > \mu(E_i) - \frac{\varepsilon}{2^i}$$

Now consider the following unions, which are compact sets too:

$$H_n = K_1 \cup \ldots \cup K_n$$

By using the additivity formula that we found in (9), recursively, we obtain:

$$\mu(E) \geq \mu(H_n) = \sum_{i=1}^n \mu(K_i) > \sum_{i=1}^n \mu(E_i) - \varepsilon$$

Now since our number $\varepsilon > 0$ was arbitrary, this proves our claim.

(11) Thus, additivity property proved, and with this in hand, the other assertions are quite standard. In order to prove these, our first claim is that for any measurable set E and any $\varepsilon > 0$, there is a compact set K, and an open set U, satisfying:

$$K \subset E \subset U \quad , \quad \mu(U - K) < \varepsilon$$

Indeed, in order to prove this claim, the first remark is that we can find indeed a compact set K and an open set U, satisfying the following conditions:

$$K \subset E \subset U \quad , \quad \mu(U) - \frac{\varepsilon}{2} < \mu(E) < \mu(K) + \frac{\varepsilon}{2}$$

Since the set $U - K$ is open, by using (8) and then (10) we obtain:

$$\mu(U) = \mu(K) + \mu(U - K)$$

But with this, we can establish our inequality, as follows:

$$\mu(U - K) = \mu(U) - \mu(K) < \varepsilon$$

(12) Our next claim is that if A, B are measurable, then so are the following sets:

$$A - B \quad , \quad A \cup B \quad , \quad A \cap B$$

In order to prove this, pick an arbitrary number $\varepsilon > 0$. By using (11) we can find certain compact sets K, H, and certain open sets U, V, satisfying:

$$K \subset A \subset U \quad , \quad \mu(U - K) < \varepsilon$$

$$H \subset B \subset V \quad , \quad \mu(V - H) < \varepsilon$$

Now observe that we have the following inclusion:

$$A - B \subset U - H \subset (U - K) \cup (K - V) \cup (V - H)$$

By using now (5), we obtain from this the following estimate:

$$\begin{aligned}\mu(A-B) &\leq \mu(U-K)+\mu(K-V)+\mu(V-H)\\ &< \varepsilon+\mu(K-V)+\varepsilon\\ &= \mu(K-V)+2\varepsilon\end{aligned}$$

Now since $K-V$ is a compact subset of $A-B$, we conclude that $A-B$ is measurable, as desired. Regarding now $A\cup B$, we can use here the following formula:

$$A\cup B=(A-B)\cup B$$

Indeed, by using (10) we obtain from this that $A\cup B$ is measurable as well. Finally, regarding $A\cap B$, we can use here the following formula:

$$A\cap B=A-(A-B)$$

Thus $A\cap B$ is measurable as well, and this finishes the proof of our claim.

(13) Our claim now is that the set M constructed in (3) is indeed an algebra, which contains all Borel sets. In order to prove this, consider an arbitrary compact set $K\subset X$. Given $A\in M$, we can write $A^c\cap K$ as a difference of two sets in M, as follows:

$$A^c\cap K=K-(A-K)$$

Thus $A^c\cap K\in M$, and by using this, we conclude that we have:

$$A\in M\implies A^c\in M$$

Next, let us look at unions. Assume $A_i\in M$, and consider their union:

$$A=\bigcup_{i=1}^\infty A_i$$

With $K\subset X$ being an arbitrary compact set, as above, set $B_1=A_1\cap K$, and then define recursively the following sets:

$$B_n=(A_n\cap K)-(B_1\cup\ldots\cup B_{n-1})$$

In terms of these sets B_n, we have the following formula:

$$A\cap K=\bigcup_{i=1}^\infty B_n$$

On the other hand, according to our definition of the sets B_n, and by using (12), we have $B_n\in M$. Thus, by using (10), we obtain $A\in M$. Thus, we have proved that:

$$A_i\in M\implies\bigcup_{i=1}^\infty A_i\in M$$

Finally, if a subset $C\subset X$ is closed, then $C\cap K$ is compact, and so $C\in M$. Thus, as claimed above, M is indeed an algebra, which contains all the Borel sets.

(14) Our next claim, which together with what we found in (10) will prove that μ is indeed a measure on M, is that, in the context of our constructions in (3), the family N constructed there consists precisely of the sets $E \in M$ satisfying:

$$\mu(E) < \infty$$

In order to prove this, let $E \in N$. By using (7) and (12) we deduce that we have $E \cap K \in N$ for any compact set $K \subset X$, and we conclude from this that we have:

$$E \in N \implies E \in M$$

Conversely, assume that $E \in M$ has finite measure, $\mu(E) < \infty$. In order to prove $E \in N$, we pick a number $\varepsilon > 0$. We can then find an open set $U \subset X$ such that:

$$E \subset U \quad , \quad \mu(U) < \infty$$

Also, by (8) and (11) we can find a compact set $K \subset X$ such that:

$$K \subset V \quad , \quad \mu(U - K) < \varepsilon$$

Now since $E \cap K \in N$, we can find a compact set $H \subset X$ such that:

$$H \subset E \cap K \quad , \quad \mu(E \cap K) < \mu(H) + \varepsilon$$

Observe now that we have the following inclusion:

$$E \subset (E \cap K) \cup (U - K)$$

By using this, we obtain the following inequality:

$$\begin{aligned} \mu(E) &\leq \mu\left((E \cap K) \cup (U - K)\right) \\ &\leq \mu(E \cap K) + \mu(U - K) \\ &< \mu(H) + 2\varepsilon \end{aligned}$$

Thus $E \in N$, which finishes the proof of our claim, and proves that μ is a measure.

(15) Time now to get into the final verification, that of the following formula:

$$I(f) = \int_X f(x) d\mu(x)$$

As a first observation, by taking real and imaginary parts, it is enough to prove this for the real functions f. Moreover, and assuming now that f is real, by using $f \to -f$, in order to prove the above equality, it is enough to prove the following inequality:

$$I(f) \leq \int_X f(x) d\mu(x)$$

(16) So, let us prove now this latter inequality. For this purpose, let $K \subset X$ be the support of our function f, choose an interval containing the range, $Im(f) \subset [a, b]$, pick an arbitrary $\varepsilon > 0$, and then numbers $y_0, \ldots, y_n$ such that $y_i - y_{i-1} < \varepsilon$, and:

$$y_0 < a < y_1 < \ldots < y_n = b$$

With this done, consider now the following sets:

$$E_i=\left\{x\in X\Big|y_{i-1}<f(x)<y_i\right\}\cap K$$

Since f is continuous, these are disjoint Borel sets, whose union is K. Thus, we can find open sets $U_i\subset X$, subject to the following conditions:

$$E_i\subset U_i\quad,\quad\mu(U_i)<\mu(E_i)+\frac{\varepsilon}{n}\quad,\quad f(x)<y_i+\varepsilon,\forall x\in U_i$$

Next, we can find certain continuous functions $h_i:X\to[0,1]$, with h_i supported on U_i, such that the following equality holds, on K:

$$h_1(x)+\ldots+h_n(x)=1$$

By multiplying by f, we conclude that the following equality holds, on K:

$$f(x)=h_1(x)f(x)+\ldots+h_n(x)f(x)$$

On the other hand, by using (7), we have the following estimate:

$$\mu(K)\leq I\left(\sum_{i=1}^nh_i\right)=\sum_{i=1}^nI(h_i)$$

Good news, with all these ingredients in hand, we can now finish. Since we have $h_if\leq(y_i+\varepsilon)h_i$, and since $y_i-\varepsilon<f(x)$ on E, we have the following estimate:

$$\begin{aligned}
I(f)&=\sum_{i=1}^nI(h_if)\\
&\leq\sum_{i=1}^n(y_i+\varepsilon)I(h_i)\\
&=\sum_{i=1}^n(|a|+y_i+\varepsilon)I(h_i)-|a|\sum_{i=1}^nI(h_i)\\
&\leq\sum_{i=1}^n(|a|+y_i+\varepsilon)\left(\mu(E_i)+\frac{\varepsilon}{n}\right)-|a|\mu(K)\\
&=\sum_{i=1}^n(y_i-\varepsilon)\mu(E_i)+2\varepsilon\mu(K)+\frac{\varepsilon}{n}\sum_{i=1}^n(|a|+y_i+\varepsilon)\\
&\leq\int_Xf(x)d\mu(x)+\varepsilon\left(2\mu(K)+|a|+b+\varepsilon\right)
\end{aligned}$$

Now since $\varepsilon>0$ was arbitrary, this finishes the proof of our latest claim, that is, of the inequality in (15), and so finishes the proof of the present theorem. □

2b. Lebesgue, Fubini

Good work that we did, in the above, and time now to enjoy what we learned. As a first consequence of the Riesz theorem, we can now formulate, following Lebesgue:

THEOREM 2.9 (Lebesgue). *We can measure all the Borel sets*

$$E\subset\mathbb{R}$$

and the corresponding integration theory extends the usual Riemann integral.

PROOF. This follows indeed from Theorem 2.8, by using the Riemann integral, that we know well from basic calculus, as the positive functional needed there:

$$f\to\int_{\mathbb{R}}f(x)dx$$

Indeed, it follows from basic calculus that the Riemann integral satisfies indeed all the needed properties in Theorem 2.8, and this gives the result. We will leave the various verifications here, which are all elementary, as an instructive exercise. □

In higher dimensions now, let us first formulate a general objective, as follows:

OBJECTIVE 2.10. *We would like to measure all the Borel sets*

$$E\subset\mathbb{R}^N$$

with the corresponding integration theory extending and clarifying the Riemann integral, in several variables. Also, we would like to clarify the validity of the Fubini formula

$$\int_{\mathbb{R}}\int_{\mathbb{R}}f(x,y)dxdy=\int_{\mathbb{R}}\int_{\mathbb{R}}f(x,y)dydx$$

in this formulation, in 2 *variables, and in general* $N\geq 2$ *variables too.*

Getting to work now, it is pretty much obvious that, in order to deal with this problem, rigorously, it is better to forget everything that we know from multivariable calculus, regarding the integration in several variables, and recreate everything from scratch.

So, here is our first result, using our abstract measure theory techniques:

PROPOSITION 2.11. *Given a compactly supported function* $f:\mathbb{R}^N\to\mathbb{C}$*, if we set*

$$I_k(f)=\frac{1}{2^{Nk}}\sum_{x\in\mathbb{Z}^N/2^k}f(x)$$

with $\mathbb{Z}^N/2^k\subset\mathbb{R}^N$ *being the points having as coordinates integer multiples of* $1/2^k$*, then*

$$I(f)=\lim_{k\to\infty}I_k(f)$$

converges, and $f\to I(f)$ *satisfies the assumptions of the Riesz theorem.*

Proof. In order to prove the first assertion, we can assume that our function is real, $f:\mathbb{R}^N\to\mathbb{R}$. But then, we can use the fact that f is uniformly continuous, and this gives the convergence in the statement. As for the second assertion, the fact that $f\to I(f)$ satisfies indeed the assumptions of the Riesz theorem is clear from definitions. □

In view of the above, we can indeed apply the Riesz theorem. We obtain:

Theorem 2.12 (Lebesgue). *There is a unique regular, translation-invariant measure on $\mathbb{R}^N$, normalized as for the volume of the unit cube to be 1. This measure is called Lebesgue measure, and appears alternatively as the measure on the Borel sets*

$$E\subset\mathbb{R}$$

whose corresponding integration theory extends the usual Riemann integral, as constructed in Proposition 2.11, or equivalently, as in multivariable calculus.

Proof. Obviously, many things going on here, and as a first remark, this fulfills indeed our requirements from Objective 2.10, save for Fubini, that we will discuss later, in the next chapter. In what regards the rest, the idea is as follows:

(1) We want to prove that the Lebesgue measure exists, and is unique, with the various properties from the statement. But here, a bit of thinking shows that this measure can only be the one coming from Proposition 2.11, via the Riesz theorem.

(2) So, let us go back to Proposition 2.11, which allows us to apply the Riesz theorem, and apply indeed this Riesz theorem. We obtain in this way a certain regular Borel measure on $\mathbb{R}^N$, that we have to prove to have the properties in the statement.

(3) But here, all the assertions follow from what we have in the Riesz theorem, and we will leave the various verifications here as an instructive exercise. □

Summarizing, we know how to integrate on $\mathbb{R}^N$. As a continuation of this, as a first task, let us go back to the Fubini formula problematics. We have:

Theorem 2.13 (Fubini, Tonelli). *Given a function $f:\mathbb{R}^N\to\mathbb{R}$ which is measurable and integrable, in the sense that the following integral is finite,*

$$\int_{\mathbb{R}^N}|f(z)|dz<\infty$$

we have the following equalities, for any decomposition $N=N_1+N_2$:

$$\int_{\mathbb{R}^{N_1}}\int_{\mathbb{R}^{N_2}}f(x,y)dydx=\int_{\mathbb{R}^{N_2}}\int_{\mathbb{R}^{N_1}}f(x,y)dxdy=\int_{\mathbb{R}^N}f(z)dz$$

Moreover, the same holds when $f:\mathbb{R}^N\to\mathbb{R}$ is assumed positive, and measurable.

PROOF. In this statement, which is something quite compact, the first assertion is called Fubini theorem, and the second assertion is called Tonelli theorem. These two theorems are related, and can be also subject to various minor improvements. In what follows we will mostly focus on proving the Fubini theorem in its basic form, as stated.

(1) In order to prove the Fubini theorem, consider the vector space L consisting of the functions $f:\mathbb{R}^N\to\mathbb{R}$ which are measurable, and integrable:

$$L=\left\{f:\mathbb{R}^N\to\mathbb{R}\Big|\int_{\mathbb{R}^N}|f(z)|dz<\infty\right\}$$

Now fix a decomposition $N=N_1+N_2$, and consider the family $F\subset L$ of functions $f:\mathbb{R}^N\to\mathbb{R}$ which satisfy the formula in the statement, namely:

$$\int_{\mathbb{R}^{N_1}}\int_{\mathbb{R}^{N_2}}f(x,y)dydx=\int_{\mathbb{R}^{N_2}}\int_{\mathbb{R}^{N_1}}f(x,y)dxdy=\int_{\mathbb{R}^N}f(z)dz$$

With this conventions, we want to prove that the inclusion $F\subset L$ is an equality.

(2) But this comes from the following sequence of observations, which all come from definitions, and from the Lebesgue dominated convergence theorem:

– F is stable under taking finite linear combinations.

– F is stable under taking pointwise limits of increasing sequences.

– $1_E\in F$, for any set of finite measure of type $E=E_1\times E_2\subset\mathbb{R}^N$.

– $1_E\in F$, for any open set of finite measure $E\subset\mathbb{R}^N$.

– $1_E\in F$, for any countable intersection of open sets, of finite measure $E\subset\mathbb{R}^N$.

– $1_E\in F$, for any set of null measure $E\subset\mathbb{R}^N$.

– $1_E\in F$, for any set of finite measure $E\subset\mathbb{R}^N$.

Indeed, the proof of all these assertions, in the above precise order, is something quite straightforward, and once we have the last assertion, the result follows, because we can approximate any given function $f\in L$ by step functions, and we obtain $f\in F$.

(3) Summarizing, Fubini proved, modulo some straightforward verifications left to you, and the same goes with Tonelli, whose proof is quite routine, based on Fubini. □

As a conclusion, we know how to rigorously integrate on $\mathbb{R}^N$, and as a bonus, the theory that we developed seems to have many follow-ups. More on this later.

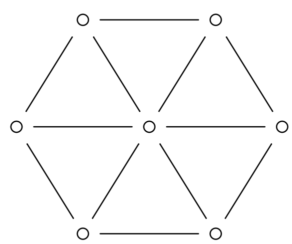
2c. Multiple integrals

Good news, time for some concrete mathematics. In the remainder of this chapter we will be interested in the explicit computation of the integrals in $\mathbb{R}^N$. Also, we would like to have more theory as well, such as a useful change of variable formula.

Let us start with some basics. In what regards the area of the circle, sure you know that, but never too late to learn the truth about it, which is as follows:

THEOREM 2.14. *The following two definitions of π are equivalent:*
(1) *The length of the unit circle is $L = 2\pi$.*
(2) *The area of the unit disk is $A = \pi$.*

PROOF. In order to prove this theorem let us cut the unit disk as a pizza, into N slices, and forgetting about gastronomy, leave aside the rounded parts:



The area to be eaten can be then computed as follows, where H is the height of the slices, S is the length of their sides, and $P = NS$ is the total length of the sides:

$$A = N \times \frac{HS}{2} = \frac{HP}{2} \simeq \frac{1 \times L}{2}$$

Thus, with $N \to \infty$ we obtain that we have $A = L/2$, as desired. □

As a second computation, the area of an ellipse can be computed as follows:

THEOREM 2.15. *The area of an ellipse, given by the equation*

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$$

with $a, b > 0$ being half the size of a box containing the ellipse, is $A = \pi ab$.

PROOF. The idea is that of cutting the ellipse into vertical slices. First observe that, according to our equation $(x/a)^2 + (y/b)^2 = 1$, the x coordinate can range as follows:

$$x \in [-a, a]$$

For any such x, the other coordinate y, satisfying $(x/a)^2 + (y/b)^2 = 1$, is given by:

$$y = \pm b\sqrt{1 - \frac{x^2}{a^2}}$$

We conclude that the area of the ellipse is given by the following formula:

$$\begin{aligned} A &= 2b\int_{-a}^{a}\sqrt{1-\frac{x^2}{a^2}}\,dx \\ &= 4ab\int_0^1\sqrt{1-y^2}\,dy \\ &= \pi ab \end{aligned}$$

Finally, as a verification, for $a=b=R$ we get $A=\pi R^2$, as we should. □

In what regards now the length of the ellipse, the "pizza" argument from the proof of Theorem 2.14 does not work, and things here get fairly complicated, as follows:

FACT 2.16. *The length of an ellipse, given by $(x/a)^2+(y/b)^2=1$, is*

$$L=4\int_0^{\pi/2}\sqrt{a^2\sin^2t+b^2\cos^2t}\,dt$$

and with this integral being generically not computable.

To be more precise, the above formula can be deduced in the following way, and more on such things later in this book, when systematically discussing the curves:

$$\begin{aligned} L &= 4\int_0^{\pi/2}\sqrt{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2}\,dt \\ &= 4\int_0^{\pi/2}\sqrt{\left(\frac{da\cos t}{dt}\right)^2+\left(\frac{db\sin t}{dt}\right)^2}\,dt \\ &= 4\int_0^{\pi/2}\sqrt{a^2\sin^2t+b^2\cos^2t}\,dt \end{aligned}$$

As for the last assertion, when $a=b=R$ we get of course $L=2\pi R$, as we should, but in general, when $a\neq b$, there is no trick for computing the above integral.

Moving now to 3D, as an obvious challenge here, we can try to compute the area and volume of the sphere, and more generally of the ellipsoids. We have here:

THEOREM 2.17. *The volume of the unit sphere in $\mathbb{R}^3$ is given by:*

$$V=\frac{4\pi}{3}$$

More generally, the volume of an ellipsoid, given by $(x/a)^2+(y/b)^2+(z/c)^2=1$, is:

$$V=\frac{4\pi abc}{3}$$

The area of the sphere is $A=4\pi$. For ellipsoids, the area is generically not computable.

Proof. There are several things going on here, as follows:

(1) Let us first compute the volume of the ellipsoid, which at $a=b=c=1$ will give the volume of the unit sphere. The range of the first coordinate x is as follows:

$$x\in[-a,a]$$

Now when the first coordinate x is fixed, the other coordinates y,z vary on an ellipse, given by the equation $(y/b)^2+(z/c)^2=1-(x/a)^2$, which can be written as follows:

$$\left(\frac{y}{\beta}\right)^2+\left(\frac{z}{\gamma}\right)^2=1\quad:\quad\beta=b\sqrt{1-\left(\frac{x}{a}\right)^2}\ ,\ \gamma=c\sqrt{1-\left(\frac{x}{a}\right)^2}$$

Thus, the vertical slice of our ellipsoid at x has area as follows:

$$A(x)=\pi\beta\gamma=\pi bc\left[1-\left(\frac{x}{a}\right)^2\right]$$

We conclude that the volume of the ellipsoid is given, as claimed, by:

$$\begin{aligned}V&=\pi bc\int_{-a}^{a}1-\left(\frac{x}{a}\right)^2dx\\&=\pi bc\left[x-\frac{x^3}{3a^2}\right]_{-a}^{a}\\&=\pi bc\left(\frac{2a}{3}+\frac{2a}{3}\right)\\&=\frac{4\pi abc}{3}\end{aligned}$$

(2) At $a=b=c=1$ we get $V=4\pi/3$, and this gives the area of the unit sphere too, because the "pizza" method from the proof of Theorem 2.14 applies, and gives:

$$A=3\times V=3\times\frac{4\pi}{3}=4\pi$$

(3) Finally, the last assertion is something quite informal, coming from Fact 2.16. □

In order to deal now with N dimensions, we will need the Wallis formula:

Theorem 2.18 (Wallis). *We have the following formulae,*

$$\int_0^{\pi/2}\cos^n t\,dt=\int_0^{\pi/2}\sin^n t\,dt=\left(\frac{\pi}{2}\right)^{\varepsilon(n)}\frac{n!!}{(n+1)!!}$$

where $\varepsilon(n)=1$ if n is even, and $\varepsilon(n)=0$ if n is odd, and where

$$m!!=(m-1)(m-3)(m-5)\ldots$$

with the product ending at 2 if m is odd, and ending at 1 if m is even.

Proof. Let us first compute the integral on the left in the statement:

$$I_n=\int_0^{\pi/2}\cos^n t\,dt$$

We do this by partial integration. We have the following formula:

$$\begin{aligned}(\cos^n t\sin t)' &= n\cos^{n-1}t(-\sin t)\sin t+\cos^n t\cos t\\ &= n\cos^{n+1}t-n\cos^{n-1}t+\cos^{n+1}t\\ &= (n+1)\cos^{n+1}t-n\cos^{n-1}t\end{aligned}$$

By integrating between 0 and $\pi/2$, we obtain the following formula:

$$(n+1)I_{n+1}=nI_{n-1}$$

Thus we can compute I_n by recurrence, and we obtain:

$$\begin{aligned}I_n &= \frac{n-1}{n}\,I_{n-2}\\ &= \frac{n-1}{n}\cdot\frac{n-3}{n-2}\,I_{n-4}\\ &= \frac{n-1}{n}\cdot\frac{n-3}{n-2}\cdot\frac{n-5}{n-4}\,I_{n-6}\\ &\ \vdots\\ &= \frac{n!!}{(n+1)!!}\,I_{1-\varepsilon(n)}\end{aligned}$$

But $I_0=\frac{\pi}{2}$ and $I_1=1$, so we get the result. As for the second formula, this follows from the first one, with $t=\frac{\pi}{2}-s$. Thus, we have proved both formulae in the statement. □

We can now compute the volumes of the N-dimensional spheres, as follows:

Theorem 2.19. *The volume of the unit sphere in $\mathbb{R}^N$ is given by*

$$V=\left(\frac{\pi}{2}\right)^{[N/2]}\frac{2^N}{(N+1)!!}$$

with our usual convention $N!!=(N-1)(N-3)(N-5)\ldots$

PROOF. If we denote by V_N the volume of the unit sphere in $\mathbb{R}^N$, we have:

$$\begin{aligned}
V_N &= \int_{-1}^1 (1-x^2)^{(N-1)/2} dx \cdot V_{N-1} \\
&= 2V_{N-1} \int_0^1 (1-x^2)^{(N-1)/2} dx \\
&= 2V_{N-1} \int_0^{\pi/2} (1-\sin^2 t)^{(N-1)/2} \cos t dt \\
&= 2V_{N-1} \int_0^{\pi/2} \cos^{N-1} t \cos t dt \\
&= 2V_{N-1} \int_0^{\pi/2} \cos^N t dt
\end{aligned}$$

Now by recurrence, and by using the formula in Theorem 2.18, we obtain:

$$\begin{aligned}
V_N &= 2^N \int_0^{\pi/2} \cos^N t dt \int_0^{\pi/2} \cos^{N-1} t dt \ldots\ldots \int_0^{\pi/2} \cos t dt \\
&= 2^N \left(\frac{\pi}{2}\right)^{\varepsilon(N)+\varepsilon(N-1)+\ldots+\varepsilon(1)} \frac{N!!}{(N+1)!!} \cdot \frac{(N-1)!!}{N!!} \cdots \frac{1!!}{2!!} \\
&= \left(\frac{\pi}{2}\right)^{\varepsilon(N)+\varepsilon(N-1)+\ldots+\varepsilon(1)} \frac{2^N}{(N+1)!!} \\
&= \left(\frac{\pi}{2}\right)^{[N/2]} \frac{2^N}{(N+1)!!}
\end{aligned}$$

Thus, we are led to the formula in the statement. □

As main particular cases of the above formula, we have:

THEOREM 2.20. *The volumes of the low-dimensional spheres are as follows:*

(1) *At $N=1$, the length of the unit interval is $V=2$.*
(2) *At $N=2$, the area of the unit disk is $V=\pi$.*
(3) *At $N=3$, the volume of the unit sphere is $V=\frac{4\pi}{3}$*
(4) *At $N=4$, the volume of the corresponding unit sphere is $V=\frac{\pi^2}{2}$.*

PROOF. Some of these results are well-known, but we can obtain all of them as particular cases of the general formula in Theorem 2.19, as follows:

(1) At $N=1$ we obtain $V=1\cdot\frac{2}{1}=2$.

(2) At $N=2$ we obtain $V=\frac{\pi}{2}\cdot\frac{4}{2}=\pi$.

(3) At $N=3$ we obtain $V=\frac{\pi}{2}\cdot\frac{8}{3}=\frac{4\pi}{3}$.

(4) At $N=4$ we obtain $V=\frac{\pi^2}{4}\cdot\frac{16}{8}=\frac{\pi^2}{2}$. □

Next, we can compute as well the volumes of general ellipsoids, as follows:

THEOREM 2.21. *The volume of an arbitrary ellipsoid in $\mathbb{R}^N$, given by*

$$\left(\frac{x_1}{a_1}\right)^2+\ldots+\left(\frac{x_N}{a_N}\right)^2=1$$

is $V=a_1\ldots a_NV_0$, with V_0 being the volume of the unit sphere in $\mathbb{R}^N$.

PROOF. We already know this at $N=2,3$, from Theorem 2.15 and Theorem 2.17, and the proof in general is similar, by suitably adapting the proof for the sphere. We will leave the computations here as an exercise, and we will come back to this in a moment, with a more conceptual argument, based on Theorem 2.19, and a change of variables. □

Finally, let us record as well the formula of the area of the sphere, as follows:

THEOREM 2.22. *The area of the unit sphere in $\mathbb{R}^N$ is given by:*

$$A=\left(\frac{\pi}{2}\right)^{[N/2]}\frac{2^N}{(N-1)!!}$$

In particular, at $N=2,3,4$ we obtain respectively $A=2\pi,4\pi,2\pi^2$.

PROOF. As shown by the pizza argument from the proof of Theorem 2.14, which extends to N dimensions, the area and volume of the sphere in $\mathbb{R}^N$ are related by:

$$A=N\cdot V$$

Together with the formula in Theorem 2.19 for V, this gives the result. □

Needless to say, in what regards the area of the general ellipsoids in $\mathbb{R}^N$, this is in general not computable, as we already know well at $N=2,3$, from the above.

2d. Change of variables

Back to the general theory, what is still missing from our bag of tools is a useful change of variable formula. Let us first recall that in one dimension we have:

PROPOSITION 2.23. *We have the change of variable formula*

$$\int_a^b f(x)dx=\int_c^d f(\varphi(t))\varphi'(t)dt$$

where $c=\varphi^{-1}(a)$ and $d=\varphi^{-1}(b)$.

PROOF. This follows with $f=F'$, from the following differentiation rule:

$$(F\varphi)'(t)=F'(\varphi(t))\varphi'(t)$$

Indeed, by integrating between c and d, we obtain the result. □

In several variables now, things are quite similar, the result being as follows:

Theorem 2.24. *Given a transformation $\varphi=(\varphi_1,\ldots,\varphi_N)$, we have*

$$\int_E f(x)dx=\int_{\varphi^{-1}(E)}f(\varphi(t))|J_\varphi(t)|dt$$

with the J_φ quantity, called Jacobian, being given by

$$J_\varphi(t)=\det\left[\left(\frac{d\varphi_i}{dx_j}(x)\right)_{ij}\right]$$

and with this generalizing the usual formula from one variable calculus.

Proof. This is something quite tricky, the idea being as follows:

(1) Observe first that this generalizes indeed the change of variable formula in 1 dimension, from Proposition 2.23, the point here being that the absolute value on the derivative appears as to compensate for the lack of explicit bounds for the integral.

(2) As a second observation, we can assume if we want, by linearity, that we are dealing with the constant function $f=1$. For this function, our formula reads:

$$vol(E)=\int_{\varphi^{-1}(E)}|J_\varphi(t)|dt$$

In terms of $D=\varphi^{-1}(E)$, this amounts in proving that we have:

$$vol(\varphi(D))=\int_D|J_\varphi(t)|dt$$

Now since this latter formula is additive with respect to D, it is enough to prove it for small cubes D. And here, as a first remark, our formula is clear for the linear maps φ, by using the definition of the determinant of real matrices, as a signed volume.

(3) However, the extension of this to the case of non-linear maps φ is something which looks non-trivial, so we will not follow this path, in what follows. So, while the above $f=1$ discussion is certainly something nice, our theorem is still in need of a proof.

(4) In order to prove the theorem, as stated, let us rather focus on the transformations used φ, instead of the functions to be integrated f. Our first claim is that the validity of the theorem is stable under taking compositions of such transformations φ.

(5) In order to prove this claim, consider a composition, as follows:

$$\varphi:E\to F\quad,\quad\psi:D\to E\quad,\quad\varphi\circ\psi:D\to F$$

Assuming that the theorem holds for φ,ψ, we have the following computation:

$$\begin{aligned}\int_F f(x)dx &= \int_E f(\varphi(s))|J_\varphi(s)|ds\\ &= \int_D f(\varphi\circ\psi(t))|J_\varphi(\psi(t))|\cdot|J_\psi(t)|dt\\ &= \int_D f(\varphi\circ\psi(t))|J_{\varphi\circ\psi}(t)|dt\end{aligned}$$

Thus, our theorem holds as well for $\varphi\circ\psi$, and we have proved our claim.

(6) Next, as a key ingredient, let us examine the case where we are in $N=2$ dimensions, and our transformation φ has one of the following special forms:

$$\varphi(x,y)=(\psi(x,y),y)\quad,\quad \varphi(x,y)=(x,\psi(x,y))$$

By symmetry, it is enough to deal with the first case. Here the Jacobian is $d\psi/dx$, and by replacing if needed $\psi\to-\psi$, we can assume that this Jacobian is positive, $d\psi/dx>0$. Now by assuming as before that $D=\varphi^{-1}(E)$ is a rectangle, $D=[a,b]\times[c,d]$, we can prove our formula by using the change of variables in 1 dimension, as follows:

$$\begin{aligned}\int_E f(s)ds &= \int_{\varphi(D)} f(x,y)dxdy\\ &= \int_c^d\int_{\psi(a,y)}^{\psi(b,y)} f(x,y)dxdy\\ &= \int_c^d\int_a^b f(\psi(x,y),y)\frac{d\psi}{dx}\,dxdy\\ &= \int_D f(\varphi(t))J_\varphi(t)dt\end{aligned}$$

(7) But with this, we can now prove the theorem, in $N=2$ dimensions. Indeed, given a transformation $\varphi=(\varphi_1,\varphi_2)$, consider the following two transformations:

$$\phi(x,y)=(\varphi_1(x,y),y)\quad,\quad \psi(x,y)=(x,\varphi_2\circ\phi^{-1}(x,y))$$

We have then $\varphi=\psi\circ\phi$, and by using (6) for ψ,ϕ, which are of the special form there, and then (3) for composing, we conclude that the theorem holds for φ, as desired.

(8) Thus, theorem proved in $N=2$ dimensions, and the extension of the above proof to arbitrary N dimensions is straightforward, that we will leave this as an exercise. □

And with this, good news, we have all the needed integration tools in our bag. To be more precise, still missing would be an analogue of the fundamental theorem of calculus, but in several variables this is something fairly complicated, to be discussed later.

As a basic application of our technology, we can recover Theorem 2.21, as follows:

THEOREM 2.25. *The volume of an arbitrary ellipsoid in $\mathbb{R}^N$, given by*

$$\left(\frac{x_1}{a_1}\right)^2+\ldots+\left(\frac{x_N}{a_N}\right)^2=1$$

is $V=a_1\ldots a_NV_0$, with V_0 being the volume of the unit sphere in $\mathbb{R}^N$.

PROOF. This is indeed something clear, with the change of variables $x_i=a_iy_i$, whose Jacobian is constant, $J=a_1\ldots a_N$. Thus, we are led to the formula in the statement. □

We will see many other applications of the change of variable formula, in the next chapter, by using polar coordinates in 2D, and spherical coordinates in general.

2e. Exercises

This was a standard chapter on integration, and as exercises, we have:

EXERCISE 2.26. *Find some other counterexamples to Fubini.*

EXERCISE 2.27. *Clarify what we said, in relation with limits of step functions.*

EXERCISE 2.28. *Check the proof of Riesz, and add topology input where needed.*

EXERCISE 2.29. *Learn a bit about inner and outer regular measures.*

EXERCISE 2.30. *Fill in the details, in the proof of the Lebesgue theorem for $\mathbb{R}^N$.*

EXERCISE 2.31. *Fill in the details too, in the proof of Fubini and Tonelli.*

EXERCISE 2.32. *Do some computations of your own, for the length of the ellipse.*

EXERCISE 2.33. *Learn some other proofs of the change of variable theorem.*

As bonus exercise, and no surprise here, learn measure theory, with full details.

CHAPTER 3

Spherical coordinates

3a. Polar coordinates

Time now do some exciting computations, with the technology that we have. In what regards the applications of the change of variable formula, these often come via:

THEOREM 3.1. *We have polar coordinates in* 2 *dimensions,*

$$\begin{cases}x &= r\cos t\\ y &= r\sin t\end{cases}$$

the corresponding Jacobian being $J = r$.

PROOF. This is elementary, the Jacobian being:

$$\begin{aligned}J &= \begin{vmatrix}\frac{d(r\cos t)}{dr} & \frac{d(r\cos t)}{dt}\\ \\ \frac{d(r\sin t)}{dr} & \frac{d(r\sin t)}{dt}\end{vmatrix}\\ &= \begin{vmatrix}\cos t & -r\sin t\\ \sin t & r\cos t\end{vmatrix}\\ &= r\cos^2 t + r\sin^2 t\\ &= r\end{aligned}$$

Thus, we have indeed the formula in the statement. □

We can now compute the Gauss integral, which is the best calculus formula ever:

THEOREM 3.2. *We have the following formula,*

$$\int_{\mathbb{R}} e^{-x^2}dx = \sqrt{\pi}$$

called Gauss integral formula.

PROOF. Let I be the above integral. By using polar coordinates, we obtain:

$$\begin{aligned} I^2 &= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{-x^2-y^2}dxdy \\ &= \int_0^{2\pi}\int_0^\infty e^{-r^2}rdrdt \\ &= 2\pi\int_0^\infty\left(-\frac{e^{-r^2}}{2}\right)'dr \\ &= 2\pi\left[0-\left(-\frac{1}{2}\right)\right] \\ &= \pi \end{aligned}$$

Thus, we are led to the formula in the statement. □

As a main application of the Gauss formula, we can now formulate:

DEFINITION 3.3. *The normal law of parameter* 1 *is the following measure:*

$$g_1 = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx$$

More generally, the normal law of parameter $t > 0$ *is the following measure:*

$$g_t = \frac{1}{\sqrt{2\pi t}}e^{-x^2/2t}dx$$

These are also called Gaussian distributions, with "g" standing for Gauss.

Observe that the above laws have indeed mass 1, as they should. This follows indeed from the Gauss formula, which gives, with $x = \sqrt{2t}\,y$:

$$\begin{aligned} \int_{\mathbb{R}} e^{-x^2/2t}dx &= \int_{\mathbb{R}} e^{-y^2}\sqrt{2t}\,dy \\ &= \sqrt{2t}\int_{\mathbb{R}} e^{-y^2}dy \\ &= \sqrt{2t}\times\sqrt{\pi} \\ &= \sqrt{2\pi t} \end{aligned}$$

Generally speaking, the normal laws appear as bit everywhere, in real life. The reasons behind this phenomenon come from the Central Limit Theorem (CLT), that we will explain in a moment, after developing some general theory. As a first result, we have:

PROPOSITION 3.4. *We have the variance formula*

$$V(g_t) = t$$

valid for any $t > 0$.

PROOF. The first moment is 0, because our normal law g_t is centered. As for the second moment, this can be computed as follows:

$$\begin{aligned} M_2 &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} x^2 e^{-x^2/2t}dx \\ &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} (tx)\left(-e^{-x^2/2t}\right)' dx \\ &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} te^{-x^2/2t}dx \\ &= t \end{aligned}$$

We conclude from this that the variance is $V = M_2 = t$. □

Here is another result, which is the key one for the study of the normal laws:

PROPOSITION 3.5. *The Fourier transform $F_f(x) = E(e^{ixf})$ of g_t is given by:*

$$F_{g_t}(x) = e^{-tx^2/2}$$

In particular, the normal laws satisfy $g_s * g_t = g_{s+t}$, for any $s, t > 0$.

PROOF. The Fourier transform formula can be established as follows:

$$\begin{aligned} F_{g_t}(x) &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} e^{-y^2/2t+ixy}dy \\ &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} e^{-(y/\sqrt{2t}-\sqrt{t/2}ix)^2-tx^2/2}dy \\ &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}} e^{-z^2-tx^2/2}\sqrt{2t}dz \\ &= \frac{1}{\sqrt{\pi}}e^{-tx^2/2}\int_{\mathbb{R}} e^{-z^2}dz \\ &= \frac{1}{\sqrt{\pi}}e^{-tx^2/2}\cdot\sqrt{\pi} \\ &= e^{-tx^2/2} \end{aligned}$$

As for the last assertion, this follows from the fact that $\log F_{g_t}$ is linear in t. □

We are now ready to state and prove the CLT, as follows:

THEOREM 3.6 (CLT). *Given random variables $f_1, f_2, f_3, \ldots \in L^\infty(X)$ which are i.i.d., centered, and with variance $t > 0$, we have, with $n \to \infty$, in moments,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i \sim g_t$$

where g_t is the Gaussian law of parameter t, having as density $\frac{1}{\sqrt{2\pi t}}e^{-y^2/2t}dy$.

PROOF. In terms of moments, the Fourier transform is given by:

$$\begin{aligned} F_f(x) &= E\left(\sum_{k=0}^\infty \frac{(ixf)^k}{k!}\right) \\ &= \sum_{k=0}^\infty \frac{(ix)^k E(f^k)}{k!} \\ &= \sum_{k=0}^\infty \frac{i^k M_k(f)}{k!}\, x^k \end{aligned}$$

Thus, the Fourier transform of the variable in the statement is:

$$\begin{aligned} F(x) &= \left[F_f\left(\frac{x}{\sqrt{n}}\right)\right]^n \\ &= \left[1 - \frac{tx^2}{2n} + O(n^{-2})\right]^n \\ &\simeq \left[1 - \frac{tx^2}{2n}\right]^n \\ &\simeq e^{-tx^2/2} \end{aligned}$$

But this latter function being the Fourier transform of g_t, we obtain the result. □

Let us discuss as well the complex analogues of the above, with a notion of complex normal, or Gaussian law. To start with, we have the following definition:

DEFINITION 3.7. *The complex normal, or Gaussian law of parameter $t > 0$ is*

$$G_t = law\left(\frac{1}{\sqrt{2}}(a + ib)\right)$$

where a, b are independent, each following the law g_t.

In short, the complex normal laws appear as natural complexifications of the real normal laws. As in the real case, these measures form convolution semigroups:

PROPOSITION 3.8. *The complex Gaussian laws have the property*

$$G_s * G_t = G_{s+t}$$

for any $s, t > 0$, and so they form a convolution semigroup.

PROOF. This follows indeed from the real result, namely $g_s * g_t = g_{s+t}$, established in Proposition 3.5, simply by taking the real and imaginary parts. □

We have as well the following complex analogue of the CLT:

THEOREM 3.9 (CCLT). *Given complex variables $f_1, f_2, f_3, \ldots \in L^\infty(X)$ which are i.i.d., centered, and with common variance $t > 0$, we have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i \sim G_t$$

with $n \to \infty$, in moments.

PROOF. This follows indeed from the real CLT, established in Theorem 3.6, simply by taking the real and imaginary parts of all the variables involved. □

We will be back to this, with some applications, at the end of this chapter.

3b. Spherical coordinates

Moving now to 3 dimensions, we have here the following result:

THEOREM 3.10. *We have spherical coordinates in 3 dimensions,*

$$\begin{cases} x &= r\cos s \\ y &= r\sin s\cos t \\ z &= r\sin s\sin t \end{cases}$$

the corresponding Jacobian being $J(r, s, t) = r^2 \sin s$.

PROOF. The fact that we have indeed spherical coordinates is clear. Regarding now the Jacobian, this is given by the following formula:

$$\begin{aligned}
& J(r,s,t) \\
&= \begin{vmatrix} \cos s & -r\sin s & 0 \\ \sin s\cos t & r\cos s\cos t & -r\sin s\sin t \\ \sin s\sin t & r\cos s\sin t & r\sin s\cos t \end{vmatrix} \\
&= r^2\sin s\sin t\begin{vmatrix} \cos s & -r\sin s \\ \sin s\sin t & r\cos s\sin t \end{vmatrix} + r\sin s\cos t\begin{vmatrix} \cos s & -r\sin s \\ \sin s\cos t & r\cos s\cos t \end{vmatrix} \\
&= r\sin s\sin^2 t\begin{vmatrix} \cos s & -r\sin s \\ \sin s & r\cos s \end{vmatrix} + r\sin s\cos^2 t\begin{vmatrix} \cos s & -r\sin s \\ \sin s & r\cos s \end{vmatrix} \\
&= r\sin s(\sin^2 t + \cos^2 t)\begin{vmatrix} \cos s & -r\sin s \\ \sin s & r\cos s \end{vmatrix} \\
&= r\sin s \times 1 \times r \\
&= r^2\sin s
\end{aligned}$$

Thus, we have indeed the formula in the statement. □

Let us work out now the general spherical coordinate formula, in arbitrary N dimensions. The formula here, which generalizes those at $N = 2, 3$, is as follows:

THEOREM 3.11. *We have spherical coordinates in N dimensions,*

$$\begin{cases}x_1 &= r\cos t_1\\ x_2 &= r\sin t_1\cos t_2\\ \vdots\\ x_{N-1} &= r\sin t_1\sin t_2\ldots\sin t_{N-2}\cos t_{N-1}\\ x_N &= r\sin t_1\sin t_2\ldots\sin t_{N-2}\sin t_{N-1}\end{cases}$$

the corresponding Jacobian being given by the following formula,

$$J(r,t)=r^{N-1}\sin^{N-2}t_1\sin^{N-3}t_2\,\ldots\,\sin^2 t_{N-3}\sin t_{N-2}$$

and with this generalizing the known formulae at $N=2,3$.

PROOF. As before, the fact that we have spherical coordinates is clear. Regarding now the Jacobian, also as before, by developing over the last column, we have:

$$\begin{aligned}J_N &= r\sin t_1\ldots\sin t_{N-2}\sin t_{N-1}\times\sin t_{N-1}J_{N-1}\\ &+ r\sin t_1\ldots\sin t_{N-2}\cos t_{N-1}\times\cos t_{N-1}J_{N-1}\\ &= r\sin t_1\ldots\sin t_{N-2}(\sin^2 t_{N-1}+\cos^2 t_{N-1})J_{N-1}\\ &= r\sin t_1\ldots\sin t_{N-2}J_{N-1}\end{aligned}$$

Thus, we obtain the formula in the statement, by recurrence. □

As a comment here, the above convention for spherical coordinates is one among many, designed to best work in arbitrary N dimensions. Also, in what regards the precise range of the angles $t_1,\ldots,t_{N-1}$, we will leave this to you, as an instructive exercise.

As an application, we can recompute the volumes of spheres, as follows:

THEOREM 3.12. *The volume of the unit sphere in $\mathbb{R}^N$ is given by*

$$V=\left(\frac{\pi}{2}\right)^{[N/2]}\frac{2^N}{(N+1)!!}$$

with our usual convention $N!!=(N-1)(N-3)(N-5)\ldots$

PROOF. Let us denote by B^+ the positive part of the unit sphere, or rather unit ball B, obtained by cutting this unit ball in 2^N parts. At the level of volumes, we have:

$$V=2^NV^+$$

We have the following computation, using spherical coordinates:

$$\begin{aligned}
V^+ &= \int_{B^+} 1 \\
&= \int_0^1 \int_0^{\pi/2} \ldots \int_0^{\pi/2} r^{N-1} \sin^{N-2} t_1 \ldots \sin t_{N-2}\, dr dt_1 \ldots dt_{N-1} \\
&= \int_0^1 r^{N-1}\, dr \int_0^{\pi/2} \sin^{N-2} t_1\, dt_1 \ldots \int_0^{\pi/2} \sin t_{N-2} dt_{N-2} \int_0^{\pi/2} 1 dt_{N-1} \\
&= \frac{1}{N} \times \left(\frac{\pi}{2}\right)^{[N/2]} \times \frac{(N-2)!!}{(N-1)!!} \cdot \frac{(N-3)!!}{(N-2)!!} \ldots \frac{2!!}{3!!} \cdot \frac{1!!}{2!!} \cdot 1 \\
&= \frac{1}{N} \times \left(\frac{\pi}{2}\right)^{[N/2]} \times \frac{1}{(N-1)!!} \\
&= \left(\frac{\pi}{2}\right)^{[N/2]} \frac{1}{(N+1)!!}
\end{aligned}$$

Thus, we obtain the formula in the statement. □

Let us record as well the asymptotics, obtained via Stirling, as follows:

THEOREM 3.13. *The volume of the unit sphere in $\mathbb{R}^N$ is given by*

$$V \simeq \left(\frac{2\pi e}{N}\right)^{N/2} \frac{1}{\sqrt{\pi N}}$$

in the $N \to \infty$ limit.

PROOF. This is something very standard, the idea being as follows:

(1) We use the exact formula found in Theorem 3.12, namely:

$$V = \left(\frac{\pi}{2}\right)^{[N/2]} \frac{2^N}{(N+1)!!}$$

(2) But the double factorials can be estimated by using the Stirling formula. Indeed, in the case where $N = 2K$ is even, we have the following computation:

$$\begin{aligned}
(N+1)!! &= 2^K K! \\
&\simeq \left(\frac{2K}{e}\right)^K \sqrt{2\pi K} \\
&= \left(\frac{N}{e}\right)^{N/2} \sqrt{\pi N}
\end{aligned}$$

As for the case where $N=2K-1$ is odd, here the estimate goes as follows:

$$\begin{aligned}
(N+1)!! &= \frac{(2K)!}{2^K K!}\\
&\simeq \frac{1}{2^K}\left(\frac{2K}{e}\right)^{2K}\sqrt{4\pi K}\left(\frac{e}{K}\right)^K\frac{1}{\sqrt{2\pi K}}\\
&= \left(\frac{2K}{e}\right)^K\sqrt{2}\\
&= \left(\frac{N+1}{e}\right)^{(N+1)/2}\sqrt{2}\\
&= \left(\frac{N}{e}\right)^{N/2}\left(\frac{N+1}{N}\right)^{N/2}\sqrt{\frac{N+1}{e}}\cdot\sqrt{2}\\
&\simeq \left(\frac{N}{e}\right)^{N/2}\sqrt{e}\cdot\sqrt{\frac{N}{e}}\cdot\sqrt{2}\\
&= \left(\frac{N}{e}\right)^{N/2}\sqrt{2N}
\end{aligned}$$

(3) Now back to the spheres, when N is even, the estimate goes as follows:

$$\begin{aligned}
V &= \left(\frac{\pi}{2}\right)^{N/2}\frac{2^N}{(N+1)!!}\\
&\simeq \left(\frac{\pi}{2}\right)^{N/2}2^N\left(\frac{e}{N}\right)^{N/2}\frac{1}{\sqrt{\pi N}}\\
&= \left(\frac{2\pi e}{N}\right)^{N/2}\frac{1}{\sqrt{\pi N}}
\end{aligned}$$

As for the case where N is odd, here the estimate goes as follows:

$$\begin{aligned}
V &= \left(\frac{\pi}{2}\right)^{(N-1)/2}\frac{2^N}{(N+1)!!}\\
&\simeq \left(\frac{\pi}{2}\right)^{(N-1)/2}2^N\left(\frac{e}{N}\right)^{N/2}\frac{1}{\sqrt{2N}}\\
&= \sqrt{\frac{2}{\pi}}\left(\frac{2\pi e}{N}\right)^{N/2}\frac{1}{\sqrt{2N}}\\
&= \left(\frac{2\pi e}{N}\right)^{N/2}\frac{1}{\sqrt{\pi N}}
\end{aligned}$$

Thus, we are led to the uniform formula in the statement. □

Getting back now to our main result so far, Theorem 3.12, we can compute in the same way the area of the sphere, the result being as follows:

THEOREM 3.14. *The area of the unit sphere in $\mathbb{R}^N$ is given by*

$$A = \left(\frac{\pi}{2}\right)^{[N/2]} \frac{2^N}{(N-1)!!}$$

with the our usual convention for double factorials, namely:

$$N!! = (N-1)(N-3)(N-5)\ldots$$

In particular, at $N = 2, 3, 4$ we obtain respectively $A = 2\pi, 4\pi, 2\pi^2$.

PROOF. Regarding the first assertion, there is no need to compute again, because the formula in the statement can be deduced from Theorem 3.12, as follows:

(1) We can either use the "pizza" argument from chapter 2, which shows that the area and volume of the sphere in $\mathbb{R}^N$ are related by the following formula:

$$A = N \cdot V$$

Together with the formula in Theorem 3.12 for V, this gives the result.

(2) Or, we can start the computation in the same way as we started the proof of Theorem 3.12, the beginning of this computation being as follows:

$$vol(S^+) = \int_0^{\pi/2} \ldots \int_0^{\pi/2} \sin^{N-2} t_1 \ldots \sin t_{N-2}\, dt_1 \ldots dt_{N-1}$$

Now by comparing with the beginning of the proof of Theorem 3.12, the only thing that changes is the following quantity, which now dissapears:

$$\int_0^1 r^{N-1}\, dr = \frac{1}{N}$$

Thus, we have $vol(S^+) = N \cdot vol(B^+)$, and so we obtain the following formula:

$$vol(S) = N \cdot vol(B)$$

But this means $A = N \cdot V$, and together with the formula in Theorem 3.12 for V, this gives the result. As for the last assertion, this can be either worked out directly, or deduced from the results for volumes that we have so far, by multiplying by N. $\square$

3c. Polynomial integrals

Let us discuss now the computation of arbitrary integrals over the sphere. We will need a technical result extending the Wallis formula from chapter 2, as follows:

THEOREM 3.15. *We have the following formula,*

$$\int_0^{\pi/2}\cos^p t\sin^q t\,dt=\left(\frac{\pi}{2}\right)^{\varepsilon(p)\varepsilon(q)}\frac{p!!q!!}{(p+q+1)!!}$$

where $\varepsilon(p)=1$ *if* p *is even, and* $\varepsilon(p)=0$ *if* p *is odd, and where*

$$m!!=(m-1)(m-3)(m-5)\ldots$$

with the product ending at 2 *if* m *is odd, and ending at* 1 *if* m *is even.*

PROOF. We use the same idea as in chapter 2. Let I_{pq} be the integral in the statement. In order to do the partial integration, observe that we have:

$$\begin{aligned}(\cos^p t\sin^q t)' &= p\cos^{p-1}t(-\sin t)\sin^q t\\ &+ \cos^p t\cdot q\sin^{q-1}t\cos t\\ &= -p\cos^{p-1}t\sin^{q+1}t+q\cos^{p+1}t\sin^{q-1}t\end{aligned}$$

By integrating between 0 and $\pi/2$, we obtain, for $p,q>0$:

$$pI_{p-1,q+1}=qI_{p+1,q-1}$$

Thus, we can compute I_{pq} by recurrence. When q is even we have:

$$\begin{aligned}I_{pq} &= \frac{q-1}{p+1}I_{p+2,q-2}\\ &= \frac{q-1}{p+1}\cdot\frac{q-3}{p+3}I_{p+4,q-4}\\ &= \frac{q-1}{p+1}\cdot\frac{q-3}{p+3}\cdot\frac{q-5}{p+5}I_{p+6,q-6}\\ &= \vdots\\ &= \frac{p!!q!!}{(p+q)!!}I_{p+q}\end{aligned}$$

But the last term comes from the formula in chapter 2, and we obtain the result:

$$\begin{aligned}I_{pq} &= \frac{p!!q!!}{(p+q)!!}I_{p+q}\\ &= \frac{p!!q!!}{(p+q)!!}\left(\frac{\pi}{2}\right)^{\varepsilon(p+q)}\frac{(p+q)!!}{(p+q+1)!!}\\ &= \left(\frac{\pi}{2}\right)^{\varepsilon(p)\varepsilon(q)}\frac{p!!q!!}{(p+q+1)!!}\end{aligned}$$

Observe that this gives the result for p even as well, by symmetry. Indeed, we have $I_{pq}=I_{qp}$, by using the following change of variables:

$$t=\frac{\pi}{2}-s$$

In the remaining case now, where both p,q are odd, we can use once again the formula $pI_{p-1,q+1}=qI_{p+1,q-1}$ established above, and the recurrence goes as follows:

$$\begin{aligned}
I_{pq} &= \frac{q-1}{p+1}\,I_{p+2,q-2}\\
&= \frac{q-1}{p+1}\cdot\frac{q-3}{p+3}\,I_{p+4,q-4}\\
&= \frac{q-1}{p+1}\cdot\frac{q-3}{p+3}\cdot\frac{q-5}{p+5}\,I_{p+6,q-6}\\
&= \ \vdots\\
&= \frac{p!!q!!}{(p+q-1)!!}\,I_{p+q-1,1}
\end{aligned}$$

In order to compute the last term, observe that we have:

$$\begin{aligned}
I_{p1} &= \int_0^{\pi/2}\cos^p t\sin t\,dt\\
&= -\frac{1}{p+1}\int_0^{\pi/2}(\cos^{p+1}t)'\,dt\\
&= \frac{1}{p+1}
\end{aligned}$$

Thus, we can finish our computation in the case p,q odd, as follows:

$$\begin{aligned}
I_{pq} &= \frac{p!!q!!}{(p+q-1)!!}\,I_{p+q-1,1}\\
&= \frac{p!!q!!}{(p+q-1)!!}\cdot\frac{1}{p+q}\\
&= \frac{p!!q!!}{(p+q+1)!!}
\end{aligned}$$

Thus, we obtain the formula in the statement, the exponent of $\pi/2$ appearing there being $\varepsilon(p)\varepsilon(q)=0\cdot 0=0$ in the present case, and this finishes the proof. $\square$

We can now integrate over the spheres, as follows:

THEOREM 3.16. *The polynomial integrals over the unit sphere $S^{N-1}_{\mathbb{R}}\subset\mathbb{R}^N$, with respect to the normalized, mass 1 measure, are given by the following formula,*

$$\int_{S^{N-1}_{\mathbb{R}}} x_1^{k_1}\ldots x_N^{k_N}\,dx=\frac{(N-1)!!k_1!!\ldots k_N!!}{(N+\Sigma k_i-1)!!}$$

valid when all exponents k_i are even. If an exponent k_i is odd, the integral vanishes.

PROOF. Assume first that one of the exponents k_i is odd. We can make then the following change of variables, which shows that the integral in the statement vanishes:

$$x_i \to -x_i$$

Assume now that all exponents k_i are even. As a first observation, the result holds indeed at $N = 2$, due to the formula from Theorem 3.15, which reads:

$$\int_0^{\pi/2} \cos^p t \sin^q t \, dt = \left(\frac{\pi}{2}\right)^{\varepsilon(p)\varepsilon(q)} \frac{p!!q!!}{(p+q+1)!!} = \frac{p!!q!!}{(p+q+1)!!}$$

In the general case now, where the dimension $N \in \mathbb{N}$ is arbitrary, the integral in the statement can be written in spherical coordinates, as follows:

$$I = \frac{2^N}{A} \int_0^{\pi/2} \ldots \int_0^{\pi/2} x_1^{k_1} \ldots x_N^{k_N} J \, dt_1 \ldots dt_{N-1}$$

Here A is the area of the sphere, J is the Jacobian, and the 2^N factor comes from the restriction to the $1/2^N$ part of the sphere where all the coordinates are positive. According to Theorem 3.14, the normalization constant in front of the integral is:

$$\frac{2^N}{A} = \left(\frac{2}{\pi}\right)^{[N/2]} (N-1)!!$$

As for the unnormalized integral, this is given by:

$$\begin{aligned} I' = \int_0^{\pi/2} \ldots \int_0^{\pi/2} \quad & (\cos t_1)^{k_1} (\sin t_1 \cos t_2)^{k_2} \\ & \vdots \\ & (\sin t_1 \sin t_2 \ldots \sin t_{N-2} \cos t_{N-1})^{k_{N-1}} \\ & (\sin t_1 \sin t_2 \ldots \sin t_{N-2} \sin t_{N-1})^{k_N} \\ & \sin^{N-2} t_1 \sin^{N-3} t_2 \ldots \sin^2 t_{N-3} \sin t_{N-2} \\ & dt_1 \ldots dt_{N-1} \end{aligned}$$

By rearranging the terms, we obtain:

$$\begin{aligned} I' &= \int_0^{\pi/2} \cos^{k_1} t_1 \sin^{k_2+\ldots+k_N+N-2} t_1\, dt_1 \\ &\quad \int_0^{\pi/2} \cos^{k_2} t_2 \sin^{k_3+\ldots+k_N+N-3} t_2\, dt_2 \\ &\quad \vdots \\ &\quad \int_0^{\pi/2} \cos^{k_{N-2}} t_{N-2} \sin^{k_{N-1}+k_N+1} t_{N-2}\, dt_{N-2} \\ &\quad \int_0^{\pi/2} \cos^{k_{N-1}} t_{N-1} \sin^{k_N} t_{N-1}\, dt_{N-1} \end{aligned}$$

Now by using the above-mentioned formula at $N=2$, this gives:

$$\begin{aligned} I' &= \frac{k_1!!(k_2+\ldots+k_N+N-2)!!}{(k_1+\ldots+k_N+N-1)!!}\left(\frac{\pi}{2}\right)^{\varepsilon(N-2)} \\ &\quad \frac{k_2!!(k_3+\ldots+k_N+N-3)!!}{(k_2+\ldots+k_N+N-2)!!}\left(\frac{\pi}{2}\right)^{\varepsilon(N-3)} \\ &\quad \vdots \\ &\quad \frac{k_{N-2}!!(k_{N-1}+k_N+1)!!}{(k_{N-2}+k_{N-1}+l_N+2)!!}\left(\frac{\pi}{2}\right)^{\varepsilon(1)} \\ &\quad \frac{k_{N-1}!!k_N!!}{(k_{N-1}+k_N+1)!!}\left(\frac{\pi}{2}\right)^{\varepsilon(0)} \end{aligned}$$

Now let F be the part involving the double factorials, and P be the part involving the powers of $\pi/2$, so that $I' = F \cdot P$. Regarding F, by cancelling terms we have:

$$F = \frac{k_1!!\ldots k_N!!}{(\Sigma k_i + N - 1)!!}$$

As in what regards P, by summing the exponents, we obtain $P = \left(\frac{\pi}{2}\right)^{[N/2]}$. We can now put everything together, and we obtain:

$$\begin{aligned} I &= \frac{2^N}{A} \times F \times P \\ &= \left(\frac{2}{\pi}\right)^{[N/2]} (N-1)!! \times \frac{k_1!!\ldots k_N!!}{(\Sigma k_i + N - 1)!!} \times \left(\frac{\pi}{2}\right)^{[N/2]} \\ &= \frac{(N-1)!!k_1!!\ldots k_N!!}{(\Sigma k_i + N - 1)!!} \end{aligned}$$

Thus, we are led to the conclusion in the statement. □

We have the following useful generalization of the above formula:

THEOREM 3.17. *We have the following integration formula over $S^{N-1}_{\mathbb{R}} \subset \mathbb{R}^N$, with respect to the normalized, mass 1 measure, valid for any exponents $k_i \in \mathbb{N}$,*

$$\int_{S^{N-1}_{\mathbb{R}}} |x_1^{k_1} \ldots x_N^{k_N}|\, dx = \left(\frac{2}{\pi}\right)^{\Sigma(k_1,\ldots,k_N)} \frac{(N-1)!!k_1!!\ldots k_N!!}{(N+\Sigma k_i - 1)!!}$$

with $\Sigma = [odds/2]$ if N is odd and $\Sigma = [(odds+1)/2]$ if N is even, where "odds" denotes the number of odd numbers in the sequence $k_1, \ldots, k_N$.

PROOF. As before, the formula holds at $N = 2$, due to Theorem 3.15. In general, the integral in the statement can be written in spherical coordinates, as follows:

$$I = \frac{2^N}{A} \int_0^{\pi/2} \ldots \int_0^{\pi/2} x_1^{k_1} \ldots x_N^{k_N}\, J\, dt_1 \ldots dt_{N-1}$$

Here A is the area of the sphere, J is the Jacobian, and the 2^N factor comes from the restriction to the $1/2^N$ part of the sphere where all the coordinates are positive. The normalization constant in front of the integral is, as before:

$$\frac{2^N}{A} = \left(\frac{2}{\pi}\right)^{[N/2]} (N-1)!!$$

As for the unnormalized integral, this can be written as before, as follows:

$$\begin{aligned} I' \;=\; & \int_0^{\pi/2} \cos^{k_1} t_1 \sin^{k_2+\ldots+k_N+N-2} t_1\, dt_1 \\ & \int_0^{\pi/2} \cos^{k_2} t_2 \sin^{k_3+\ldots+k_N+N-3} t_2\, dt_2 \\ & \vdots \\ & \int_0^{\pi/2} \cos^{k_{N-2}} t_{N-2} \sin^{k_{N-1}+k_N+1} t_{N-2}\, dt_{N-2} \\ & \int_0^{\pi/2} \cos^{k_{N-1}} t_{N-1} \sin^{k_N} t_{N-1}\, dt_{N-1} \end{aligned}$$

Now by using the formula at $N=2$, we get:

$$\begin{aligned}
I' &= \frac{\pi}{2}\cdot\frac{k_1!!(k_2+\ldots+k_N+N-2)!!}{(k_1+\ldots+k_N+N-1)!!}\left(\frac{2}{\pi}\right)^{\delta(k_1,k_2+\ldots+k_N+N-2)}\\
&\quad \frac{\pi}{2}\cdot\frac{k_2!!(k_3+\ldots+k_N+N-3)!!}{(k_2+\ldots+k_N+N-2)!!}\left(\frac{2}{\pi}\right)^{\delta(k_2,k_3+\ldots+k_N+N-3)}\\
&\quad \vdots\\
&\quad \frac{\pi}{2}\cdot\frac{k_{N-2}!!(k_{N-1}+k_N+1)!!}{(k_{N-2}+k_{N-1}+k_N+2)!!}\left(\frac{2}{\pi}\right)^{\delta(k_{N-2},k_{N-1}+k_N+1)}\\
&\quad \frac{\pi}{2}\cdot\frac{k_{N-1}!!k_N!!}{(k_{N-1}+k_N+1)!!}\left(\frac{2}{\pi}\right)^{\delta(k_{N-1},k_N)}
\end{aligned}$$

In order to compute this quantity, let us denote by F the part involving the double factorials, and by P the part involving the powers of $\pi/2$, so that we have:

$$I' = F\cdot P$$

Regarding F, there are many cancellations there, and we end up with:

$$F = \frac{k_1!!\ldots k_N!!}{(\Sigma k_i+N-1)!!}$$

As in what regards P, the δ exponents on the right sum up to the following number:

$$\Delta(k_1,\ldots,k_N) = \sum_{i=1}^{N-1}\delta(k_i,k_{i+1}+\ldots+k_N+N-i-1)$$

In other words, with this notation, the above formula reads:

$$\begin{aligned}
I' &= \left(\frac{\pi}{2}\right)^{N-1}\frac{k_1!!k_2!!\ldots k_N!!}{(k_1+\ldots+k_N+N-1)!!}\left(\frac{2}{\pi}\right)^{\Delta(k_1,\ldots,k_N)}\\
&= \left(\frac{2}{\pi}\right)^{\Delta(k_1,\ldots,k_N)-N+1}\frac{k_1!!k_2!!\ldots k_N!!}{(k_1+\ldots+k_N+N-1)!!}\\
&= \left(\frac{2}{\pi}\right)^{\Sigma(k_1,\ldots,k_N)-[N/2]}\frac{k_1!!k_2!!\ldots k_N!!}{(k_1+\ldots+k_N+N-1)!!}
\end{aligned}$$

To be more precise, the formula relating Δ to Σ follows from a number of simple observations, the first of which being the fact that, due to obvious parity reasons, the sequence of δ numbers appearing in the definition of Δ cannot contain two consecutive zeroes. Together with $I=(2^N/V)I'$, this gives the formula in the statement. □

Let us discuss as well the complex versions of the above results. We have the following variation of the formula in Theorem 3.16, dealing with the complex sphere:

THEOREM 3.18. *We have the following integration formula over the complex sphere* $S^{N-1}_\mathbb{C}\subset\mathbb{C}^N$*, with respect to the normalized uniform measure,*

$$\int_{S^{N-1}_\mathbb{C}}|z_1|^{2k_1}\ldots|z_N|^{2k_N}\,dz=\frac{(N-1)!k_1!\ldots k_n!}{(N+\sum k_i-1)!}$$

valid for any exponents $k_i\in\mathbb{N}$*. As for the other polynomial integrals in* $z_1,\ldots,z_N$ *and their conjugates* $\bar{z}_1,\ldots,\bar{z}_N$*, these all vanish.*

PROOF. Consider an arbitrary polynomial integral over $S^{N-1}_\mathbb{C}$, containing the same number of plain and conjugated variables, as to not vanish trivially, written as follows:

$$I=\int_{S^{N-1}_\mathbb{C}}z_{i_1}\bar{z}_{i_2}\ldots z_{i_{2k-1}}\bar{z}_{i_{2k}}\,dz$$

By using transformations of type $p\to\lambda p$ with $|\lambda|=1$, we see that this integral I vanishes, unless each z_a appears as many times as $\bar{z}_a$ does, and this gives the last assertion. So, assume now that we are in the non-vanishing case. Then the k_a copies of z_a and the k_a copies of $\bar{z}_a$ produce by multiplication a factor $|z_a|^{2k_a}$, so we have:

$$I=\int_{S^{N-1}_\mathbb{C}}|z_1|^{2k_1}\ldots|z_N|^{2k_N}\,dz$$

Now by using the standard identification $S^{N-1}_\mathbb{C}\simeq S^{2N-1}_\mathbb{R}$, we obtain:

$$\begin{aligned}I&=\int_{S^{2N-1}_\mathbb{R}}(x_1^2+y_1^2)^{k_1}\ldots(x_N^2+y_N^2)^{k_N}\,d(x,y)\\&=\sum_{r_1\ldots r_N}\binom{k_1}{r_1}\ldots\binom{k_N}{r_N}\int_{S^{2N-1}_\mathbb{R}}x_1^{2k_1-2r_1}y_1^{2r_1}\ldots x_N^{2k_N-2r_N}y_N^{2r_N}\,d(x,y)\end{aligned}$$

By using the formula in Theorem 3.16, we obtain:

$$\begin{aligned}&I\\=&\sum_{r_1\ldots r_N}\binom{k_1}{r_1}\ldots\binom{k_N}{r_N}\frac{(2N-1)!!(2r_1)!!\ldots(2r_N)!!(2k_1-2r_1)!!\ldots(2k_N-2r_N)!!}{(2N+2\sum k_i-1)!!}\\=&\sum_{r_1\ldots r_N}\binom{k_1}{r_1}\ldots\binom{k_N}{r_N}\frac{2^{N-1}(N-1)!\prod(2r_i)!/(2^{r_i}r_i!)\prod(2k_i-2r_i)!/(2^{k_i-r_i}(k_i-r_i)!)}{2^{N+\sum k_i-1}(N+\sum k_i-1)!}\\=&\sum_{r_1\ldots r_N}\binom{k_1}{r_1}\ldots\binom{k_N}{r_N}\frac{(N-1)!(2r_1)!\ldots(2r_N)!(2k_1-2r_1)!\ldots(2k_N-2r_N)!}{4^{\sum k_i}(N+\sum k_i-1)!r_1!\ldots r_N!(k_1-r_1)!\ldots(k_N-r_N)!}\end{aligned}$$

Now observe that can rewrite this quantity in the following way:

$$\begin{aligned}
I &= \sum_{r_1\ldots r_N}\frac{k_1!\ldots k_N!(N-1)!(2r_1)!\ldots(2r_N)!(2k_1-2r_1)!\ldots(2k_N-2r_N)!}{4^{\sum k_i}(N+\sum k_i-1)!(r_1!\ldots r_N!(k_1-r_1)!\ldots(k_N-r_N)!)^2}\\
&= \sum_{r_1}\binom{2r_1}{r_1}\binom{2k_1-2r_1}{k_1-r_1}\ldots\sum_{r_N}\binom{2r_N}{r_N}\binom{2k_N-2r_N}{k_N-r_N}\frac{(N-1)!k_1!\ldots k_N!}{4^{\sum k_i}(N+\sum k_i-1)!}\\
&= 4^{k_1}\times\ldots\times 4^{k_N}\times\frac{(N-1)!k_1!\ldots k_N!}{4^{\sum k_i}(N+\sum k_i-1)!}\\
&= \frac{(N-1)!k_1!\ldots k_N!}{(N+\sum k_i-1)!}
\end{aligned}$$

Thus, we are led to the formula in the statement. □

3d. Hyperspherical laws

In order to process the above results, in the $N\to\infty$ limit, we need to know more about the normal laws. In the real case, we first have the following result:

Proposition 3.19. *The even moments of the normal law are the numbers*

$$M_k(g_t)=t^{k/2}\times k!!$$

where $k!!=(k-1)(k-3)(k-5)\ldots$, and the odd moments vanish.

Proof. We have the following computation, valid for any integer $k\in\mathbb{N}$:

$$\begin{aligned}
M_k &= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}}y^ke^{-y^2/2t}dy\\
&= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}}(ty^{k-1})\left(-e^{-y^2/2t}\right)'dy\\
&= \frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}}t(k-1)y^{k-2}e^{-y^2/2t}dy\\
&= t(k-1)\times\frac{1}{\sqrt{2\pi t}}\int_{\mathbb{R}}y^{k-2}e^{-y^2/2t}dy\\
&= t(k-1)M_{k-2}
\end{aligned}$$

Thus by recurrence, we are led to the formula in the statement. □

We have the following alternative formulation of the above result:

Proposition 3.20. *The moments of the normal law are the numbers*

$$M_k(g_t)=t^{k/2}|P_2(k)|$$

where $P_2(k)$ is the set of pairings of $\{1,\ldots,k\}$.

Proof. Let us count the pairings of $\{1,\ldots,k\}$. In order to have such a pairing, we must pair 1 with one of the numbers $2,\ldots,k$, and then use a pairing of the remaining $k-2$ numbers. Thus, we have the following recurrence formula:

$$|P_2(k)|=(k-1)|P_2(k-2)|$$

As for the initial data, this is $P_1=0$, $P_2=1$. Thus, we are led to the result. □

We are not done yet, and here is one more improvement of the above:

Theorem 3.21. *The moments of the normal law are the numbers*

$$M_k(g_t)=\sum_{\pi\in P_2(k)}t^{|\pi|}$$

where $P_2(k)$ is the set of pairings of $\{1,\ldots,k\}$, and $|.|$ is the number of blocks.

Proof. This follows indeed from Proposition 3.20, because the number of blocks of a pairing of $\{1,\ldots,k\}$ is trivially $k/2$, independently of the pairing. □

We can now go back to the spherical integrals, and we have:

Theorem 3.22. *The moments of the hyperspherical variables are*

$$\int_{S^{N-1}_\mathbb{R}}x_i^p\,dx=\frac{(N-1)!!p!!}{(N+p-1)!!}$$

and the rescaled variables $y_i=\sqrt{N}x_i$ become normal and independent with $N\to\infty$.

Proof. The moment formula in the statement follows from the general formula from Theorem 3.16. As a consequence, with $N\to\infty$ we have the following estimate:

$$\begin{aligned}\int_{S^{N-1}_\mathbb{R}}x_i^p\,dx&\simeq N^{-p/2}\times p!!\\&=N^{-p/2}M_p(g_1)\end{aligned}$$

Thus, the rescaled variables $\sqrt{N}x_i$ become normal with $N\to\infty$, as claimed. As for the proof of the asymptotic independence, this is standard too, once again by using the formula in Theorem 3.16. Indeed, the joint moments of $x_1,\ldots,x_N$ are given by:

$$\begin{aligned}\int_{S^{N-1}_\mathbb{R}}x_1^{k_1}\ldots x_N^{k_N}\,dx&=\frac{(N-1)!!k_1!!\ldots k_N!!}{(N+\Sigma k_i-1)!!}\\&\simeq N^{-\Sigma k_i}\times k_1!!\ldots k_N!!\end{aligned}$$

By rescaling, the joint moments of the variables $y_i=\sqrt{N}x_i$ are given by:

$$\int_{S^{N-1}_\mathbb{R}}y_1^{k_1}\ldots y_N^{k_N}\,dx\simeq k_1!!\ldots k_N!!$$

Thus, we have multiplicativity, and so independence with $N\to\infty$, as claimed. □

Importantly, we can recover the normal laws as well in connection with the rotation groups. Indeed, we have the following reformulation of Theorem 3.22:

THEOREM 3.23. *We have the integration formula*

$$\int_{O_N} U_{ij}^p \, dU = \frac{(N-1)!!p!!}{(N+p-1)!!}$$

and the rescaled variables $V_{ij} = \sqrt{N} U_{ij}$ *become normal and independent with* $N \to \infty$.

PROOF. We use the basic fact that the rotations $U \in O_N$ act on the points of the real sphere $z \in S^{N-1}_{\mathbb{R}}$, with the stabilizer of $z = (1, 0, \ldots, 0)$ being the subgroup $O_{N-1} \subset O_N$. In algebraic terms, this gives an identification as follows:

$$S^{N-1}_{\mathbb{R}} = O_N / O_{N-1}$$

In functional analytic terms, this result provides us with an embedding as follows, for any i, which makes correspond the respective integration functionals:

$$C(S^{N-1}_{\mathbb{R}}) \subset C(O_N) \quad , \quad x_i \to U_{1i}$$

With this identification made, the result follows from Theorem 3.22. □

In the complex case now, let us first formulate the following definition:

DEFINITION 3.24. *The moments a complex variable* $f \in L^\infty(X)$ *are the numbers*

$$M_k = E(f^k)$$

depending on colored integers $k = \circ \bullet \bullet \circ \ldots$, *with the conventions*

$$f^\emptyset = 1 \quad , \quad f^\circ = f \quad , \quad f^\bullet = \bar{f}$$

and multiplicativity, in order to define the colored powers f^k.

Observe that, since $f, \bar{f}$ commute, we can permute terms, and restrict the attention to exponents of type $k = \ldots \circ \circ \circ \bullet \bullet \bullet \ldots$, if we want to. However, our results about the complex Gaussian laws, and other complex laws too, later on, will actually look better without doing is, so we will use Definition 3.24 as stated. We first have:

THEOREM 3.25. *The moments of the complex normal law are given by*

$$M_k(G_t) = \begin{cases} t^p p! & (k \text{ uniform, of length } 2p) \\ 0 & (k \text{ not uniform}) \end{cases}$$

where $k = \circ \bullet \bullet \circ \ldots$ *is called uniform when it contains the same number of* $\circ$ *and* $\bullet$.

PROOF. We must compute the moments, with respect to colored integer exponents $k = \circ \bullet \bullet \circ \ldots$, of the variable from Definition 3.7, namely:

$$f = \frac{1}{\sqrt{2}}(a + ib)$$

We can assume that we are in the case $t=1$, and the proof here goes as follows:

(1) As a first observation, in the case where our exponent $k=\circ\bullet\bullet\circ\ldots$ is not uniform, a standard rotation argument shows that the corresponding moment of f vanishes. To be more precise, the variable $f'=wf$ is complex Gaussian too, for any complex number $w\in\mathbb T$, and from $M_k(f)=M_k(f')$ we obtain $M_k(f)=0$, in this case.

(2) In the uniform case now, where the exponent $k=\circ\bullet\bullet\circ\ldots$ consists of p copies of $\circ$ and p copies of $\bullet$, the corresponding moment can be computed as follows:

$$\begin{aligned}
M_k &= \int (f\bar{f})^p\\
&= \frac{1}{2^p}\int (a^2+b^2)^p\\
&= \frac{1}{2^p}\sum_r \binom{p}{r}\int a^{2r}\int b^{2p-2r}\\
&= \frac{1}{2^p}\sum_r \binom{p}{r}(2r)!!(2p-2r)!!\\
&= \frac{1}{2^p}\sum_r \frac{p!}{r!(p-r)!}\cdot\frac{(2r)!}{2^r r!}\cdot\frac{(2p-2r)!}{2^{p-r}(p-r)!}\\
&= \frac{p!}{4^p}\sum_r \binom{2r}{r}\binom{2p-2r}{p-r}
\end{aligned}$$

(3) In order to finish now the computation, let us recall that we have the following formula, coming from the generalized binomial formula, or from the Taylor formula:

$$\frac{1}{\sqrt{1+t}}=\sum_{q=0}^\infty \binom{2q}{q}\left(\frac{-t}{4}\right)^q$$

By taking the square of this series, we obtain the following formula:

$$\frac{1}{1+t}=\sum_p \left(\frac{-t}{4}\right)^p\sum_r \binom{2r}{r}\binom{2p-2r}{p-r}$$

Now by looking at the coefficient of t^p on both sides, we conclude that the sum on the right equals 4^p. Thus, we can finish the moment computation in (2), as follows:

$$M_k=\frac{p!}{4^p}\times 4^p=p!$$

We are therefore led to the conclusion in the statement. □

As before with the real Gaussian laws, a better-looking statement is in terms of partitions. Given a colored integer $k = \circ\bullet\bullet\circ\ldots$, we say that a pairing $\pi \in P_2(k)$ is matching when it pairs $\circ - \bullet$ symbols. With this convention, we have the following result:

THEOREM 3.26. *The moments of the complex normal law are the numbers*

$$M_k(G_t) = \sum_{\pi \in \mathcal{P}_2(k)} t^{|\pi|}$$

where $\mathcal{P}_2(k)$ are the matching pairings of $\{1,\ldots,k\}$, and $|.|$ is the number of blocks.

PROOF. This is a reformulation of Theorem 3.25. Indeed, we can assume that we are in the case $t = 1$, and here we know from Theorem 3.25 that the moments are:

$$M_k = \begin{cases} (|k|/2)! & (k \text{ uniform}) \\ 0 & (k \text{ not uniform}) \end{cases}$$

On the other hand, the numbers $|\mathcal{P}_2(k)|$ are given by exactly the same formula. Indeed, in order to have a matching pairing of k, our exponent $k = \circ\bullet\bullet\circ\ldots$ must be uniform, consisting of p copies of $\circ$ and p copies of $\bullet$, with $p = |k|/2$. But then the matching pairings of k correspond to the permutations of the $\bullet$ symbols, as to be matched with $\circ$ symbols, and so we have $p!$ such pairings. Thus, we have the same formula as for the moments of f, and we are led to the conclusion in the statement. □

In practice, we also need to know how to compute joint moments. We have here:

THEOREM 3.27 (Wick formula). *Given independent variables f_i, each following the complex normal law G_t, with $t > 0$ being a fixed parameter, we have the formula*

$$E\left(f_{i_1}^{k_1}\ldots f_{i_s}^{k_s}\right) = t^{s/2}\#\left\{\pi \in \mathcal{P}_2(k)\Big|\pi \leq \ker i\right\}$$

where $k = k_1\ldots k_s$ and $i = i_1\ldots i_s$, for the joint moments of these variables, where $\pi \leq \ker i$ means that the indices of i must fit into the blocks of π, in the obvious way.

PROOF. This is something well-known, which can be proved as follows:

(1) Let us first discuss the case where we have a single variable f, which amounts in taking $f_i = f$ for any i in the formula in the statement. What we have to compute here are the moments of f, with respect to colored integer exponents $k = \circ\bullet\bullet\circ\ldots$, and the formula in the statement tells us that these moments must be:

$$E(f^k) = t^{|k|/2}|\mathcal{P}_2(k)|$$

But this is the formula in Theorem 3.26, so we are done with this case.

(2) In general now, when expanding the product $f_{i_1}^{k_1}\ldots f_{i_s}^{k_s}$ and rearranging the terms, we are left with doing a number of computations as in (1), and then making the product of the expectations that we found. But this amounts in counting the partitions in the

statement, with the condition $\pi \leq \ker i$ there standing for the fact that we are doing the various type (1) computations independently, and then making the product. □

The above statement is one of the possible formulations of the Wick formula, and there are many more formulations, which are all useful. For instance, we have:

THEOREM 3.28 (Wick formula 2). *Given independent variables f_i, each following the complex normal law G_t, with $t > 0$ being a fixed parameter, we have the formula*

$$E\left(f_{i_1}\ldots f_{i_k}f_{j_1}^*\ldots f_{j_k}^*\right)=t^k\#\left\{\pi\in S_k\Big|i_{\pi(r)}=j_r,\forall r\right\}$$

for the non-vanishing joint moments of these variables.

PROOF. This follows from the usual Wick formula, from Theorem 3.27. With some changes in the indices and notations, the formula there reads:

$$E\left(f_{I_1}^{K_1}\ldots f_{I_s}^{K_s}\right)=t^{s/2}\#\left\{\sigma\in\mathcal{P}_2(K)\Big|\sigma\leq\ker I\right\}$$

Now observe that we have $\mathcal{P}_2(K)=\emptyset$, unless the colored integer $K=K_1\ldots K_s$ is uniform, in the sense that it contains the same number of $\circ$ and $\bullet$ symbols. Up to permutations, the non-trivial case, where the moment is non-vanishing, is the case where the colored integer $K=K_1\ldots K_s$ is of the following special form:

$$K=\underbrace{\circ\circ\ldots\circ}_k\ \underbrace{\bullet\bullet\ldots\bullet}_k$$

So, let us focus on this case, which is the non-trivial one. Here we have $s=2k$, and we can write the multi-index $I=I_1\ldots I_s$ in the following way:

$$I=i_1\ldots i_k\ j_1\ldots j_k$$

With these changes made, the above usual Wick formula reads:

$$E\left(f_{i_1}\ldots f_{i_k}f_{j_1}^*\ldots f_{j_k}^*\right)=t^k\#\left\{\sigma\in\mathcal{P}_2(K)\Big|\sigma\leq\ker(ij)\right\}$$

The point now is that the matching pairings $\sigma\in\mathcal{P}_2(K)$, with $K=\circ\ldots\circ\bullet\ldots\bullet$, of length $2k$, as above, correspond to the permutations $\pi\in S_k$, in the obvious way. With this identification made, the above modified usual Wick formula becomes:

$$E\left(f_{i_1}\ldots f_{i_k}f_{j_1}^*\ldots f_{j_k}^*\right)=t^k\#\left\{\pi\in S_k\Big|i_{\pi(r)}=j_r,\forall r\right\}$$

Thus, we have reached to the formula in the statement, and we are done. □

Finally, here is one more formulation of the Wick formula, useful as well:

THEOREM 3.29 (Wick formula 3). *Given independent variables f_i, each following the complex normal law G_t, with $t > 0$ being a fixed parameter, we have the formula*

$$E\left(f_{i_1}f_{j_1}^*\ldots f_{i_k}f_{j_k}^*\right)=t^k\#\left\{\pi\in S_k\Big|i_{\pi(r)}=j_r,\forall r\right\}$$

for the non-vanishing joint moments of these variables.

PROOF. This follows from our second Wick formula, from Theorem 3.28, simply by permuting the terms, as to have an alternating sequence of plain and conjugate variables. Alternatively, we can start with Theorem 3.27, and then perform the same manipulations as in the proof of Theorem 3.28, but with the exponent being this time as follows:

$$K = \underbrace{\circ\bullet\circ\bullet\ldots\ldots\circ\bullet}_{2k}$$

Thus, we are led to the conclusion in the statement. □

We can go back now to the spherical integrals, and we have the following result:

THEOREM 3.30. *The rescalings $\sqrt{N}z_i$ of the unit complex sphere coordinates*

$$z_i : S^{N-1}_{\mathbb{C}} \to \mathbb{C}$$

as well as the rescalings $\sqrt{N}U_{ij}$ of the unitary group coordinates

$$U_{ij} : U_N \to \mathbb{C}$$

become complex Gaussian and independent with $N \to \infty$.

PROOF. We have several assertions to be proved, the idea being as follows:

(1) According to the formula in Theorem 3.18, the polynomials integrals in $z_i, \bar{z}_i$ vanish, unless the number of $z_i, \bar{z}_i$ is the same. In this latter case these terms can be grouped together, by using $z_i\bar{z}_i = |z_i|^2$, and the relevant integration formula is:

$$\int_{S^{N-1}_{\mathbb{C}}} |z_i|^{2k}\,dz = \frac{(N-1)!k!}{(N+k-1)!}$$

Now with $N \to \infty$, we obtain from this the following estimate:

$$\int_{S^{N-1}_{\mathbb{C}}} |z_i|^{2k}dx \simeq N^{-k} \times k!$$

Thus, the rescaled variables $\sqrt{N}z_i$ become normal with $N \to \infty$, as claimed.

(2) As for the proof of asymptotic independence, this is standard too, again by using the formula in Theorem 3.18. Indeed, the joint moments of $z_1, \ldots, z_N$ are given by:

$$\begin{aligned}\int_{S^{N-1}_{\mathbb{R}}} |z_1|^{2k_1}\ldots|z_N|^{2k_N}\,dx &= \frac{(N-1)!k_1!\ldots k_n!}{(N+\sum k_i - 1)!}\\ &\simeq N^{-\Sigma k_i} \times k_1!\ldots k_N!\end{aligned}$$

By rescaling, the joint moments of the variables $y_i = \sqrt{N}z_i$ are given by:

$$\int_{S^{N-1}_{\mathbb{R}}} |y_1|^{2k_1}\ldots|y_N|^{2k_N}\,dx \simeq k_1!\ldots k_N!$$

Thus, we have multiplicativity, and so independence with $N \to \infty$, as claimed.

(3) Regarding the last assertion, we can use here the basic fact that the rotations $U\in U_N$ act on the points of the sphere $z\in S^{N-1}_\mathbb{C}$, with the stabilizer of $z=(1,0,\ldots,0)$ being the subgroup $U_{N-1}\subset U_N$. In algebraic terms, this gives an equality as follows:

$$S^{N-1}_\mathbb{C}=U_N/U_{N-1}$$

In functional analytic terms, this result provides us with an embedding as follows, for any i, which makes correspond the respective integration functionals:

$$C(S^{N-1}_\mathbb{C})\subset C(U_N)\quad,\quad x_i\to U_{1i}$$

With this identification made, the result follows from (1,2). □

3e. Exercises

This was a standard multivariable calculus chapter, and as exercises, we have:

Exercise 3.31. *Can you find an elementary, 1D proof for the Gauss formula?*

Exercise 3.32. *Learn if needed more about the Fourier transform.*

Exercise 3.33. *Clarify the range of angles, in the spherical coordinate formula.*

Exercise 3.34. *Learn if needed the proof of the Stirling formula.*

Exercise 3.35. *Estimate, using Stirling, the central binomial coefficients.*

Exercise 3.36. *Learn the Stone-Weierstrass theorem, and apply it to spheres.*

Exercise 3.37. *Compute the densities of hyperspherical laws, at small N.*

Exercise 3.38. *Have a look at random matrices, heavily using the Wick formula.*

As bonus exercise, find and compute 100 double or triple integrals.

CHAPTER 4

Basic applications

4a. Waves and heat

With the mathematics that we learned, and more specifically with spherical coordinates, we are now ready for some physics. Let us start with a key result, as follows:

THEOREM 4.1. *The wave equation in* $\mathbb{R}^N$ *is*

$$\ddot{f} = v^2 \Delta f$$

where dots denote time derivatives, Δ *is the Laplace operator, given by*

$$\Delta f = \sum_i \frac{d^2 f}{dx_i^2}$$

and $v > 0$ *is the propagation speed.*

PROOF. The equation in the statement is of course what comes out of physics experiments. However, allowing us a bit of imagination, and trust in this imagination, we can mathematically "prove" this equation, by discretizing, as follows:

(1) Let us first consider the 1D case. In order to understand the propagation of waves, we will model $\mathbb{R}$ as a network of balls, with springs between them, as follows:

$$\cdots \times\!\times\!\times \bullet \times\!\times\!\times \bullet \times\!\times\!\times \bullet \times\!\times\!\times \bullet \times\!\times\!\times \bullet \times\!\times\!\times \cdots$$

Now let us send an impulse, and see how the balls will be moving. For this purpose, we zoom on one ball. The situation here is as follows, l being the spring length:

$$\cdots\cdots \bullet_{f(x-l)} \times\!\times\!\times \bullet_{f(x)} \times\!\times\!\times \bullet_{f(x+l)} \cdots\cdots$$

We have two forces acting at x. First is the Newton motion force, mass times acceleration, which is as follows, with m being the mass of each ball:

$$F_n = m \cdot \ddot{f}(x)$$

And second is the Hooke force, displacement of the spring, times spring constant. Since we have two springs at x, this is as follows, k being the spring constant:

$$\begin{aligned} F_h &= F_h^r - F_h^l \\ &= k(f(x+l) - f(x)) - k(f(x) - f(x-l)) \\ &= k(f(x+l) - 2f(x) + f(x-l)) \end{aligned}$$

We conclude that the equation of motion, in our model, is as follows:

$$m \cdot \ddot{f}(x) = k(f(x+l) - 2f(x) + f(x-l))$$

(2) Now let us take the limit of our model, as to reach to continuum. For this purpose we will assume that our system consists of $B >> 0$ balls, having a total mass M, and spanning a total distance L. Thus, our previous infinitesimal parameters are as follows, with K being the spring constant of the total system, which is of course lower than k:

$$m = \frac{M}{B} \quad , \quad k = KB \quad , \quad l = \frac{L}{B}$$

With these changes, our equation of motion found in (1) reads:

$$\ddot{f}(x) = \frac{KB^2}{M}(f(x+l) - 2f(x) + f(x-l))$$

Now observe that this equation can be written, more conveniently, as follows:

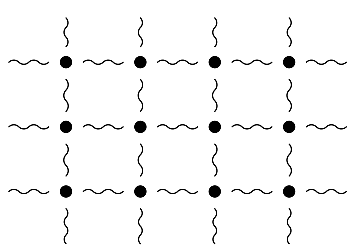
$$\ddot{f}(x) = \frac{KL^2}{M} \cdot \frac{f(x+l) - 2f(x) + f(x-l)}{l^2}$$

With $N \to \infty$, and therefore $l \to 0$, we obtain in this way:

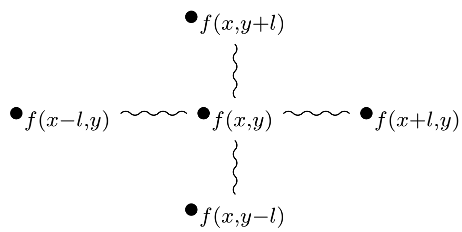
$$\ddot{f}(x) = \frac{KL^2}{M} \cdot \frac{d^2 f}{dx^2}(x)$$

We are therefore led to the wave equation in the statement, which is $\ddot{f} = v^2 f''$ in our present $N = 1$ dimensional case, the propagation speed being $v = \sqrt{K/M} \cdot L$.

(3) In 2 dimensions now, the same argument carries on. Indeed, we can use here a lattice model as follows, with all the edges standing for small springs:



As before in one dimension, we send an impulse, and we zoom on one ball. The situation here is as follows, with l being the spring length:



We have two forces acting at (x,y). First is the Newton motion force, mass times acceleration, which is as follows, with m being the mass of each ball:

$$F_n = m \cdot \ddot{f}(x,y)$$

And second is the Hooke force, displacement of the spring, times spring constant. Since we have four springs at (x,y), this is as follows, k being the spring constant:

$$\begin{aligned}
F_h &= F_h^r - F_h^l + F_h^u - F_h^d \\
&= k(f(x+l,y) - f(x,y)) - k(f(x,y) - f(x-l,y)) \\
&+ k(f(x,y+l) - f(x,y)) - k(f(x,y) - f(x,y-l)) \\
&= k(f(x+l,y) - 2f(x,y) + f(x-l,y)) \\
&+ k(f(x,y+l) - 2f(x,y) + f(x,y-l))
\end{aligned}$$

We conclude that the equation of motion, in our model, is as follows:

$$\begin{aligned}
m \cdot \ddot{f}(x,y) &= k(f(x+l,y) - 2f(x,y) + f(x-l,y)) \\
&+ k(f(x,y+l) - 2f(x,y) + f(x,y-l))
\end{aligned}$$

(4) Now let us take the limit of our model, as to reach to continuum. For this purpose we will assume that our system consists of $B^2 >> 0$ balls, having a total mass M, and spanning a total area L^2. Thus, our previous infinitesimal parameters are as follows, with K being the spring constant of the total system, taken to be equal to k:

$$m = \frac{M}{B^2} \quad , \quad k = K \quad , \quad l = \frac{L}{B}$$

With these changes, our equation of motion found in (3) reads:

$$\begin{aligned}
\ddot{f}(x,y) &= \frac{KB^2}{M}(f(x+l,y) - 2f(x,y) + f(x-l,y)) \\
&+ \frac{KB^2}{M}(f(x,y+l) - 2f(x,y) + f(x,y-l))
\end{aligned}$$

Now observe that this equation can be written, more conveniently, as follows:

$$\begin{aligned}\ddot{f}(x,y) &= \frac{KL^2}{M}\times\frac{f(x+l,y)-2f(x,y)+f(x-l,y)}{l^2}\\ &+ \frac{KL^2}{M}\times\frac{f(x,y+l)-2f(x,y)+f(x,y-l)}{l^2}\end{aligned}$$

With $N\to\infty$, and therefore $l\to 0$, we obtain in this way:

$$\ddot{f}(x,y)=\frac{KL^2}{M}\left(\frac{d^2f}{dx^2}+\frac{d^2f}{dy^2}\right)(x,y)$$

Thus, we are led in this way to the following wave equation in two dimensions, with $v=\sqrt{K/M}\cdot L$ being the propagation speed of our wave:

$$\ddot{f}(x,y)=v^2\left(\frac{d^2f}{dx^2}+\frac{d^2f}{dy^2}\right)(x,y)$$

But we recognize at right the Laplace operator, and we are done. As before in 1D, there is of course some discussion to be made here, arguing that our spring model in (3) is indeed the correct one. But do not worry, experiments confirm our findings.

(5) In 3 dimensions now, which is the case of the main interest, corresponding to our real-life world, the same argument carries over, and the wave equation is as follows:

$$\ddot{f}(x,y,z)=v^2\left(\frac{d^2f}{dx^2}+\frac{d^2f}{dy^2}+\frac{d^2f}{dz^2}\right)(x,y,z)$$

(6) Finally, the same argument, namely a lattice model, carries on in arbitrary N dimensions, and the wave equation here is as follows:

$$\ddot{f}(x_1,\ldots,x_N)=v^2\sum_{i=1}^N\frac{d^2f}{dx_i^2}(x_1,\ldots,x_N)$$

Thus, we are led to the conclusion in the statement. $\square$

In order to discuss the heat equation as well, we will need the following fact:

PROPOSITION 4.2. *Intuitively, the Laplacian of a function f, given by*

$$\Delta f=\sum_i\frac{d^2f}{dx_i^2}$$

computes how much different is $f(x)$, compared to the average of $f(y)$, with $y\simeq x$.

PROOF. Before anything, you might wonder what kind of mathematical statement this is, talking about intuition, instead of precise things, as we are supposed to. In answer, please relax and calm down, we are doing physics here, right now in this chapter. Getting now to the proof, let us write the Taylor formula for f at order 2, as follows:

$$f(x+h) \simeq f(x) + f'(x)h + \frac{< f''(x)h, h >}{2}$$

With the change $h \to -h$, the following approximation formula holds too:

$$f(x-h) \simeq f(x) - f'(x)h + \frac{< f''(x)h, h >}{2}$$

Now by making the average of our formulae, we obtain the following formula:

$$\frac{f(x+h) + f(x-h)}{2} \simeq f(x) + \frac{< f''(x)h, h >}{2}$$

Thus, thinking a bit, we are led to the conclusion in the statement. It is of course possible to say more here, but we will not really need all the details, in what follows. □

Now back to physics, regarding heat diffusion, we have the following result:

THEOREM 4.3. *Heat diffusion in $\mathbb{R}^N$ is described by the heat equation*

$$\dot{f} = \alpha \Delta f$$

where $\alpha > 0$ is the thermal diffusivity of the medium, and Δ is the Laplace operator.

PROOF. The study here is quite similar to the study of waves, as follows:

(1) To start with, as an intuitive explanation for the equation, since the second derivative f'' in one dimension, or the quantity Δf in general, computes the average value of a function f around a point, minus the value of f at that point, the heat equation as formulated above tells us that the rate of change $\dot{f}$ of the temperature of the material at any given point must be proportional, with proportionality factor $\alpha > 0$, to the average difference of temperature between that given point and the surrounding material.

(2) The heat equation as formulated above is of course something approximative, and several improvements can be made to it, first by incorporating a term accounting for heat radiation, and then doing several fine-tunings, depending on the material involved. But more on this later, for the moment let us focus on the heat equation above.

(3) In relation with our modeling questions, we can recover this equation a bit as we did for the wave equation before, by using a basic lattice model. Indeed, let us first assume, for simplifying, that we are in the one-dimensional case, $N = 1$. Here our model looks as follows, with distance $l > 0$ between neighbors:

$$—— \circ_{x-l} \overset{l}{——} \circ_x \overset{l}{——} \circ_{x+l} ——$$

In order to model heat diffusion, we have to implement the intuitive mechanism explained above, namely "the rate of change of the temperature of the material at any given point must be proportional, with proportionality factor $\alpha > 0$, to the average difference of temperature between that given point and the surrounding material".

(4) In practice, this leads to a condition as follows, expressing the change of the temperature φ, over a small period of time $\delta > 0$:

$$f(x,t+\delta) = f(x,t) + \frac{\alpha\delta}{l^2}\sum_{x\sim y}[f(y,t) - f(x,t)]$$

To be more precise, we have made several assumptions here, as follows:

– General heat diffusion assumption: the change of temperature at any given point x is proportional to the average over neighbors, $y \sim x$, of the differences $f(y,t) - f(x,t)$ between the temperatures at x, and at these neighbors y.

– Infinitesimal time and length conditions: in our model, the change of temperature at a given point x is proportional to small period of time involved, $\delta > 0$, and is inverse proportional to the square of the distance between neighbors, l^2.

(5) Regarding these latter assumptions, the one regarding the proportionality with the time elapsed $\delta > 0$ is something quite natural, physically speaking, and mathematically speaking too, because we can rewrite our equation as follows, making it clear that we have here an equation regarding the rate of change of temperature at x:

$$\frac{f(x,t+\delta) - f(x,t)}{\delta} = \frac{\alpha}{l^2}\sum_{x\sim y}[f(y,t) - f(x,t)]$$

As for the second assumption that we made above, namely inverse proportionality with l^2, this can be justified on physical grounds too, but again, perhaps the best is to do the math, which will show right away where this proportionality comes from.

(6) So, let us do the math. In the context of our 1D model the neighbors of x are the points $x \pm l$, and so the equation that we wrote above takes the following form:

$$\frac{f(x,t+\delta) - f(x,t)}{\delta} = \frac{\alpha}{l^2}\Big[(f(x+l,t) - f(x,t)) + (f(x-l,t) - f(x,t))\Big]$$

Now observe that we can write this equation as follows:

$$\frac{f(x,t+\delta) - f(x,t)}{\delta} = \alpha \cdot \frac{f(x+l,t) - 2f(x,t) + f(x-l,t)}{l^2}$$

(7) As it was the case with the wave equation before, we recognize on the right the usual approximation of the second derivative, coming from calculus. Thus, when taking the continuous limit of our model, $l \to 0$, we obtain the following equation:

$$\frac{f(x,t+\delta) - f(x,t)}{\delta} = \alpha \cdot f''(x,t)$$

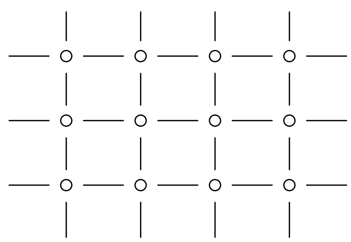
Now with $t \to 0$, we are led in this way to the heat equation, namely:

$$\dot{f}(x,t) = \alpha \cdot f''(x,t)$$

Summarizing, we are done with the 1D case, with our proof being quite similar to the one for the wave equation, done earlier in this section.

(8) In practice now, there are of course still a few details to be discussed, in relation with all this, for instance at the end, in relation with the precise order of the limiting operations $l \to 0$ and $\delta \to 0$ to be performed, but these remain minor aspects, because our equation makes it clear, right from the beginning, that time and space are separated, and so that there is no serious issue with all this. And so, fully done with 1D.

(9) With this done, let us discuss now 2 dimensions. Here, as before for the waves, we can use a lattice model as follows, with all lengths being $l > 0$, for simplifying:



(10) We have to implement now the physical heat diffusion mechanism, namely "the rate of change of the temperature of the material at any given point must be proportional, with proportionality factor $\alpha > 0$, to the average difference of temperature between that given point and the surrounding material". In practice, this leads to a condition as follows, expressing the change of the temperature f, over a small period of time $\delta > 0$:

$$f(x,y,t+\delta) = f(x,y,t) + \frac{\alpha\delta}{l^2} \sum_{(x,y)\sim(u,v)} [f(u,v,t) - f(x,y,t)]$$

In fact, we can rewrite our equation as follows, making it clear that we have here an equation regarding the rate of change of temperature at x:

$$\frac{f(x,y,t+\delta) - f(x,y,t)}{\delta} = \frac{\alpha}{l^2} \sum_{(x,y)\sim(u,v)} [f(u,v,t) - f(x,y,t)]$$

(11) So, let us do the math. In the context of our 2D model the neighbors of x are the points $(x \pm l, y \pm l)$, so the equation above takes the following form:

$$\begin{aligned} & \frac{f(x,y,t+\delta)-f(x,y,t)}{\delta} \\ = & \frac{\alpha}{l^2}\Big[(f(x+l,y,t)-f(x,y,t))+(f(x-l,y,t)-f(x,y,t))\Big] \\ + & \frac{\alpha}{l^2}\Big[(f(x,y+l,t)-f(x,y,t))+(f(x,y-l,t)-f(x,y,t))\Big] \end{aligned}$$

Now observe that we can write this equation as follows:

$$\begin{aligned} \frac{f(x,y,t+\delta)-f(x,y,t)}{\delta} &= \alpha\cdot\frac{f(x+l,y,t)-2f(x,y,t)+f(x-l,y,t)}{l^2} \\ &+ \alpha\cdot\frac{f(x,y+l,t)-2f(x,y,t)+f(x,y-l,t)}{l^2} \end{aligned}$$

(12) As it was the case when modeling the wave equation before, we recognize on the right the usual approximation of the second derivative, coming from calculus. Thus, when taking the continuous limit of our model, $l \to 0$, we obtain the following equation:

$$\frac{f(x,y,t+\delta)-f(x,y,t)}{\delta} = \alpha\left(\frac{d^2f}{dx^2}+\frac{d^2f}{dy^2}\right)(x,y,t)$$

Now with $t \to 0$, we are led in this way to the heat equation, namely:

$$\dot{f}(x,y,t) = \alpha\cdot\Delta f(x,y,t)$$

Finally, in arbitrary N dimensions the same argument carries over, namely a straightforward lattice model, and gives the heat equation, as formulated in the statement. $\square$

So long for the basic physics of waves and heat. In practice, many other things can be said about waves and heat, with a summary here being as follows:

(1) We have been talking in the above about mechanical waves, but quite remarkably, the other types of waves, such as the electromagnetic ones, are described by the same equation, namely $\ddot{f} = v^2\Delta f$. This follows indeed from the Maxwell equations.

(2) Talking electromagnetism, the Laplace operator appears in fact naturally before moving charges, magnetism and waves, in the context of electrostatics. To be more precise, we have there the Laplace equation $\Delta f = 0$, appearing in relation with potentials.

(3) Still talking electromagnetism, the refined version of the theory, which is no longer statistical, and applies at very small scales, is quantum mehanics, where the key equation is the Schrödinger one $i\dot{f} = a\Delta f + bf$, still making use of the Laplace operator Δ.

(4) Finally, regarding the heat equation $\dot{f}=\alpha\Delta f$, this needs in practice a number of corrections, due to Stefan-Boltzmann and others. By the way, observe also the striking similarity between the Schrödinger equation, and the heat equation.

And we will refer here to any standard physics book, such as Feynman [32], for more on all this. In any case, convinced I hope, it is all about the Laplace operator Δ.

4b. Harmonic functions

Getting now to the mathematics of Δ, previous experience with linear algebra and linear operators suggests looking first into the eigenvectors of Δ. But the simplest such eigenvectors are those corresponding to the eigenvalue $\lambda=0$, called harmonic:

$$\Delta f=0$$

So, let us first try to find the functions $f:\mathbb{R}^2\to\mathbb{C}$ which are harmonic. And here, as a good surprise, we have an interesting link with the holomorphic functions:

Theorem 4.4. *Any holomorphic function $f:\mathbb{C}\to\mathbb{C}$, when regarded as real function*

$$f:\mathbb{R}^2\to\mathbb{C}$$

is harmonic. Moreover, the conjugates $\bar{f}$ of holomorphic functions are harmonic too.

Proof. The first assertion follows from the following computation, for the power functions $f(z)=z^n$, with the usual notation $z=x+iy$:

$$\begin{aligned}\Delta z^n &= \frac{d^2z^n}{dx^2}+\frac{d^2z^n}{dy^2}\\ &= \frac{d(nz^{n-1})}{dx}+\frac{d(inz^{n-1})}{dy}\\ &= n(n-1)z^{n-2}-n(n-1)z^{n-2}\\ &= 0\end{aligned}$$

As for the second assertion, this follows from $\Delta\bar{f}=\overline{\Delta f}$, which is clear from definitions, and which shows that if f is harmonic, then so is its conjugate $\bar{f}$. □

In order to understand the harmonic functions, we can try to find the homogeneous polynomials $P\in\mathbb{R}[x,y]$ which are harmonic. In order to do so, the most convenient is to use the variable $z=x+iy$, and think of these polynomials as being homogeneous polynomials $P\in\mathbb{R}[z,\bar{z}]$. With this convention, the result is as follows:

Theorem 4.5. *The degree n homogeneous polynomials $P\in\mathbb{R}[x,y]$ which are harmonic are precisely the linear combinations of*

$$P=z^n\quad,\quad P=\bar{z}^n$$

with the usual identification $z=x+iy$.

PROOF. As explained above, any homogeneous polynomial $P \in \mathbb{R}[x,y]$ can be regarded as an homogeneous polynomial $P \in \mathbb{R}[z,\bar{z}]$, with the change of variables $z = x+iy$, and in this picture, the degree n homogeneous polynomials are as follows:

$$P(z) = \sum_{k+l=n} c_{kl} z^k \bar{z}^l$$

In oder to solve now the Laplace equation $\Delta P = 0$, we must compute the quantities $\Delta(z^k\bar{z}^l)$, for any k,l. But the computation here is routine. We first have the following formula, with the derivatives being computed with respect to the variable x:

$$\begin{aligned}\frac{d(z^k\bar{z}^l)}{dx} &= (z^k)'\bar{z}^l + z^k(\bar{z}^l)' \\ &= kz^{k-1}\bar{z}^l + lz^k\bar{z}^{l-1}\end{aligned}$$

By taking one more time the derivative with respect to x, we obtain:

$$\begin{aligned}\frac{d^2(z^k\bar{z}^l)}{dx^2} &= k(z^{k-1}\bar{z}^l)' + l(z^k\bar{z}^{l-1})' \\ &= k\left[(z^{k-1})'\bar{z}^l + z^{k-1}(\bar{z}^l)'\right] + l\left[(z^k)'\bar{z}^{l-1} + z^k(\bar{z}^{l-1})'\right] \\ &= k\left[(k-1)z^{k-2}\bar{z}^l + lz^{k-1}\bar{z}^{l-1}\right] + l\left[kz^{k-1}\bar{z}^{l-1} + (l-1)z^k\bar{z}^{l-2}\right] \\ &= k(k-1)z^{k-2}\bar{z}^l + 2klz^{k-1}\bar{z}^{l-1} + l(l-1)z^k\bar{z}^{l-2}\end{aligned}$$

With respect to the variable y, the computations are similar, but some $\pm i$ factors appear, due to $z' = i$ and $\bar{z}' = -i$, coming from $z = x+iy$. We first have:

$$\begin{aligned}\frac{d(z^k\bar{z}^l)}{dy} &= (z^k)'\bar{z}^l + z^k(\bar{z}^l)' \\ &= ikz^{k-1}\bar{z}^l - ilz^k\bar{z}^{l-1}\end{aligned}$$

By taking one more time the derivative with respect to y, we obtain:

$$\begin{aligned}\frac{d^2(z^k\bar{z}^l)}{dy^2} &= ik(z^{k-1}\bar{z}^l)' - il(z^k\bar{z}^{l-1})' \\ &= ik\left[(z^{k-1})'\bar{z}^l + z^{k-1}(\bar{z}^l)'\right] - il\left[(z^k)'\bar{z}^{l-1} + z^k(\bar{z}^{l-1})'\right] \\ &= ik\left[i(k-1)z^{k-2}\bar{z}^l - ilz^{k-1}\bar{z}^{l-1}\right] - il\left[ikz^{k-1}\bar{z}^{l-1} - i(l-1)z^k\bar{z}^{l-2}\right] \\ &= -k(k-1)z^{k-2}\bar{z}^l + 2klz^{k-1}\bar{z}^{l-1} - l(l-1)z^k\bar{z}^{l-2}\end{aligned}$$

We can now sum the formulae that we found, and we obtain:

$$\begin{aligned}\Delta(z^k\bar{z}^l) &= \frac{d^2(z^k\bar{z}^l)}{dx^2}+\frac{d^2(z^k\bar{z}^l)}{dy^2}\\ &= k(k-1)z^{k-2}\bar{z}^l+2klz^{k-1}\bar{z}^{l-1}+l(l-1)z^k\bar{z}^{l-2}\\ &\quad -k(k-1)z^{k-2}\bar{z}^l+2klz^{k-1}\bar{z}^{l-1}-l(l-1)z^k\bar{z}^{l-2}\\ &= 4klz^{k-1}\bar{z}^{l-1}\end{aligned}$$

In other words, we have reached to the following formula:

$$f=z^k\bar{z}^l\implies\Delta f=\frac{4klf}{|z|^2}$$

Now let us get back to our homogeneous polynomial P, written as follows:

$$P(z)=\sum_{k+l=n}c_{kl}z^k\bar{z}^l$$

By using the above formula, the Laplacian of P is given by:

$$\Delta P(z)=\frac{4}{|z|^2}\sum_{k+l=n}klc_{kl}z^k\bar{z}^l$$

We conclude that the Laplace equation for P takes the following form:

$$\begin{aligned}\Delta P=0 &\iff klc_{kl}=0,\forall k,l\\ &\iff [k,l\neq 0\implies c_{kl}=0]\\ &\iff P=c_{n0}z^n+c_{0n}\bar{z}^n\end{aligned}$$

Thus, we are led to the conclusion in the statement. And with the observation that the real formulation of the final result is something quite complicated, and so, for one more time, the use of the complex variable $z=x+iy$ is something very useful. □

As our next objective, coming as a continuation of the above, let us try now to find the harmonic functions which are radial, in the following sense:

$$f(z)=\varphi(|z|)$$

However, things are quite tricky here, involving a blowup phenomenon precisely at the dimension value $N=2$. So, moving now to arbitrary N dimensions, we have here:

THEOREM 4.6. *The fundamental radial solutions of $\Delta f=0$ are*

$$f(x)=\begin{cases}||x||^{2-N} & (N\neq 2)\\ \log||x|| & (N=2)\end{cases}$$

with the log *at $N=2$ basically coming from* $\log'=1/x$.

Proof. Consider indeed a radial function, defined outside the origin $x = 0$. This function can be written as follows, with $\varphi : (0,\infty) \to \mathbb{C}$ being a certain function:

$$f : \mathbb{R}^N - \{0\} \to \mathbb{C} \quad , \quad f(x) = \varphi(||x||)$$

Our first goal will be that of reformulating the Laplace equation $\Delta f = 0$ in terms of the one-variable function $\varphi : (0,\infty) \to \mathbb{C}$. For this purpose, observe that we have:

$$\begin{aligned}
\frac{d||x||}{dx_i} &= \frac{d\sqrt{\sum_{i=1}^N x_i^2}}{dx_i} \\
&= \frac{1}{2} \cdot \frac{1}{\sqrt{\sum_{i=1}^N x_i^2}} \cdot \frac{d\left(\sum_{i=1}^N x_i^2\right)}{dx_i} \\
&= \frac{1}{2} \cdot \frac{1}{||x||} \cdot 2x_i \\
&= \frac{x_i}{||x||}
\end{aligned}$$

By using this formula, we have the following computation:

$$\begin{aligned}
\frac{df}{dx_i} &= \frac{d\varphi(||x||)}{dx_i} \\
&= \varphi'(||x||) \cdot \frac{d||x||}{dx_i} \\
&= \varphi'(||x||) \cdot \frac{x_i}{||x||}
\end{aligned}$$

By differentiating one more time, we obtain the following formula:

$$\begin{aligned}
\frac{d^2 f}{dx_i^2} &= \frac{d}{dx_i}\left(\varphi'(||x||) \cdot \frac{x_i}{||x||}\right) \\
&= \frac{d\varphi'(||x||)}{dx_i} \cdot \frac{x_i}{||x||} + \varphi'(||x||) \cdot \frac{d}{dx_i}\left(\frac{x_i}{||x||}\right) \\
&= \left(\varphi''(||x||) \cdot \frac{x_i}{||x||}\right) \cdot \frac{x_i}{||x||} + \varphi'(||x||) \cdot \frac{||x|| - x_i \cdot x_i/||x||}{||x||^2} \\
&= \varphi''(||x||) \cdot \frac{x_i^2}{||x||^2} + \varphi'(||x||) \cdot \frac{||x||^2 - x_i^2}{||x||^3}
\end{aligned}$$

Now by summing over $i \in \{1, \ldots, N\}$, this gives the following formula:

$$\begin{aligned}\Delta f &= \sum_{i=1}^N \varphi''(||x||) \cdot \frac{x_i^2}{||x||^2} + \sum_{i=1}^N \varphi'(||x||) \cdot \frac{||x||^2 - x_i^2}{||x||^3} \\ &= \varphi''(||x||) \cdot \frac{||x||^2}{||x||^2} + \varphi'(||x||) \cdot \frac{(N-1)||x||^2}{||x||^3} \\ &= \varphi''(||x||) + \varphi'(||x||) \cdot \frac{N-1}{||x||}\end{aligned}$$

Thus, with $r = ||x||$, the Laplace equation $\Delta f = 0$ can be reformulated as follows:

$$\varphi''(r) + \frac{(N-1)\varphi'(r)}{r} = 0$$

Equivalently, the equation that we want to solve is as follows:

$$r\varphi'' + (N-1)\varphi' = 0$$

Now observe that we have the following formula:

$$\begin{aligned}(r^{N-1}\varphi')' &= (N-1)r^{N-2}\varphi' + r^{N-1}\varphi'' \\ &= r^{N-2}((N-1)\varphi' + r\varphi'')\end{aligned}$$

Thus, the equation to be solved can be simply written as follows:

$$(r^{N-1}\varphi')' = 0$$

We conclude that $r^{N-1}\varphi'$ must be a constant K, and so, that we must have:

$$\varphi' = Kr^{1-N}$$

But the fundamental solutions of this latter equation are as follows:

$$\varphi(r) = \begin{cases} r^{2-N} & (N \neq 2) \\ \log r & (N = 2) \end{cases}$$

Thus, we are led to the conclusion in the statement. □

As a last piece of theory, let us examine now the general harmonic functions, in general N dimensions. Many things can be said here, with a summary being as follows:

THEOREM 4.7. *The harmonic functions $f : \mathbb{R}^N \to \mathbb{C}$ obey to the same general principles as the holomorphic functions, namely:*

1. *The maximum modulus principle.*
2. *The plain mean value formula.*
3. *The boundary mean value formula.*
4. *The Liouville theorem.*

PROOF. This is something quite tricky, the idea being as follows:

(1) Regarding the plain mean value formula, here the statement is that given an harmonic function $f : X \to \mathbb{C}$, and a ball B, the following happens:

$$f(x) = \int_B f(y)dy$$

In order to prove this, we can assume that B is centered at 0, of radius $r > 0$. If we denote by χ_r the characteristic function of this ball, nomalized as to integrate up to 1, in terms of the standard convolution operation $*$, we want to prove that we have:

$$f = f * \chi_r$$

For doing so, let us pick a number $0 < s < r$, and a solution w of the following equation, supported on B, which can be constructed explicitly:

$$\Delta w = \chi_r - \chi_s$$

By using now the standard properties of the convolution operation $*$, we have:

$$\begin{aligned} f * \chi_r - f * \chi_s &= f * (\chi_r - \chi_s) \\ &= f * \Delta w \\ &= \Delta f * w \\ &= 0 \end{aligned}$$

Thus $f * \chi_r = f * \chi_s$, and by letting now $s \to 0$, we get $f * \chi_r = f$, as desired.

(2) Regarding the boundary mean value formula, here the statement is that given an harmonic function $f : X \to \mathbb{C}$, and a ball B, with boundary γ, the following happens:

$$f(x) = \int_\gamma f(y)dy$$

But this follows as a consequence of the plain mean value formula in (1), with our two mean value formulae, the one there and the one here, being in fact equivalent, by using annuli and radial integration for the proof of the equivalence, in the obvious way.

(3) Regarding the maximum modulus principle, the statement here is that any holomorphic function $f : X \to \mathbb{C}$ has the property that the maximum of $|f|$ over a domain is attained on its boundary. That is, given a domain D, with boundary γ, we have:

$$\exists x \in \gamma \quad , \quad |f(x)| = \max_{y \in D} |f(y)|$$

But this is something which follows again from the mean value formula in (1), first for the balls, and then in general, by using a standard division argument.

(4) Regarding now the Liouville theorem, as in the holomorphic case, the precise statement here is that an entire, bounded harmonic function must be constant:

$$f:\mathbb{R}^N\to\mathbb{C}\quad,\quad \Delta f=0\quad,\quad |f|\leq M\quad\implies\quad f=\text{constant}$$

As a slightly weaker statement, again called Liouville theorem, we have the fact that an entire harmonic function which vanishes at ∞ must vanish globally:

$$f:\mathbb{R}^N\to\mathbb{C}\quad,\quad \Delta f=0\quad,\quad \lim_{x\to\infty}f(x)=0\quad\implies\quad f=0$$

But can view these as a consequence of the mean value formula in (1), because given two points $x\neq y$, we can view the values of f at these points as averages over big balls centered at these points, say $B=B_x(R)$ and $C=B_y(R)$, with $R>>0$:

$$f(x)=\int_B f(z)dz\quad,\quad f(y)=\int_C f(z)dz$$

Indeed, the point is that when the radius goes to ∞, these averages tend to be equal, and so we have $f(x)\simeq f(y)$, which gives $f(x)=f(y)$ in the limit, as desired. □

Many other things can be said about harmonic functions, notably with the fundamental fact that, locally, the harmonic functions $f:\mathbb{R}^2\to\mathbb{R}$ appear as the real parts of the holomorphic functions. For more on this, and related topics, we refer to Rudin [79].

4c. Polar decomposition

Getting back now to physics and equations, regarding the waves, we have:

THEOREM 4.8. *The solution of the 1D wave equation with initial value conditions $\varphi(x,0)=f(x)$ and $\dot{\varphi}(x,0)=g(x)$ is given by the d'Alembert formula, namely:*

$$\varphi(x,t)=\frac{f(x-vt)+f(x+vt)}{2}+\frac{1}{2v}\int_{x-vt}^{x+vt}g(s)ds$$

In the context of our previous lattice model discretizations, what happens is more or less that the above d'Alembert integral gets computed via Riemann sums.

PROOF. There are several things going on here, the idea being as follows:

(1) Let us first check that the d'Alembert solution is indeed a solution of the wave equation $\ddot{\varphi}=v^2\varphi''$. The first time derivative is computed as follows:

$$\dot{\varphi}(x,t)=\frac{-vf'(x-vt)+vf'(x+vt)}{2}+\frac{1}{2v}(vg(x+vt)+vg(x-vt))$$

The second time derivative is computed as follows:

$$\ddot{\varphi}(x,t)=\frac{v^2f''(x-vt)+v^2f(x+vt)}{2}+\frac{vg'(x+vt)-vg'(x-vt)}{2}$$

Regarding now space derivatives, the first one is computed as follows:

$$\varphi'(x,t)=\frac{f'(x-vt)+f'(x+vt)}{2}+\frac{1}{2v}(g'(x+vt)-g'(x-vt))$$

As for the second space derivative, this is computed as follows:

$$\varphi''(x,t)=\frac{f''(x-vt)+f''(x+vt)}{2}+\frac{g''(x+vt)-g''(x-vt)}{2v}$$

Thus we have indeed $\ddot{\varphi}=v^2\varphi''$. As for the initial conditions, $\varphi(x,0)=f(x)$ is clear from our definition of φ, and $\dot{\varphi}(x,0)=g(x)$ is clear from our above formula of $\dot{\varphi}$.

(2) Conversely now, we must show that our solution is unique, but instead of going here into abstract arguments, we will simply solve our equation, which among others will doublecheck the computations in (1). Let us make the following change of variables:

$$\xi=x-vt\quad,\quad\eta=x+vt$$

With this change of variables, which is quite tricky, mixing space and time variables, our wave equation $\ddot{\varphi}=v^2\varphi''$ reformulates in a very simple way, as follows:

$$\frac{d^2\varphi}{d\xi d\eta}=0$$

But this latter equation tells us that our new ξ,η variables get separated, and we conclude from this that the solution must be of the following special form:

$$\varphi(x,t)=F(\xi)+G(\eta)=F(x-vt)+G(x+vt)$$

Now by taking into account the intial conditions $\varphi(x,0)=f(x)$ and $\dot{\varphi}(x,0)=g(x)$, and then integrating, we are led to the d'Alembert formula in the statement.

(3) In regards now with our discretization questions, by using a 1D lattice model with balls and springs as before, what happens to all the above is more or less that the above d'Alembert integral gets computed via Riemann sums, in our model, as stated. □

In $N\geq 2$ dimensions things are more complicated. In the case $N=3$, which is where most physics happens, of great use is the following technical result:

Theorem 4.9. *The Laplace operator in spherical coordinates is*

$$\Delta=\frac{1}{r^2}\cdot\frac{d}{dr}\left(r^2\cdot\frac{d}{dr}\right)+\frac{1}{r^2\sin s}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{d}{ds}\right)+\frac{1}{r^2\sin^2 s}\cdot\frac{d^2}{dt^2}$$

with our standard conventions for these coordinates, in 3D.

Proof. There are several proofs here, a short, elementary one being as follows:

(1) Let us first see how Δ behaves under a change of coordinates $\{x_i\} \to \{y_i\}$, in arbitrary N dimensions. Our starting point is the chain rule for derivatives:

$$\frac{d}{dx_i} = \sum_j \frac{d}{dy_j} \cdot \frac{dy_j}{dx_i}$$

By using this rule, then Leibnitz for products, then again this rule, we obtain:

$$\begin{aligned}
\frac{d^2 f}{dx_i^2} &= \sum_j \frac{d}{dx_i}\left(\frac{df}{dy_j} \cdot \frac{dy_j}{dx_i}\right) \\
&= \sum_j \frac{d}{dx_i}\left(\frac{df}{dy_j}\right) \cdot \frac{dy_j}{dx_i} + \frac{df}{dy_j} \cdot \frac{d}{dx_i}\left(\frac{dy_j}{dx_i}\right) \\
&= \sum_j \left(\sum_k \frac{d}{dy_k} \cdot \frac{dy_k}{dx_i}\right)\left(\frac{df}{dy_j}\right) \cdot \frac{dy_j}{dx_i} + \frac{df}{dy_j} \cdot \frac{d^2 y_j}{dx_i^2} \\
&= \sum_{jk} \frac{d^2 f}{dy_k dy_j} \cdot \frac{dy_k}{dx_i} \cdot \frac{dy_j}{dx_i} + \sum_j \frac{df}{dy_j} \cdot \frac{d^2 y_j}{dx_i^2}
\end{aligned}$$

(2) Now by summing over i, we obtain the following formula, with A being the derivative of $x \to y$, that is to say, the matrix of partial derivatives dy_i/dx_j:

$$\begin{aligned}
\Delta f &= \sum_{ijk} \frac{d^2 f}{dy_k dy_j} \cdot \frac{dy_k}{dx_i} \cdot \frac{dy_j}{dx_i} + \sum_{ij} \frac{df}{dy_j} \cdot \frac{d^2 y_j}{dx_i^2} \\
&= \sum_{ijk} A_{ki} A_{ji} \frac{d^2 f}{dy_k dy_j} + \sum_{ij} \frac{d^2 y_j}{dx_i^2} \cdot \frac{df}{dy_j} \\
&= \sum_{jk} (AA^t)_{jk} \frac{d^2 f}{dy_k dy_j} + \sum_j \Delta(y_j) \frac{df}{dy_j}
\end{aligned}$$

(3) So, this will be the formula that we will need. Observe that this formula can be further compacted as follows, with all the notations being self-explanatory:

$$\Delta f = Tr(AA^t H_y(f)) + < \Delta(y), \nabla_y(f) >$$

(4) Getting now to spherical coordinates, $(x, y, z) \to (r, s, t)$, the derivative of the inverse, obtained by differentiating x, y, z with respect to r, s, t, is given by:

$$A^{-1} = \begin{pmatrix} \cos s & -r \sin s & 0 \\ \sin s \cos t & r \cos s \cos t & -r \sin s \sin t \\ \sin s \sin t & r \cos s \sin t & r \sin s \cos t \end{pmatrix}$$

The product $(A^{-1})^tA^{-1}$ of the transpose of this matrix with itself is then:

$$\begin{pmatrix}\cos s&\sin s\cos t&\sin s\sin t\\ -r\sin s&r\cos s\cos t&r\cos s\sin t\\ 0&-r\sin s\sin t&r\sin s\cos t\end{pmatrix}\begin{pmatrix}\cos s&-r\sin s&0\\ \sin s\cos t&r\cos s\cos t&-r\sin s\sin t\\ \sin s\sin t&r\cos s\sin t&r\sin s\cos t\end{pmatrix}$$

But everything simplifies here, and we have the following remarkable formula, which by the way is something very useful, worth to be memorized:

$$(A^{-1})^tA^{-1}=\begin{pmatrix}1&0&0\\ 0&r^2&0\\ 0&0&r^2\sin^2 s\end{pmatrix}$$

Now by inverting, we obtain the following formula, in relation with the above:

$$AA^t=\begin{pmatrix}1&0&0\\ 0&1/r^2&0\\ 0&0&1/(r^2\sin^2 s)\end{pmatrix}$$

(5) Let us compute now the Laplacian of r,s,t. We first have the following formula, that we will use many times in what follows, and is worth to be memorized:

$$\begin{aligned}\frac{dr}{dx}&=\frac{d}{dx}\sqrt{x^2+y^2+z^2}\\ &=\frac{1}{2}\cdot\frac{2x}{\sqrt{x^2+y^2+z^2}}\\ &=\frac{x}{r}\end{aligned}$$

Of course the same computation works for y,z too, and we therefore have:

$$\frac{dr}{dx}=\frac{x}{r}\quad,\quad\frac{dr}{dy}=\frac{y}{r}\quad,\quad\frac{dr}{dz}=\frac{z}{r}$$

(6) By using the above formulae, twice, we can compute the Laplacian of r:

$$\begin{aligned}\Delta(r)&=\Delta\left(\sqrt{x^2+y^2+z^2}\right)\\ &=\frac{d}{dx}\left(\frac{x}{r}\right)+\frac{d}{dy}\left(\frac{y}{r}\right)+\frac{d}{dz}\left(\frac{z}{r}\right)\\ &=\frac{r^2-x^2}{r^3}+\frac{r^2-y^2}{r^3}+\frac{r^2-z^2}{r^3}\\ &=\frac{2}{r}\end{aligned}$$

(7) In what regards now s, the computation here goes as follows:

$$\begin{aligned}
\Delta(s) &= \Delta\left(\arccos\left(\frac{x}{r}\right)\right)\\
&= \frac{d}{dx}\left(-\frac{\sqrt{r^2-x^2}}{r^2}\right)+\frac{d}{dy}\left(\frac{xy}{r^2\sqrt{r^2-x^2}}\right)+\frac{d}{dz}\left(\frac{xz}{r^2\sqrt{r^2-x^2}}\right)\\
&= \frac{2x\sqrt{r^2-x^2}}{r^4}+\frac{r^2(z^2-2y^2)+2x^2y^2}{r^4\sqrt{r^2-x^2}}+\frac{r^2(y^2-2z^2)+2x^2z^2}{r^4\sqrt{r^2-x^2}}\\
&= \frac{2x\sqrt{r^2-x^2}}{r^4}+\frac{x(2x^2-r^2)}{r^4\sqrt{r^2-x^2}}\\
&= \frac{x}{r^2\sqrt{r^2-x^2}}\\
&= \frac{\cos s}{r^2\sin s}
\end{aligned}$$

(8) Finally, in what regards t, the computation here goes as follows:

$$\begin{aligned}
\Delta(t) &= \Delta\left(\arctan\left(\frac{z}{y}\right)\right)\\
&= \frac{d}{dx}(0)+\frac{d}{dy}\left(-\frac{z}{y^2+z^2}\right)+\frac{d}{dz}\left(\frac{y}{y^2+z^2}\right)\\
&= 0-\frac{2yz}{(y^2+z^2)^2}+\frac{2yz}{(y^2+z^2)^2}\\
&= 0
\end{aligned}$$

(9) We can now plug the data from (4) and (6,7,8) in the general formula that we found in (2) above, and we obtain in this way:

$$\begin{aligned}
\Delta f &= \frac{d^2f}{dr^2}+\frac{1}{r^2}\cdot\frac{d^2f}{ds^2}+\frac{1}{r^2\sin^2 s}\cdot\frac{d^2f}{dt^2}+\frac{2}{r}\cdot\frac{df}{dr}+\frac{\cos s}{r^2\sin s}\cdot\frac{df}{ds}\\
&= \frac{2}{r}\cdot\frac{df}{dr}+\frac{d^2f}{dr^2}+\frac{\cos s}{r^2\sin s}\cdot\frac{df}{ds}+\frac{1}{r^2}\cdot\frac{d^2f}{ds^2}+\frac{1}{r^2\sin^2 s}\cdot\frac{d^2f}{dt^2}\\
&= \frac{1}{r^2}\cdot\frac{d}{dr}\left(r^2\cdot\frac{df}{dr}\right)+\frac{1}{r^2\sin s}\cdot\frac{d}{ds}\left(\sin s\cdot\frac{df}{ds}\right)+\frac{1}{r^2\sin^2 s}\cdot\frac{d^2f}{dt^2}
\end{aligned}$$

Thus, we are led to the formula in the statement. □

And with this, good news, we have all needed tools for investigating both the wave and heat equations, in 3D. The story here, however, is quite technical, and in what follows we will rather apply this technology to a related equation, the Schrödinger one.

4d. Quantum mechanics

Let us discuss now some applications of the above to quantum mechanics. The main idea of Heisenberg was to use infinite matrices, based on the following fact:

FACT 4.10 (Rydberg, Ritz). *The spectral lines of the hydrogen atom are given by the Rydberg formula, as follows, depending on integer parameters $n_1 < n_2$:*

$$\frac{1}{\lambda_{n_1n_2}} = R\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$$

These spectral lines combine according to the Ritz-Rydberg principle, as follows:

$$\frac{1}{\lambda_{n_1n_2}} + \frac{1}{\lambda_{n_2n_3}} = \frac{1}{\lambda_{n_1n_3}}$$

Similar formulae hold for the other atoms, with suitable fine-tunings of the constant R.

To be more precise, all this is based on some key experiments and observations of Lyman, Balmer, Paschen, around 1890-1900. Mathematically now, the point is that the above combination principle reminds the multiplication formula $e_{n_1n_2}e_{n_2n_3} = e_{n_1n_3}$ for the elementary matrices $e_{ij} : e_j \to e_i$, which leads to the following principle:

PRINCIPLE 4.11 (Heisenberg). *Observables in quantum mechanics should be some sort of infinite matrices, generalizing the Lyman, Balmer, Paschen lines of the hydrogen atom, and multiplying between them as the matrices do, as to produce further observables.*

Before further developing this key idea, let us hear as well the point of view of Schrödinger, which came a few years later. His idea was to forget about exact things, and try to investigate the hydrogen atom statistically. Let us start with:

QUESTION 4.12. *In the context of the hydrogen atom, assuming that the proton is fixed, what is the probability density $\varphi_t(x)$ of the position of the electron e, at time t,*

$$P_t(e \in V) = \int_V \varphi_t(x)dx$$

as function of an intial probability density $\varphi_0(x)$? Moreover, can the corresponding equation be solved, and will this prove the Bohr claims for hydrogen, statistically?

In order to get familiar with this question, let us first look at examples coming from classical mechanics. In the context of a particle whose position at time t is given by $x_0 + \gamma(t)$, the evolution of the probability density will be given by:

$$\varphi_t(x) = \varphi_0(x) + \gamma(t)$$

However, such examples are somewhat trivial, of course not in relation with the computation of γ, usually a difficult question, but in relation with our questions, and do not apply to the electron. The point indeed is that, in what regards the electron, we have:

FACT 4.13. *In respect with various simple interference experiments:*

(1) *The electron is definitely not a particle in the usual sense.*
(2) *But in most situations it behaves exactly like a wave.*
(3) *But in other situations it behaves like a particle.*

Getting back now to the Schrödinger question, all this suggests to use, as for the waves, an amplitude function $\psi_t(x) \in \mathbb{C}$, related to the density $\varphi_t(x) > 0$ by the formula $\varphi_t(x) = |\psi_t(x)|^2$. Not that a big deal, you would say, because the two are related by simple formulae as follows, with $\theta_t(x)$ being an arbitrary phase function:

$$\varphi_t(x) = |\psi_t(x)|^2 \quad , \quad \psi_t(x) = e^{i\theta_t(x)}\sqrt{\varphi_t(x)}$$

However, such manipulations can be crucial, raising for instance the possibility that the amplitude function satisfies some simple equation, while the density itself, maybe not. And this is what happens indeed. Schrödinger was led in this way to:

CLAIM 4.14 (Schrödinger). *In the context of the hydrogen atom, the amplitude function of the electron $\psi = \psi_t(x)$ is subject to the Schrödinger equation*

$$ih\dot{\psi} = -\frac{h^2}{2m}\Delta\psi + V\psi$$

m being the mass, $h = h_0/2\pi$ the reduced Planck constant, and V the Coulomb potential of the proton. The same holds for movements of the electron under any potential V.

Observe the similarity with the wave equation $\ddot{f} = v^2\Delta f$, and with the heat equation $\dot{f} = \alpha\Delta f$ too. Many things can be said here. Following now Heisenberg and Schrödinger, and then especially Dirac, who did the axiomatization work, we have:

DEFINITION 4.15. *In quantum mechanics the states of the system are vectors of a Hilbert space H, and the observables of the system are linear operators*

$$T : H \to H$$

which can be densely defined, and are taken self-adjoint, $T = T^*$. The average value of such an observable T, evaluated on a state $\xi \in H$, is given by:

$$< T > = < T\xi, \xi >$$

In the context of the Schrödinger mechanics of the hydrogen atom, the Hilbert space is the space $H = L^2(\mathbb{R}^3)$ where the wave function ψ lives, and we have

$$< T > = \int_{\mathbb{R}^3} T(\psi) \cdot \bar{\psi}\, dx$$

which is called "sandwiching" formula, with the operators

$$x \quad , \quad -\frac{ih}{m}\nabla \quad , \quad -ih\nabla \quad , \quad -\frac{h^2\Delta}{2m} \quad , \quad -\frac{h^2\Delta}{2m} + V$$

representing the position, speed, momentum, kinetic energy, and total energy.

In other words, we are declaring here by axiom that various "sandwiching" formulae found before by Heisenberg, involving the operators at the end, hold true. And also, we are raising the possibility for other quantum mechanical systems, more complicated, to be described as well by the operators on a certain Hilbert space H, as above.

Let us go back now to the Schrödinger equation, from Claim 4.14. As a first observation, coming from some standard calculus, this equation conserves probability amplitudes, in agreement with the basic probabilistic requirement, $P = 1$ overall:

$$\int_{\mathbb{R}^3} |\psi_0|^2 = 1 \implies \int_{\mathbb{R}^3} |\psi_t|^2 = 1$$

In order to solve now the hydrogen atom, by using the Schrödinger equation, the idea will be that of reformulating this equation in spherical coordinates. We have:

Theorem 4.16. *The time-independent Schrödinger equation in spherical coordinates separates, for solutions of type $\phi = \rho(r)\alpha(s,t)$, into two equations, as follows,*

$$\frac{d}{dr}\left(r^2 \cdot \frac{d\rho}{dr}\right) - \frac{2mr^2}{h^2}(V-E)\rho = K\rho$$

$$\sin s \cdot \frac{d}{ds}\left(\sin s \cdot \frac{d\alpha}{ds}\right) + \frac{d^2\alpha}{dt^2} = -K \sin^2 s \cdot \alpha$$

with K being a constant, called radial equation, and angular equation.

Proof. We use the formula from Theorem 4.9 for the Laplace operator in spherical coordinates. The time-independent Schrödinger equation reformulates as:

$$(V-E)\phi = \frac{h^2}{2m}\left[\frac{1}{r^2} \cdot \frac{d}{dr}\left(r^2 \cdot \frac{d\phi}{dr}\right) + \frac{1}{r^2 \sin s} \cdot \frac{d}{ds}\left(\sin s \cdot \frac{d\phi}{ds}\right) + \frac{1}{r^2 \sin^2 s} \cdot \frac{d^2\phi}{dt^2}\right]$$

Let us look now for separable solutions for this latter equation, consisting of a radial part and an angular part, as in the statement, namely:

$$\phi(r,s,t) = \rho(r)\alpha(s,t)$$

By plugging this function into our equation, we obtain:

$$(V-E)\rho\alpha = \frac{h^2}{2m}\left[\frac{\alpha}{r^2} \cdot \frac{d}{dr}\left(r^2 \cdot \frac{d\rho}{dr}\right) + \frac{\rho}{r^2 \sin s} \cdot \frac{d}{ds}\left(\sin s \cdot \frac{d\alpha}{ds}\right) + \frac{\rho}{r^2 \sin^2 s} \cdot \frac{d^2\alpha}{dt^2}\right]$$

By multiplying everything by $2mr^2/(h^2\rho\alpha)$, and then moving the radial terms to the left, and the angular terms to the right, this latter equation can be written as follows:

$$\frac{2mr^2}{h^2}(V-E) - \frac{1}{\rho} \cdot \frac{d}{dr}\left(r^2 \cdot \frac{d\rho}{dr}\right) = \frac{1}{\alpha \sin^2 s}\left[\sin s \cdot \frac{d}{ds}\left(\sin s \cdot \frac{d\alpha}{ds}\right) + \frac{d^2\alpha}{dt^2}\right]$$

Since this latter equation is now separated between radial and angular variables, both sides must be equal to a certain constant $-K$, and this gives the result. $\square$

Let us first study the angular equation. The result here is as follows:

THEOREM 4.17. *The separated solutions $\alpha = \sigma(s)\theta(t)$ of the angular equation,*

$$\sin s \cdot \frac{d}{ds}\left(\sin s \cdot \frac{d\alpha}{ds}\right) + \frac{d^2\alpha}{dt^2} = -K \sin^2 s \cdot \alpha$$

are given by the following formulae, where $l \in \mathbb{N}$ is such that $K = l(l+1)$,

$$\sigma(s) = P_l^m(\cos s) \quad , \quad \theta(t) = e^{imt}$$

and where $m \in \mathbb{Z}$ is a constant, and with P_l^m being the Legendre function,

$$P_l^m(x) = (-1)^m (1-x^2)^{m/2} \left(\frac{d}{dx}\right)^m P_l(x)$$

where P_l are the Legendre polynomials, given by the following formula:

$$P_l(x) = \frac{1}{2^l l!}\left(\frac{d}{dx}\right)^l (x^2-1)^l$$

These solutions $\alpha = \sigma(s)\theta(t)$ are called spherical harmonics.

PROOF. This follows from some standard study, and with the comment that everything is taken up to linear combinations. We will normalize the wave function later. □

In relation now with the radial equation, we have the following manipulation:

PROPOSITION 4.18. *The radial equation, written with $K = l(l+1)$,*

$$(r^2\rho')' - \frac{2mr^2}{h^2}(V-E)\rho = l(l+1)\rho$$

takes with $\rho = u/r$ the following form, called modified radial equation,

$$Eu = -\frac{h^2}{2m} \cdot u'' + \left(V + \frac{h^2 l(l+1)}{2mr^2}\right) u$$

which is a time-independent 1D Schrödinger equation.

PROOF. With $\rho = u/r$ as in the statement, we have:

$$\rho = \frac{u}{r} \quad , \quad \rho' = \frac{u'r - u}{r^2} \quad , \quad (r^2\rho')' = u''r$$

By plugging this data into the radial equation, this becomes:

$$u''r - \frac{2mr}{h^2}(V-E)u = \frac{l(l+1)}{r} \cdot u$$

By multiplying everything by $h^2/(2mr)$, this latter equation becomes:

$$\frac{h^2}{2m} \cdot u'' - (V-E)u = \frac{h^2 l(l+1)}{2mr^2} \cdot u$$

But this gives the formula in the statement. □

It remains to solve the above equation, for the Coulomb potential of the proton. And we have here the following result, which proves the original claims by Bohr:

THEOREM 4.19 (Schrödinger). *In the case of the hydrogen atom, where V is the Coulomb potential of the proton, the modified radial equation, which reads*

$$Eu = -\frac{h^2}{2m} \cdot u'' + \left(-\frac{Ke^2}{r} + \frac{h^2 l(l+1)}{2mr^2}\right) u$$

leads to the Bohr formula for allowed energies,

$$E_n = -\frac{m}{2}\left(\frac{Ke^2}{h}\right)^2 \cdot \frac{1}{n^2}$$

with $n \in \mathbb{N}$, the binding energy being

$$E_1 \simeq -2.177 \times 10^{-18}$$

with means $E_1 \simeq -13.591$ eV.

PROOF. This is again something quite tricky. With $\gamma = \sqrt{-2mE}/h$, then $p = \gamma r$, and $S = -\gamma Ke^2/E$, the modified radial equation takes the following form:

$$u'' = \left(1 - \frac{S}{p} + \frac{l(l+1)}{p^2}\right) u$$

But a power series study of this latter equation shows that the solutions blow up, unless we have $S = 2n$, for a certain integer $n > l$, and this gives the result. □

For more on all this, we refer to any standard quantum mechanics book.

4e. Exercises

This was an exciting physics chapter, and as exercises on this, we have:

EXERCISE 4.20. *Learn about the various types of waves, appearing in real life.*

EXERCISE 4.21. *Learn about deducing the wave equation from linear elasticity.*

EXERCISE 4.22. *Have a look at chains and whips, with the wave equation in mind.*

EXERCISE 4.23. *Learn about Stefan-Boltzmann, and corrections to the heat equation.*

EXERCISE 4.24. *Learn about potentials, and the Laplace equation in electrostatics.*

EXERCISE 4.25. *Clarify our intuitive interpretation of the Laplace operator.*

EXERCISE 4.26. *Learn more, say from Rudin, about the harmonic functions.*

EXERCISE 4.27. *Learn as well about soap films, and related topics.*

As bonus exercise, and no surprise here, learn more quantum mechanics.

Part II

Smooth manifolds

CHAPTER 5

Smooth manifolds

5a. Ellipses, conics

Looking up to the skies, things are quite fascinating. The first thing that you see is the Sun, seemingly moving around the Earth on a circle, but a more careful study reveals that this circle is in fact an ellipse. As for the other stars and planets, these have all sort of weird trajectories, but a more careful study reveals that, with due attention to what the best "center" is, replacing our Earth, the trajectories are often ellipses:

(1) Indeed, this applies to all the planets in our Solar System, which move around the biggest object in the system, which is by far the Sun, on ellipses.

(2) The same trick applies to the trajectories of various distant stars, the rule being always the same, "small moves around big, on an ellipse".

(3) However, there are counterexamples too, such as asteroids reaching our Solar system, but then traveling outwards, never to be seen again.

Summarizing, modulo some annoying asteroids that we will leave for later, we are led in this way to ellipses, and their mathematics. And good news, a full theory of ellipses is available, and this since the ancient Greeks, whose main findings were as follows:

THEOREM 5.1. *The ellipses, taken centered at the origin* 0*, and squarely oriented with respect to* Oxy*, can be defined in* 4 *possible ways, as follows:*

(1) *As the curves given by an equation as follows, with* $a,b>0$*:*

$$\left(\frac{x}{a}\right)^2+\left(\frac{y}{b}\right)^2=1$$

(2) *Or given by an equation as follows, with* $q>0$*,* $p=-q$*, and* $l\in(0,2q)$*:*

$$d(z,p)+d(z,q)=l$$

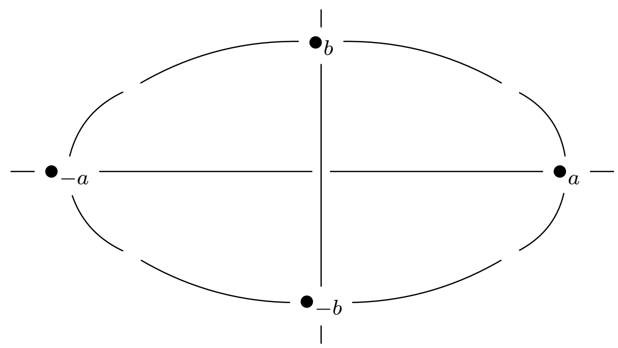
(3) *As the curves appearing when drawing a circle, from various perspectives:*

$$\bigcirc\quad\to\quad ?$$

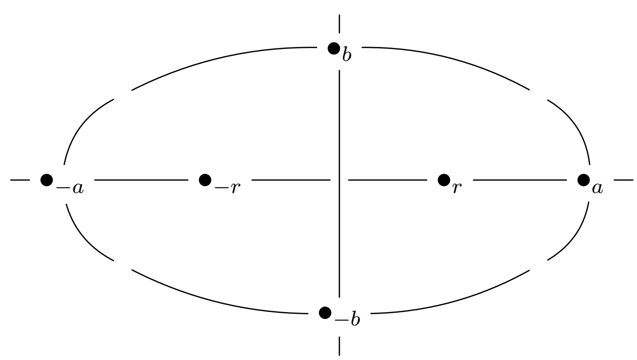
(4) *As the closed non-degenerate curves appearing by cutting a cone with a plane.*

Proof. This might look a bit confusing, and you might say, what exactly is to be proved here. Good point, and in answer, what is to be proved is that the above constructions (1-4) give rise to the same class of curves. And this can be done as follows:

(1) To start with, let us draw a picture from what comes out of (1), which will be our main definition for the ellipses, in what follows. Here that is, making it clear what the parameters $a, b > 0$ stand for, with $2a \times 2b$ being the gift box size for our ellipse:



(2) Let us prove now that such an ellipse has two focal points, as stated in (2). We must look for a number $r > 0$, and a number $l > 0$, such that our ellipse appears as $d(z,p) + d(z,q) = l$, with $p = (0,-r)$ and $q = (0,r)$, according to the following picture:



(3) Let us first compute these numbers $r, l > 0$. Assuming that our result holds indeed as stated, by taking $z = (0,a)$, we see that the length l is:

$$l = (a-r) + (a+r) = 2a$$

As for the parameter r, by taking $z = (b,0)$, we conclude that we must have:

$$2\sqrt{b^2 + r^2} = 2a \implies r = \sqrt{a^2 - b^2}$$

(4) With these observations made, let us prove now the result. Given $l,r>0$, and setting $p=(0,-r)$ and $q=(0,r)$, we have the following computation, with $z=(x,y)$:

$$\begin{aligned}
& d(z,p)+d(z,q)=l\\
\iff\ & \sqrt{(x+r)^2+y^2}+\sqrt{(x-r)^2+y^2}=l\\
\iff\ & \sqrt{(x+r)^2+y^2}=l-\sqrt{(x-r)^2+y^2}\\
\iff\ & (x+r)^2+y^2=(x-r)^2+y^2+l^2-2l\sqrt{(x-r)^2+y^2}\\
\iff\ & 2l\sqrt{(x-r)^2+y^2}=l^2-4xr\\
\iff\ & 4l^2(x^2+r^2-2xr+y^2)=l^4+16x^2r^2-8l^2xr\\
\iff\ & 4l^2x^2+4l^2r^2+4l^2y^2=l^4+16x^2r^2\\
\iff\ & (4x^2-l^2)(4r^2-l^2)=4l^2y^2
\end{aligned}$$

(5) Now observe that we can further process the equation that we found as follows:

$$\begin{aligned}
(4x^2-l^2)(4r^2-l^2)=4l^2y^2 \iff\ & \frac{4x^2-l^2}{l^2}=\frac{4y^2}{4r^2-l^2}\\
\iff\ & \frac{4x^2-l^2}{l^2}=\frac{y^2}{r^2-l^2/4}\\
\iff\ & \left(\frac{x}{2l}\right)^2-1=\left(\frac{y}{\sqrt{r^2-l^2/4}}\right)^2\\
\iff\ & \left(\frac{x}{2l}\right)^2+\left(\frac{y}{\sqrt{r^2-l^2/4}}\right)^2=1
\end{aligned}$$

(6) Thus, our result holds indeed, and with the numbers $l,r>0$ appearing, and no surprise here, via the formulae $l=2a$ and $r=\sqrt{a^2-b^2}$, found in (3) above.

(7) Getting back now to our theorem, we have two other assertions there at the end, labeled (3,4). But, thinking a bit, these assertions are in fact equivalent, and in what concerns us, we will rather focus on (4), which looks more mathematical. And in what regards this assertion (4), this can be established indeed, by doing some 3D computations, that we will leave here as an instructive exercise, for you. And with the promise that we will come back to this in a moment, with a full proof, in a more general setting. □

All this is very nice, but before getting into physics, with some explanations for the fact that planets travel indeed on ellipses, which is something that we must surely understand, before going with some further math, let us settle as well the question of wandering asteroids. Observations show that these can travel on parabolas and hyperbolas, so what we need as mathematics is a unified theory of ellipses, parabolas and hyperbolas. And fortunately, this theory exists, also since the ancient Greeks, summarized as follows:

Theorem 5.2. *The conics, which are the algebraic curves of degree* 2 *in the plane,*

$$C = \left\{(x,y)\in\mathbb{R}^2\Big|P(x,y)=0\right\}$$

with $\deg P\leq 2$*, appear modulo degeneration by cutting a* 2*-sided cone with a plane, and can be classified into ellipses, parabolas and hyperbolas.*

Proof. This follows by further building on Theorem 5.1, as follows:

(1) Let us first classify the conics up to non-degenerate linear transformations of the plane, which are by definition transformations as follows, with $\det A\neq 0$:

$$\begin{pmatrix}x\\y\end{pmatrix}\to A\begin{pmatrix}x\\y\end{pmatrix}$$

Our claim is that as solutions we have the circles, parabolas, hyperbolas, along with some degenerate solutions, namely $\emptyset$, points, lines, pairs of lines, $\mathbb{R}^2$.

(2) As a first remark, it looks like we forgot precisely the ellipses, but via linear transformations these become circles, so things fine. As a second remark, all our claimed solutions can appear. Indeed, the circles, parabolas, hyperbolas can appear as follows:

$$x^2+y^2=1\quad,\quad x^2=y\quad,\quad xy=1$$

As for $\emptyset$, points, lines, pairs of lines, $\mathbb{R}^2$, these can appear too, as follows, and with our polynomial P chosen, whenever possible, to be of degree exactly 2:

$$x^2=-1\quad,\quad x^2+y^2=0\quad,\quad x^2=0\quad,\quad xy=0\quad,\quad 0=0$$

Observe here that, when dealing with these degenerate cases, assuming $\deg P=2$ instead of $\deg P\leq 2$ would only rule out $\mathbb{R}^2$ itself, which is not worth it.

(3) Getting now to the proof of our claim in (1), classification up to linear transformations, consider an arbitrary conic, written as follows, with $a,b,c,d,e,f\in\mathbb{R}$:

$$ax^2+by^2+cxy+dx+ey+f=0$$

Assume first $a\neq 0$. By making a square out of ax^2, up to a linear transformation in (x,y), we can get rid of the term cxy, and we are left with:

$$ax^2+by^2+dx+ey+f=0$$

In the case $b\neq 0$ we can make two obvious squares, and again up to a linear transformation in (x,y), we are left with an equation as follows:

$$x^2\pm y^2=k$$

In the case of positive sign, $x^2+y^2=k$, the solutions are the circle, when $k\geq 0$, the point, when $k=0$, and $\emptyset$, when $k<0$. As for the case of negative sign, $x^2-y^2=k$, which reads $(x-y)(x+y)=k$, here once again by linearity our equation becomes $xy=l$, which is a hyperbola when $l\neq 0$, and two lines when $l=0$.

(4) In the case $b\neq 0$ the study is similar, with the same solutions, so we are left with the case $a=b=0$. Here our conic is as follows, with $c,d,e,f\in\mathbb R$:

$$cxy+dx+ey+f=0$$

If $c\neq 0$, by linearity our equation becomes $xy=l$, which produces a hyperbola or two lines, as explained before. As for the remaining case, $c=0$, here our equation is:

$$dx+ey+f=0$$

But this is generically the equation of a line, unless we are in the case $d=e=0$, where our equation is $f=0$, having as solutions $\emptyset$ when $f\neq 0$, and $\mathbb R^2$ when $f=0$.

(5) Thus, done with the classification, up to linear transformations as in (1). But this classification leads to the classification in general too, by applying now linear transformations to the solutions that we found. So, done with this, and very good.

(6) It remains to discuss the cone cutting. By suitably choosing our coordinate axes (x,y,z), we can assume that our cone is given by an equation as follows, with $k>0$:

$$x^2+y^2=kz^2$$

In order to prove the result, we must in principle intersect this cone with an arbitrary plane, which has an equation as follows, with $(a,b,c)\neq(0,0,0)$:

$$ax+by+cz=d$$

(7) However, before getting into computations, observe that what we want to find is a certain degree 2 equation in the above plane, for the intersection. Thus, it is convenient to change the coordinates, as for our plane to be given by the following equation:

$$z=0$$

(8) But with this done, what we have to do is to see how the cone equation $x^2+y^2=kz^2$ changes, under this change of coordinates, and then set $z=0$, as to get the (x,y) equation of the intersection. But this leads, via some thinking or computations, to the conclusion that the cone equation $x^2+y^2=kz^2$ becomes in this way a degree 2 equation in (x,y), which can be arbitrary, and so to the final conclusion in the statement. □

Ready for some physics? Here is the main result here, due to Kepler and Newton:

THEOREM 5.3. *Planets and other celestial bodies move around the Sun on conics,*

$$C=\left\{(x,y)\in\mathbb R^2\Big|P(x,y)=0\right\}$$

with $P\in\mathbb R[x,y]$ being of degree 2, which can be ellipses, parabolas or hyperbolas.

PROOF. This is something quite long, due to Kepler and Newton, but no fear, we know calculus, and threrefore what can resist us. The proof goes as follows:

(1) According to countless observations and calculations, first formalized by Newton, the force of attraction between two bodies of masses M, m is given by:

$$||F|| = G \cdot \frac{Mm}{d^2}$$

Here d is the distance between the two bodies, and $G \simeq 6.674 \times 10^{-11}$ is a constant. Now assuming that M is fixed at $0 \in \mathbb{R}^3$, the force exterted on m positioned at $x \in \mathbb{R}^3$, regarded as a vector $F \in \mathbb{R}^3$, is given by the following formula:

$$\begin{aligned} F &= -||F|| \cdot \frac{x}{||x||} \\ &= -\frac{GMm}{||x||^2} \cdot \frac{x}{||x||} \\ &= -\frac{GMmx}{||x||^3} \end{aligned}$$

But $F = ma = m\ddot{x}$, with $a = \ddot{x}$ being the acceleration, second derivative of the position, so the equation of motion of m, assuming that M is fixed at 0, is:

$$\ddot{x} = -\frac{GMx}{||x||^3}$$

Obviously, the problem happens in 2 dimensions, and you can even find, as an exercise, a formal proof of that, based on the above equation, if your really want to. Now here the most convenient is to use standard x, y coordinates, and denote our point as $z = (x, y)$. With this change made, and by setting $K = GM$, the equation of motion becomes:

$$\ddot{z} = -\frac{Kz}{||z||^3}$$

(2) The idea now is that the problem can be solved via some calculus. Let us write indeed our vector $z = (x, y)$ in polar coordinates, as follows:

$$x = r\cos\theta \quad , \quad y = r\sin\theta$$

We have then $||z|| = r$, and our equation of motion becomes:

$$\ddot{z} = -\frac{Kz}{r^3}$$

Let us differentiate now x, y. By using the standard calculus rules, we have:

$$\dot{x} = \dot{r}\cos\theta - r\sin\theta \cdot \dot{\theta}$$

$$\dot{y} = \dot{r}\sin\theta + r\cos\theta \cdot \dot{\theta}$$

Differentiating one more time gives the following formulae:

$$\ddot{x} = \ddot{r}\cos\theta - 2\dot{r}\sin\theta \cdot \dot{\theta} - r\cos\theta \cdot \dot{\theta}^2 - r\sin\theta \cdot \ddot{\theta}$$

$$\ddot{y} = \ddot{r}\sin\theta + 2\dot{r}\cos\theta \cdot \dot{\theta} - r\sin\theta \cdot \dot{\theta}^2 + r\cos\theta \cdot \ddot{\theta}$$

Consider now the following two quantities, appearing as coefficients in the above:

$$a=\ddot{r}-r\dot{\theta}^2\quad,\quad b=2\dot{r}\dot{\theta}+r\ddot{\theta}$$

In terms of these quantities, our second derivative formulae read:

$$\ddot{x}=a\cos\theta-b\sin\theta$$

$$\ddot{y}=a\sin\theta+b\cos\theta$$

(3) We can now solve the equation of motion from (1). Indeed, with the formulae that we found for $\ddot{x},\ddot{y}$, our equation of motion takes the following form:

$$a\cos\theta-b\sin\theta=-\frac{K}{r^2}\cos\theta$$

$$a\sin\theta+b\cos\theta=-\frac{K}{r^2}\sin\theta$$

But these two formulae can be written in the following way:

$$\left(a+\frac{K}{r^2}\right)\cos\theta=b\sin\theta\quad,\quad\left(a+\frac{K}{r^2}\right)\sin\theta=-b\cos\theta$$

By making now the product, and assuming that we are in a non-degenerate case, where the angle θ varies indeed, we obtain by positivity that we must have:

$$a+\frac{K}{r^2}=b=0$$

(4) Let us first examine the second equation, $b=0$. This can be solved as follows:

$$\begin{aligned}b=0&\iff 2\dot{r}\dot{\theta}+r\ddot{\theta}=0\\&\iff\frac{\ddot{\theta}}{\dot{\theta}}=-2\frac{\dot{r}}{r}\\&\iff(\log\dot{\theta})'=(-2\log r)'\\&\iff\log\dot{\theta}=-2\log r+c\\&\iff\dot{\theta}=\frac{\lambda}{r^2}\end{aligned}$$

As for the first equation the we found, namely $a+K/r^2=0$, this becomes:

$$\ddot{r}-\frac{\lambda^2}{r^3}+\frac{K}{r^2}=0$$

As a conclusion to all this, in polar coordinates, $x=r\cos\theta$, $y=r\sin\theta$, our equations of motion are as follows, with λ being a constant, not depending on t:

$$\ddot{r}=\frac{\lambda^2}{r^3}-\frac{K}{r^2}\quad,\quad\dot{\theta}=\frac{\lambda}{r^2}$$

Even better now, by writing $K = \lambda^2/c$, these equations read:

$$\ddot{r} = \frac{\lambda^2}{r^2}\left(\frac{1}{r} - \frac{1}{c}\right) \quad , \quad \dot{\theta} = \frac{\lambda}{r^2}$$

(5) In order to study the first equation, we use a trick. Let us write:

$$r(t) = \frac{1}{f(\theta(t))}$$

Abbreviated, and by reminding that f takes $\theta = \theta(t)$ as variable, this reads:

$$r = \frac{1}{f}$$

With the convention that dots mean as usual derivatives with respect to t, and that the primes will denote derivatives with respect to $\theta = \theta(t)$, we have:

$$\dot{r} = -\frac{f'\dot{\theta}}{f^2} = -\frac{f'}{f^2} \cdot \frac{\lambda}{r^2} = -\lambda f'$$

By differentiating one more time with respect to t, we obtain:

$$\ddot{r} = -\lambda f''\dot{\theta} = -\lambda f'' \cdot \frac{\lambda}{r^2} = -\frac{\lambda^2}{r^2} f''$$

On the other hand, our equation for $\ddot{r}$ found in (4) above reads:

$$\ddot{r} = \frac{\lambda^2}{r^2}\left(\frac{1}{r} - \frac{1}{c}\right) = \frac{\lambda^2}{r^2}\left(f - \frac{1}{c}\right)$$

Thus, in terms of $f = 1/r$ as above, our equation for $\ddot{r}$ simply reads:

$$f'' + f = \frac{1}{c}$$

But this latter equation is elementary to solve. Indeed, both functions $\cos t, \sin t$ satisfy $g" + g = 0$, so any linear combination of them satisfies as well this equation. But the solutions of $f'' + f = 1/c$ being those of $g'' + g = 0$ shifted by $1/c$, we obtain:

$$f = \frac{1 + \varepsilon \cos\theta + \delta \sin\theta}{c}$$

Now by inverting, we obtain the following formula:

$$r = \frac{c}{1 + \varepsilon \cos\theta + \delta \sin\theta}$$

(6) But this leads to the conclusion that the trajectory is a conic. Indeed, in terms of the parameter θ, the formulae of the coordinates are:

$$x = \frac{c\cos\theta}{1 + \varepsilon \cos\theta + \delta \sin\theta} \quad , \quad y = \frac{c\sin\theta}{1 + \varepsilon \cos\theta + \delta \sin\theta}$$

Now observe that these two functions x, y satisfy the following formula:

$$x^2 + y^2 = \frac{c^2(\cos^2\theta + \sin^2\theta)}{(1 + \varepsilon\cos\theta + \delta\sin\theta)^2} = \frac{c^2}{(1 + \varepsilon\cos\theta + \delta\sin\theta)^2}$$

On the other hand, these two functions satisfy as well the following formula:

$$\begin{aligned}(\varepsilon x + \delta y - c)^2 &= \frac{c^2\big(\varepsilon\cos\theta + \delta\sin\theta - (1 + \varepsilon\cos\theta + \delta\sin\theta)\big)^2}{(1 + \varepsilon\cos\theta + \delta\sin\theta)^2}\\ &= \frac{c^2}{(1 + \varepsilon\cos\theta + \delta\sin\theta)^2}\end{aligned}$$

We conclude that our coordinates x, y satisfy the following equation:

$$x^2 + y^2 = (\varepsilon x + \delta y - c)^2$$

But what we have here is an equation of a conic, and we are done. □

With the above formulae in hand, we can work out how various inital speeds and accelerations lead to various types of conics. The computations here are many, and very interesting, and we will leave them as a long, pleasant and instructive exercise.

5b. Curves, surfaces

As a conclusion to what we did so far, conics are at the core of everything, mathematics, physics, life. But, what is next? A natural answer to this question comes from:

Definition 5.4. *An algebraic curve in $\mathbb{R}^2$ is the vanishing set*

$$C = \left\{(x, y) \in \mathbb{R}^2 \Big| P(x, y) = 0\right\}$$

of a polynomial $P \in \mathbb{R}[X, Y]$ of arbitrary degree.

We already know well the algebraic curves in degree 2, which are the conics, and a first problem is, what results from what we learned about conics have a chance to be relevant to the arbitrary algebraic curves. And normally none, because the ellipses, parabolas and hyperbolas are obviously very particular curves, having very particular properties.

Let us record however a useful statement here, as follows:

Proposition 5.5. *The conics can be written in cartesian, polar, parametric or complex coordinates, with the equations for the unit circle being*

$$x^2 + y^2 = 1 \quad , \quad r = 1 \quad , \quad x = \cos t\,,\ y = \sin t \quad , \quad |z| = 1$$

and with the equations for ellipses, parabolas and hyperbolas being similar.

Proof. The equations for the circle are clear, those for ellipses can be found in the above, and we will leave as an exercise those for parabolas and hyperbolas. □

As a true answer to our question now, coming this time from a very modest conic, namely $xy = 0$, that we dismissed in the above as being "degenerate", we have:

PROPOSITION 5.6. *The following happen, for curves C defined by polynomials P:*

(1) *In degree $d = 2$, curves can have singularities, such as $xy = 0$ at $(0, 0)$.*
(2) *In general, assuming $P = P_1 \ldots P_k$, we have $C = C_1 \cup \ldots\ldots \cup C_k$.*
(3) *A union of curves $C_i \cup C_j$ is generically non-smooth, unless disjoint.*
(4) *Due to this, we say that C is non-degenerate when P is irreducible.*

PROOF. All this is self-explanatory, the details being as follows:

(1) This is something obvious, just the story of two lines crossing.

(2) This comes from the following trivial fact, with the notation $z = (x, y)$:

$$P_1 \ldots P_k(z) = 0 \iff P_1(z) = 0, \text{ or } P_2(z) = 0, \ \ldots\ , \text{ or } P_k(z) = 0$$

(3) This is something very intuitive, and it actually takes a bit of time to imagine a situation where $C_1 \cap C_2 \neq \emptyset$, $C_1 \not\subset C_2$, $C_2 \not\subset C_1$, but $C_1 \cup C_2$ is smooth. In practice now, "generically" has of course a mathematical meaning, in relation with probability, and our assertion does say something mathematical, that we are supposed to prove. But, we will not insist on this, and leave this as an instructive exercise, precise formulation of the claim, and its proof, in the case you are familiar with probability theory.

(4) This is just a definition, based on the above, that we will use in what follows. □

With degree 1 and 2 investigated, and our conclusions recorded, let us get now to degree 3, see what new phenomena appear here. And here, to start with, we have the following remarkable curve, well-known from calculus, because 0 is not a maximum or minimum of the function $x \to y$, despite the derivative vanishing there:

$$x^3 = y$$

Also, in relation with set theory and logic, and with the foundations of mathematics in general, we have the following curve, which looks like the empyset $\emptyset$:

$$(x - y)(x^2 + y^2 - 1) = 0$$

But, it is not about counterexamples to calculus, or about logic, that we want to talk about here. As a first truly remarkable degree 3 curve, or cubic, we have the cusp:

PROPOSITION 5.7. *The standard cusp, which is the cubic given by*

$$x^3 = y^2$$

has a singularity at $(0, 0)$, with only 1 tangent line at that singularity.

Proof. The two branches of the cusp are indeed both tangent to Ox, because:

$$y' = \pm\frac{3}{2}\sqrt{x} \implies y'(0) = 0$$

Observe also that what happens for the cusp is different from what happens for $xy = 0$, precisely because we have 1 line tangent at the singularity, instead of 2. □

As a second remarkable cubic, which gets the crown, and the right to have a Theorem about it, we have the Tschirnhausen curve, which is as follows:

Theorem 5.8. *The Tschirnhausen cubic, given by the following equation,*

$$x^3 = x^2 - 3y^2$$

makes the dream of $xy = 0$ come true, by self-intersecting, and being non-degenerate.

Proof. This is something self-explanatory, by drawing a picture, but there are several other interesting things that can be said about this curve, and the family of curves containing it, depending on a parameter, and up to basic transformations, as follows:

(1) Let us start with the curve written in polar coordinates as follows:

$$r\cos^3\left(\frac{\theta}{3}\right) = a$$

With $t = \tan(\theta/3)$, the equations of the coordinates are as follows:

$$x = a(1-3t^2) \quad , \quad y = at(3-t^2)$$

Now by eliminating t, we reach to the following equation:

$$(a-x)(8a+x)^2 = 27ay^2$$

(2) By translating horizontally by $8a$, and changing signs of variables, we have:

$$x = 3a(3-t^2) \quad , \quad y = at(3-t^2)$$

Now by eliminating t, we reach to the following equation:

$$x^3 = 9a(x^2 - 3y^2)$$

But with $a = 1/9$ this is precisely the equation in the statement. □

Let us get now to $\mathbb{R}^3$. Here we are right away into a dillema, because the plane curves have two possible generalizations. First we have the algebraic curves in $\mathbb{R}^3$:

Definition 5.9. *An algebraic curve in $\mathbb{R}^3$ is a curve as follows,*

$$C = \left\{(x,y,z) \in \mathbb{R}^3 \Big| P(x,y,z) = 0,\ Q(x,y,z) = 0\right\}$$

appearing as the joint zeroes of two polynomials P, Q.

These curves look of course like the usual plane curves, and at the level of the phenomena that can appear, these are similar to those in the plane, involving singularities and so on, but also knotting, which is a new phenomenon. However, it is hard to say something with bare hands about knots. We will be back to this, later in this book.

On the other hand, as another natural generalization of the plane curves, and this might sound a bit surprising, we have the surfaces in $\mathbb{R}^3$, constructed as follows:

Definition 5.10. *An algebraic surface in $\mathbb{R}^3$ is a surface as follows,*

$$S=\left\{(x,y,z)\in\mathbb{R}^3\Big|P(x,y,z)=0\right\}$$

appearing as the zeroes of a polynomial P.

The point indeed is that, as it was the case with the plane curves, what we have here is something defined by a single equation. And with respect to many questions, having a single equation matters a lot, and this is why surfaces in $\mathbb{R}^3$ are "simpler" than curves in $\mathbb{R}^3$. In fact, believe me, they are even the correct generalization of the curves in $\mathbb{R}^2$.

As an example of what can be done with surfaces, which is very similar to what we did with the conics $C\subset\mathbb{R}^2$ earlier in this chapter, we have the following result:

Theorem 5.11. *The degree 2 surfaces $S\subset\mathbb{R}^3$, called quadrics, are the ellipsoid*

$$\left(\frac{x}{a}\right)^2+\left(\frac{y}{b}\right)^2+\left(\frac{z}{c}\right)^2=1$$

which is the only compact one, plus 16 more, which can be explicitly listed.

Proof. We will be quite brief here, because we intend to rediscuss all this in a moment, with full details, in arbitrary N dimensions, the idea being as follows:

(1) The equations for a quadric $S\subset\mathbb{R}^2$ are best written as follows, with $A\in M_3(\mathbb{R})$ being a matrix, $B\in M_{1\times3}(\mathbb{R})$ being a row vector, and $C\in\mathbb{R}$ being a constant:

$$<Au,u>+Bu+C=0$$

(2) By doing now the linear algebra, and we will come back to this in a moment, with details, or by invoking the theorem of Sylvester on quadratic forms, we are left, modulo degeneracy and linear transformations, with signed sums of squares, as follows:

$$\pm x^2\pm y^2\pm z^2=0,1$$

(3) Thus the sphere is the only compact quadric, up to linear transformations, and by applying now linear transformations to it, we are led to the ellipsoids in the statement.

(4) As for the other quadrics, there are many of them, a bit similar to the parabolas and hyperbolas in 2 dimensions, and some work here leads to a 16 item list. □

With this done, instead of further insisting on the surfaces $S \subset \mathbb{R}^3$, or getting into their rivals, the curves $C \subset \mathbb{R}^3$, which appear as intersections of such surfaces, $C = S \cap S'$, let us get instead to arbitrary N dimensions, see what the axiomatics looks like there, with the hope that this will clarify our dimensionality dillema, curves vs surfaces.

So, moving to N dimensions, we have here the following definition, to start with:

DEFINITION 5.12. *An algebraic hypersurface in $\mathbb{R}^N$ is a space of the form*

$$S = \left\{ (x_1, \ldots, x_N) \in \mathbb{R}^N \middle| P(x_1, \ldots, x_N) = 0, \forall i \right\}$$

appearing as the zeroes of a polynomial $P \in \mathbb{R}[x_1, \ldots, x_N]$.

Again, this is a quite general definition, covering both the plane curves $C \subset \mathbb{R}$ and the surfaces $S \subset \mathbb{R}^2$, which is certainly worth a systematic exploration. But, no hurry with this, for the moment we are here for talking definitons and axiomatics.

In order to have now a full collection of beasts, in all possible dimensions $N \in \mathbb{N}$, and of all possible dimensions $k \in \mathbb{N}$, we must intersect such algebraic hypersurfaces. We are led in this way to the zeroes of families of polynomials, as follows:

DEFINITION 5.13. *An algebraic manifold in $\mathbb{R}^N$ is a space of the form*

$$X = \left\{ (x_1, \ldots, x_N) \in \mathbb{R}^N \middle| P_i(x_1, \ldots, x_N) = 0, \forall i \right\}$$

with $P_i \in \mathbb{R}[x_1, \ldots, x_N]$ being a family of polynomials.

As a first observation, as already mentioned, such a manifold appears as an intersection of hypersurfaces S_i, those associated to the various polynomials P_i:

$$X = S_1 \cap \ldots \cap S_r$$

There is actually a bit of a discussion needed here, regarding the parameter $r \in \mathbb{N}$, shall we allow this parameter to be $r = \infty$ too, or not. We will discuss this later, with some algebra helping, the idea being that allowing $r = \infty$ forces in fact $r < \infty$.

As an announcement now, good news, what we have in Definition 5.13 is the good and final notion of algebraic manifold, very general, and with the branch of mathematics studying such manifolds being called algebraic geometry. In what follows we will discuss a bit what can be done with this, as a continuation of our previous work on the plane curves, at the elementary level. All this will lead us into the conclusion that we must first develop commutative algebra, and come back to algebraic geometry afterwards.

Let us first look more in detail at the hypersurfaces. We have here:

Theorem 5.14. *The degree 2 hypersurfaces $S\subset\mathbb{R}^N$, called quadrics, are up to degeneracy and to linear transformations the hypersurfaces of the following form,*

$$\pm x_1^2\pm\ldots\pm x_N^2=0,1$$

and with the sphere being the only compact one.

Proof. We have two statements here, the idea being as follows:

(1) The equations for a quadric $S\subset\mathbb{R}^N$ are best written as follows, with $A\in M_N(\mathbb{R})$ being a matrix, $B\in M_{1\times N}(\mathbb{R})$ being a row vector, and $C\in\mathbb{R}$ being a constant:

$$< Ax,x > +Bx+C=0$$

(2) By doing the linear algebra, or by invoking the theorem of Sylvester on quadratic forms, we are left, modulo linear transformations, with signed sums of squares:

$$\pm x_1^2\pm\ldots\pm x_N^2=0,1$$

(3) To be more precise, with linear algebra, by evenly distributing the terms x_ix_j above and below the diagonal, we can assume that our matrix $A\in M_N(\mathbb{R})$ is symmetric. Thus A must be diagonalizable, and by changing the basis of $\mathbb{R}^N$, as to have it diagonal, our equation becomes as follows, with $D\in M_N(\mathbb{R})$ being now diagonal:

$$< Dx,x > +Ex+F=0$$

(4) But now, by making squares in the obvious way, which amounts in applying yet another linear transformation to our quadric, the equation takes the following form, with $G\in M_N(-1,0,1)$ being diagonal, and with $H\in\{0,1\}$ being a constant:

$$< Gx,x >= H$$

(5) Now barring the degenerate cases, we can further assume $G\in M_N(-1,1)$, and we are led in this way to the equation claimed in (2) above, namely:

$$\pm x_1^2\pm\ldots\pm x_N^2=0,1$$

(6) In particular we see that, up to some degenerate cases, namely emptyset and point, the only compact quadric, up to linear transformations, is the one given by:

$$x_1^2+\ldots+x_N^2=1$$

(7) But this is the unit sphere, so are led to the conclusions in the statement. □

Regarding now the examples of hypersurfaces $S\subset\mathbb{R}^N$, or of more general algebraic manifolds $X\subset\mathbb{R}^N$, there are countless of them, and it is impossible to have some discussion started here, without being subjective. The unit sphere $S^{N-1}_\mathbb{R}\subset\mathbb{R}^N$ gets of course the crown from everyone, as being the most important manifold after $\mathbb{R}^N$ itself. But then, passed this sphere, things ramify, depending on what exact applications of algebraic geometry you have in mind. In what concerns me, here is my next favorite example:

THEOREM 5.15. *The invertible matrices $A \in M_N(\mathbb{R})$ lie outside the hypersurface*

$$\det A = 0$$

and are therefore dense, in the space of all matrices $M_N(\mathbb{R})$.

PROOF. This is something self-explanatory, but with this result being some key in linear algebra, all this is worth a detailed discussion, as follows:

(1) We certainly know from basic linear algebra that a matrix $A \in M_N(\mathbb{R})$ is invertible precisely when it has nonzero determinant, $\det A \neq 0$. Thus, the invertible matrices $A \in M_N(\mathbb{R})$ are located precisely in the complement of the following space:

$$S = \left\{A \in M_N(\mathbb{R}) \Big| \det A = 0\right\}$$

(2) We also know from basic linear algebra, or perhaps not so basic linear algebra, that the determinant $\det A$ is a certain polynomial in the entries of A, of degree N:

$$\det \in \mathbb{R}[X_{11}, \ldots, X_{NN}]$$

(3) We conclude from this that the above set S is a degree N algebraic hypersurface in our sense, in the Euclidean space $M_N(\mathbb{R}) \simeq \mathbb{R}^n$, with $n = N^2$.

(4) Now since the complements of non-trivial hypersurfaces $S \subset \mathbb{R}^n$ are obviously dense, and if needing a formal proof here, for our above hypersurface S this is clear, simply by suitably perturbing the matrix, and in general do not worry, we will be back to this, with full details, we are led to the conclusions in the statement. □

5c. Smooth manifolds

We already know about algebraic curves, then surfaces and other algebraic manifolds, generalizing the conics, from the above. A second idea now, in order to generalize the conics, is to look at the smooth manifolds, in the following sense:

DEFINITION 5.16. *A smooth manifold is a space X which is locally isomorphic to $\mathbb{R}^N$. To be more precise, this space X must be covered by charts, bijectively mapping open pieces of it to open pieces of $\mathbb{R}^N$, with the changes of charts being C^∞ functions.*

It is of course possible to talk as well about C^k manifolds, with $k < \infty$, but this is rather technical material, that we will not get into, in this book.

As basic examples of smooth manifolds, we have $\mathbb{R}^N$ itself, or any open subset $X \subset \mathbb{R}^N$, with only 1 chart being needed here. Other basic examples include the circle, or curves like ellipses and so on, for obvious reasons. To be more precise, the unit circle can be covered by 2 charts as above, by using polar coordinates, in the obvious way, and then by applying dilations, translations and other such transformations, namely bijections which are smooth, we obtain a whole menagery of circle-looking manifolds.

Here is a more precise statement in this sense, covering the conics:

THEOREM 5.17. *The following are smooth manifolds, in the plane:*
(1) *The circles.*
(2) *The ellipses.*
(3) *The non-degenerate conics.*
(4) *Smooth deformations of these.*

PROOF. All this is quite intuitive, the idea being as follows:

(1) Consider the unit circle, $x^2 + y^2 = 1$. We can write then $x = \cos t$, $y = \sin t$, with $t \in [0, 2\pi)$, and we seem to have here the solution to our problem, just using 1 chart. But this is of course wrong, because $[0, 2\pi)$ is not open, and we have a problem at 0. In practice we need to use 2 such charts, say with the first one being with $t \in (0, 3\pi/2)$, and the second one being with $t \in (\pi, 5\pi/2)$. As for the fact that the change of charts is indeed smooth, this comes by writing down the formulae, or just thinking a bit, and arguing that this change of chart being actually a translation, it is automatically linear.

(2) This follows from (1), by pulling the circle in both the Ox and Oy directions, and the formulae here, based on those for ellipses from chapter 1, are left to you reader.

(3) We already have the ellipses, and the case of the parabolas and hyperbolas is elementary as well, and in fact simpler than the case of the ellipses. Indeed, a parablola is clearly homeomorphic to $\mathbb{R}$, and a hyperbola, to two copies of $\mathbb{R}$.

(4) This is something which is clear too, depending of course on what exactly we mean by "smooth deformation", and by using a bit of multivariable calculus if needed. $\square$

In higher dimensions, as basic examples, we have the spheres, as shown by:

THEOREM 5.18. *The sphere is a smooth manifold.*

PROOF. There are several proofs for this, all instructive, as follows:

(1) A first idea is to use spherical coordinates, which are as follows:

$$\begin{cases} x_1 &= r\cos t_1 \\ x_2 &= r\sin t_1\cos t_2 \\ \vdots \\ x_{N-1} &= r\sin t_1\sin t_2\ldots\sin t_{N-2}\cos t_{N-1} \\ x_N &= r\sin t_1\sin t_2\ldots\sin t_{N-2}\sin t_{N-1} \end{cases}$$

Indeed, these coordinates produce explicit charts for the sphere.

(2) A second idea, which makes use of less charts, and to be more precise, only 2 charts, is to use the stereographic projection, given by inverse maps as follows:

$$\Phi : \mathbb{R}^N \to S^N_{\mathbb{R}} - \{\infty\} \quad , \quad \Psi : S^N_{\mathbb{R}} - \{\infty\} \to \mathbb{R}^N$$

To be more precise, we have an isomorphism $\mathbb{R}^N \simeq S^N_\mathbb{R} - \{\infty\}$, obtained by identifying $\mathbb{R}^N = \mathbb{R}^N \times \{0\} \subset \mathbb{R}^{N+1}$ with the unit sphere $S^N_\mathbb{R} \subset \mathbb{R}^{N+1}$, with the convention that the point which is added is $\infty = (1, 0, \ldots, 0)$, and the correspondences are given by:

$$\Phi(v) = (1,0) + \frac{2}{1+||v||^2}(-1,v) \quad , \quad \Psi(c,x) = \frac{x}{1-c}$$

Indeed, in one sense, we must have a formula as follows for our map Φ, with the parameter $t \in (0,1)$ being such that $||\Phi(v)|| = 1$:

$$\Phi(v) = t(0,v) + (1-t)(1,0)$$

The equation for the parameter $t \in (0,1)$ can be solved as follows:

$$\begin{aligned}(1-t)^2 + t^2||v||^2 = 1 \quad &\iff \quad t^2(1+||v||^2) = 2t\\ &\iff \quad t = \frac{2}{1+||v||^2}\end{aligned}$$

Thus, we obtain Φ. In the other sense now we must have, for a certain $\alpha \in \mathbb{R}$:

$$(0, \Psi(c,x)) = \alpha(c,x) + (1-\alpha)(1,0)$$

But from $\alpha c + 1 - \alpha = 0$ we get the following formula for the parameter α:

$$\alpha = \frac{1}{1-c}$$

Thus, we are led to the above formula for the inverse map Ψ.

(3) Next, we have as well cylindrical coordinates for the sphere, as well as many other types of more specialized coordinates, which can be useful in physics, plus of course, in disciplines like geography, economics and so on. There are many interesting computations that can be done here, and we will be back to these, on a regular basis in what follows, once we will know more about smooth manifolds, and their properties. □

As another basic examples of smooth manifolds, we have the projective spaces:

THEOREM 5.19. *The projective space $P^{N-1}_\mathbb{R}$ is a smooth manifold, with charts*

$$(x_1, \ldots, x_N) \to \left(\frac{x_1}{x_i}, \ldots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \ldots, \frac{x_N}{x_i}\right)$$

where $x_i \neq 0$. This manifold is compact, and of dimension $N-1$.

PROOF. We know that $P^{N-1}_\mathbb{R}$ appears as the space of lines in $\mathbb{R}^N$ passing through the origin, so we have the following formula, with $\sim$ being the proportionality of vectors, given as usual by $x \sim y$ when $x = \lambda y$, for some scalar $\lambda \neq 0$:

$$P^{N-1}_\mathbb{R} = \mathbb{R}^N - \{0\} / \sim$$

Alternatively, we can restrict if we want the attention to the vectors on the unit sphere $S^{N-1}_\mathbb{R} \subset \mathbb{R}^N$, and this because any line in $\mathbb{R}^N$ passing through the origin will certainly cross this sphere. Moreover, it is clear that our line will cross the sphere in exactly two points $\pm x$, and we conclude that we have the following formula, with $\sim$ being now the proportionality of vectors on the sphere, given by $x \sim y$ when $x = \pm y$:

$$P^{N-1}_{\mathbb{R}} = S^{N-1}_{\mathbb{R}} / \sim$$

With this discussion made, let us get now to what is to be proved. Obviously, once we fix an index $i \in \{1, \ldots, N\}$, the condition $x_i \neq 0$ on the vectors $x \in \mathbb{R}^N - \{0\}$ defines an open subset $U_i \subset P^{N-1}_{\mathbb{R}}$, and the open subsets that we get in this way cover $P^{N-1}_{\mathbb{R}}$:

$$P^{N-1}_{\mathbb{R}} = U_1 \cup \ldots \cup U_N$$

Moreover, the map in the statement is injective $U_i \to \mathbb{R}^{N-1}$, and it is clear too that the changes of charts are C^∞. Thus, we have our smooth manifold, as claimed. □

5d. Tangent spaces

Tangent spaces.

5e. Exercises

This chapter was a standard introduction to geometry at large, both algebraic and differential, and as exercises on all this, we have:

Exercise 5.20. *Clarify all the details, in the basic properties of conics.*

Exercise 5.21. *Work out some more on gravity, based on the above.*

Exercise 5.22. *Find the formulae for all conics, in all types of coordinates.*

Exercise 5.23. *Read about the higher degree algebraic plane curves.*

Exercise 5.24. *Read also about the surfaces in $\mathbb{R}^3$, beyond the above.*

Exercise 5.25. *What is the best way to draw a map of the Earth?*

Exercise 5.26. *Learn more about projective spaces, and their properties.*

Exercise 5.27. *Compute tangent spaces, for all manifolds that we know.*

As bonus exercise, complementary to what we will be doing here, read some algebraic geometry. This is the true classical geometry, that must be known too.

CHAPTER 6

Embedded manifolds

6a. Implicit functions

Implicit functions.

6b. Embedded manifolds

Embedded manifolds.

6c. Lagrange multipliers

We discuss now some more specialized topics, in relation with optimization. First of all, thinking well, the functions that we have to minimize or maximize, in the real life, are often defined on a manifold, instead of being defined on the whole $\mathbb{R}^N$. Fortunately, the good old principle $f'(x)=0$ can be adapted to the manifold case, as follows:

PRINCIPLE 6.1. *In order for a function $f:X\to\mathbb{R}$ defined on a manifold X to have a local extremum at $x\in X$, we must have, as usual*

$$f'(x)=0$$

but with this taking into account the fact that the equations defining the manifold count as well as "zero", and so must be incorporated into the formula $f'(x)=0$.

Obviously, we are punching here about our weight, because our discussion about manifolds from chapter 5 was quite amateurish, and we have no tools in our bag for proving such things, or even for properly formulating them. However, we can certainly talk about all this, a bit like physicists do. So, our principle will be the one above, and in practice, the idea is that we must have a formula as follows, with g_i being the constraint functions for our manifold X, and with $\lambda_i\in\mathbb{R}$ being certain scalars, called Lagrange multipliers:

$$f'(x)=\sum_i\lambda_ig_i'(x)$$

As a basic illustration for this, our claim is that, by using a suitable manifold, and a suitable function, and Lagrange multipliers, we can prove in this way the Hölder inequality, that we know well of course, but without any computation. Let us start with:

PROPOSITION 6.2. *For any exponent $p > 1$, the following set*

$$S_p = \left\{ x \in \mathbb{R}^N \Big| \sum_i |x_i|^p = 1 \right\}$$

is a submanifold of $\mathbb{R}^N$.

PROOF. We know from chapter 5 that the unit sphere in $\mathbb{R}^N$ is a manifold. In our terms, this solves our problem at $p = 2$, because this unit sphere is:

$$S_2 = \left\{ x \in \mathbb{R}^N \Big| \sum_i x_i^2 = 1 \right\}$$

Now observe that we have a bijection $S_p \simeq S_2$, at least on the part where all the coordinates are positive, $x_i > 0$, given by the following function:

$$x_i \to x_i^{2/p}$$

Thus we obtain that S_p is indeed a manifold, as claimed. □

We already know that the manifold S_p constructed above is the unit sphere, in the case $p = 2$. In order to have a better geometric picture of what is going on, in general, observe that S_p can be constructed as well at $p = 1$, as follows:

$$S_1 = \left\{ x \in \mathbb{R}^N \Big| \sum_i |x_i| = 1 \right\}$$

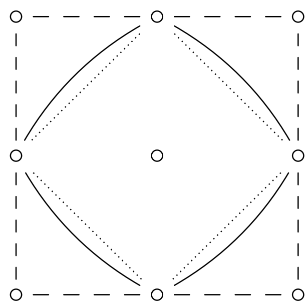
However, this is no longer a manifold, as we can see for instance at $N = 2$, where we obtain a square. Now observe that we can talk as well about $p = \infty$, as follows:

$$S_\infty = \left\{ x \in \mathbb{R}^N \Big| \sup_i |x_i| = 1 \right\}$$

This letter set is no longer a manifold either, as we can see for instance at $N = 2$, where we obtain again a square, containing the previous square, the one at $p = 1$.

With these limiting constructions in hand, we can have now a better geometric picture of what is going on, in the general context of Proposition 6.2. Indeed, let us draw, at

$N = 2$ for simplifying, our sets S_p at the values $p = 1, 2, \infty$ of the exponent:



We can see that what we have is a small square, at $p = 1$, becoming smooth and inflating towards the circle, in the parameter range $p \in (1, 2]$, and then further inflating, in the parameter range $p \in [2, \infty)$, towards the big square appearing at $p = \infty$.

With these preliminaries in hand, we can formulate our result, as follows:

THEOREM 6.3. *The local extrema over S_p of the function*

$$f(x) = \sum_i x_i y_i$$

can be computed by using Lagrange multipliers, and this gives

$$\left|\sum_i x_i y_i\right| \leq \left(\sum_i |x_i|^p\right)^{1/p} \left(\sum_i |y_i|^q\right)^{1/q}$$

with $1/p + 1/q = 1$, that is, the Hölder inequality, with a purely geometric proof.

PROOF. We can restrict the attention to the case where all the coordinates are positive, $x_i > 0$ and $y_i > 0$. The derivative of the function in the statement is:

$$f'(x) = (y_1, \ldots, y_N)$$

On the other hand, we know that the manifold S_p appears by definition as the set of zeroes of the function $\varphi(x) = \sum_i x_i^p - 1$, having derivative as follows:

$$\varphi'(x) = p(x_1^{p-1}, \ldots, x_N^{p-1})$$

Thus, by using Lagrange multipliers, the critical points of f must satisfy:

$$(y_1, \ldots, y_N) \sim (x_1^{p-1}, \ldots, x_N^{p-1})$$

In other words, the critical points must satisfy $x_i = \lambda y_i^{1/(p-1)}$, for some $\lambda > 0$, and by using now $\sum_i x_i^p = 1$ we can compute the precise value of λ, and we get:

$$\lambda = \left(\sum_i y_i^{p/(p-1)}\right)^{-1/p}$$

Now let us see what this means. Since the critical point is unique, this must be a maximum of our function, and we conclude that for any $x\in S_p$, we have:

$$\begin{aligned}\sum_i x_iy_i &\leq \sum_i\lambda y_i^{1/(p-1)}\cdot y_i\\ &= \left(\sum_i y_i^{p/(p-1)}\right)^{1-1/p}\\ &= \left(\sum_i y_i^q\right)^{1/q}\end{aligned}$$

Thus we have Hölder, and the general case follows from this, by rescaling. □

As a second illustration for the method of Lagrange multipliers, this time in relation with certain questions from linear algebra, let us discuss the Hadamard matrices. In the real case, the basic theory of these matrices is as follows:

Theorem 6.4. *The real Hadamard matrices, $H\in M_N(-1,1)$ having pairwise orthogonal rows, have the following properties:*

1. *The set of Hadamard matrices is $X_N=M_N(-1,1)\cap\sqrt{N}O_N$.*
2. *In order to have $X_N\neq\emptyset$, the matrix size must be $N\in\{2\}\cup 4\mathbb{N}$.*
3. *For $H\in M_N(-1,1)$ we have $|\det H|\leq N^{N/2}$, with equality when H is Hadamard.*
4. *For $U\in O_N$ we have $||U||_1\leq N\sqrt{N}$, with equality when $H=\sqrt{N}U$ is Hadamard.*

Proof. Many things going on here, the idea being as follows:

(1) This is just a reformulation of the Hadamard matrix condition.

(2) This follows by playing with the first 3 rows, exercise for you.

(3) This follows from our definition of the determinant, as a signed volume.

(4) This follows from $||U||_2=\sqrt{N}$ and Cauchy-Schwarz, easy exercise for you. □

All the above is quite interesting, and (1,2) raise the question of finding the correct generalizations of the Hadamard matrices, at $N\notin\{2\}\cup 4\mathbb{N}$. But the answer here comes from (3,4), which suggest looking either at the maximizers of $|\det|$ on $M_N(-1,1)$, or of $||.||_1$ on $\sqrt{N}O_N$. By following this latter way, we are led to the following question:

Question 6.5. *What are the critical points of the 1-norm on O_N?*

And, good news, we can solve this latter question by using the theory of Lagrange multipliers developed in the above, the result here being as follows:

THEOREM 6.6. *An orthogonal matrix with nonzero entries is a critical point of*

$$||U||_1 = \sum_{ij} |U_{ij}|$$

precisely when SU^t is symmetric, where $S_{ij} = \mathrm{sgn}(U_{ij})$.

PROOF. We regard O_N as a real algebraic manifold, with coordinates U_{ij}. This manifold consists by definition of the zeroes of the following polynomials:

$$A_{ij} = \sum_k U_{ik}U_{jk} - \delta_{ij}$$

Thus $U \in O_N$ is a critical point of $F(U) = ||U||_1$ when the following is satisfied:

$$dF \in span(dA_{ij})$$

Regarding the space $span(dA_{ij})$, this consists of the following quantities:

$$\begin{aligned}\sum_{ij} M_{ij} dA_{ij} &= \sum_{ijk} M_{ij}(U_{ik}dU_{jk} + U_{jk}dU_{ik})\\ &= \sum_{ij} (M^tU)_{ij} dU_{ij} + \sum_{ij} (MU)_{ij} dU_{ij}\end{aligned}$$

In order to compute dF, observe first that, with $S_{ij} = sgn(U_{ij})$, we have:

$$d|U_{ij}| = d\sqrt{U_{ij}^2} = \frac{U_{ij}dU_{ij}}{|U_{ij}|} = S_{ij}dU_{ij}$$

Thus $dF = \sum_{ij} S_{ij}dU_{ij}$, and so $U \in O_N$ is a critical point of F precisely when there exists a matrix $M \in M_N(\mathbb{R})$ such that the following two conditions are satisfied:

$$S = M^tU \quad , \quad S = MU$$

Now observe that these two equations can be written as follows:

$$M^t = SU^t \quad , \quad M = SU^t$$

Thus, the matrix SU^t must be symmetric, as claimed. □

6d. Gradient method

Gradient method.

6e. Exercises

Exercises:

Exercise 6.7.

Exercise 6.8.

Exercise 6.9.

Exercise 6.10.

Exercise 6.11.

Exercise 6.12.

Exercise 6.13.

Exercise 6.14.

Bonus exercise.

CHAPTER 7

Symmetric spaces

7a. Lie groups

Our aim here is to discuss some key applications of differential geometry techniques, to the study of the continuous groups of matrices. Before that, however, let us do some ad-hoc computations, in low dimensions, which are both instructive, and very useful. We first have the following result, regarding the main group here, which is SU_2:

THEOREM 7.1. *We have the formula*

$$SU_2=\left\{\begin{pmatrix}x+iy&z+it\\-z+it&x-iy\end{pmatrix}\ \Big|\ x^2+y^2+z^2+t^2=1\right\}$$

which makes SU_2 isomorphic to the unit real sphere $S^3_\mathbb{R}\subset\mathbb{R}^3$.

PROOF. By solving $U^*=U^{-1}$, under the assumption $\det U=1$, we obtain:

$$SU_2=\left\{\begin{pmatrix}a&b\\-\bar{b}&\bar{a}\end{pmatrix}\ \Big|\ |a|^2+|b|^2=1\right\}$$

Now let us write our parameters $a,b\in\mathbb{C}$, which belong to the complex unit sphere $S^1_\mathbb{C}\subset\mathbb{C}^2$, in terms of their real and imaginary parts, as follows:

$$a=x+iy\quad,\quad b=z+it$$

In terms of $x,y,z,t\in\mathbb{R}$, our formula for a generic matrix $U\in SU_2$ becomes the one in the statement. As for the condition to be satisfied by the parameters $x,y,z,t\in\mathbb{R}$, this comes the condition $|a|^2+|b|^2=1$ to be satisfied by $a,b\in\mathbb{C}$, which reads:

$$x^2+y^2+z^2+t^2=1$$

Thus, we are led to the conclusion in the statement. Regarding now the last assertion, recall that the unit sphere $S^3_\mathbb{R}\subset\mathbb{R}^4$ is given by:

$$S^3_\mathbb{R}=\left\{(x,y,z,t)\ \Big|\ x^2+y^2+z^2+t^2=1\right\}$$

Thus, we have an isomorphism of compact spaces, as follows:

$$SU_2\simeq S^3_\mathbb{R}\quad,\quad\begin{pmatrix}x+iy&z+it\\-z+it&x-iy\end{pmatrix}\to(x,y,z,t)$$

We have therefore proved our theorem. □

As a continuation of this material, we have the following result:

THEOREM 7.2. *We have the following formula,*

$$SU_2 = \left\{ xc_1 + yc_2 + zc_3 + tc_4 \,\Big|\, x^2 + y^2 + z^2 + t^2 = 1 \right\}$$

where c_1, c_2, c_3, c_4 are the Pauli matrices, given by:

$$c_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad c_2 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

$$c_3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad , \quad c_4 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

PROOF. We recall from Theorem 7.1 that the group SU_2 can be parametrized by the real sphere $S^3_{\mathbb{R}} \subset \mathbb{R}^4$, in the following way:

$$SU_2 = \left\{ \begin{pmatrix} x+iy & z+it \\ -z+it & x-iy \end{pmatrix} \,\Big|\, x^2 + y^2 + z^2 + t^2 = 1 \right\}$$

Thus, the elements $U \in SU_2$ are precisely the matrices as follows, depending on parameters $x, y, z, t \in \mathbb{R}$ satisfying $x^2 + y^2 + z^2 + t^2 = 1$:

$$U = x \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + y \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} + z \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + t \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

But this gives the formula for SU_2 in the statement. □

The above result is often the most convenient one, when dealing with SU_2. This is because the Pauli matrices have a number of remarkable properties, which are very useful when doing computations. These properties can be summarized as follows:

PROPOSITION 7.3. *The Pauli matrices multiply according to the formulae*

$$c_2^2 = c_3^2 = c_4^2 = -1$$

$$c_2c_3 = -c_3c_2 = c_4$$

$$c_3c_4 = -c_4c_3 = c_2$$

$$c_4c_2 = -c_2c_4 = c_3$$

they conjugate according to the following rules,

$$c_1^* = c_1 \ , \ c_2^* = -c_2 \ , \ c_3^* = -c_3 \ , \ c_4^* = -c_4$$

and they form an orthonormal basis of $M_2(\mathbb{C})$, with respect to the scalar product

$$< a, b >= tr(ab^*)$$

with $tr : M_2(\mathbb{C}) \to \mathbb{C}$ being the normalized trace of 2×2 matrices, $tr = Tr/2$.

PROOF. The first two assertions, regarding the multiplication and conjugation rules for the Pauli matrices, follow from some elementary computations. As for the last assertion, this follows by using these rules. Indeed, the fact that the Pauli matrices are pairwise orthogonal follows from computations of the following type, for $i \neq j$:

$$< c_i, c_j >= tr(c_i c_j^*) = tr(\pm c_i c_j) = tr(\pm c_k) = 0$$

As for the fact that the Pauli matrices have norm 1, this follows from:

$$< c_i, c_i >= tr(c_i c_i^*) = tr(\pm c_i^2) = tr(c_1) = 1$$

Thus, we are led to the conclusion in the statement. □

Regarding now the basic unitary groups in 3 or more dimensions, the situation here becomes fairly complicated. It is possible however to explicitly compute the rotation groups SO_3 and O_3, and explaining this result, due to Euler-Rodrigues, which is something non-trivial and very useful, for all sorts of practical purposes, will be our next goal.

The proof of the Euler-Rodrigues formula is something quite tricky. Let us start with the following construction, whose usefulness will become clear in a moment:

PROPOSITION 7.4. *The adjoint action $SU_2 \curvearrowright M_2(\mathbb{C})$, given by*

$$T_U(M) = UMU^*$$

leaves invariant the following real vector subspace of $M_2(\mathbb{C})$,

$$E = span_{\mathbb{R}}(c_1, c_2, c_3, c_4)$$

and we obtain in this way a group morphism $SU_2 \to GL_4(\mathbb{R})$.

PROOF. We have two assertions to be proved, as follows:

(1) We must first prove that, with $E \subset M_2(\mathbb{C})$ being the real vector space in the statement, we have the following implication:

$$U \in SU_2, M \in E \implies UMU^* \in E$$

But this is clear from the multiplication rules for the Pauli matrices, from Proposition 7.3. Indeed, let us write our matrices U, M as follows:

$$U = xc_1 + yc_2 + zc_3 + tc_4$$

$$M = ac_1 + bc_2 + cc_3 + dc_4$$

We know that the coefficients x, y, z, t and a, b, c, d are real, due to $U \in SU_2$ and $M \in E$. The point now is that when computing UMU^*, by using the various rules from Proposition 7.3, we obtain a matrix of the same type, namely a combination of c_1, c_2, c_3, c_4, with real coefficients. Thus, we have $UMU^* \in E$, as desired.

(2) In order to conclude, let us identify $E \simeq \mathbb{R}^4$, by using the basis c_1, c_2, c_3, c_4. The result found in (1) shows that we have a correspondence as follows:

$$SU_2 \to M_4(\mathbb{R}) \quad , \quad U \to (T_U)_{|E}$$

Now observe that for any $U \in SU_2$ and any $M \in M_2(\mathbb{C})$ we have:

$$T_{U^*}T_U(M) = U^*UMU^*U = M$$

Thus $T_{U^*} = T_U^{-1}$, and so the correspondence that we found can be written as:

$$SU_2 \to GL_4(\mathbb{R}) \quad , \quad U \to (T_U)_{|E}$$

But this a group morphism, due to the following computation:

$$T_UT_V(M) = UVMV^*U^* = T_{UV}(M)$$

Thus, we are led to the conclusion in the statement. □

The point now, which makes the link with SO_3, and which will ultimate elucidate the structure of SO_3, is that Proposition 7.4 can be improved as follows:

THEOREM 7.5. *The adjoint action $SU_2 \curvearrowright M_2(\mathbb{C})$, given by*

$$T_U(M) = UMU^*$$

leaves invariant the following real vector subspace of $M_2(\mathbb{C})$,

$$F = span_{\mathbb{R}}(c_2, c_3, c_4)$$

and we obtain in this way a group morphism $SU_2 \to SO_3$.

PROOF. We can do this in several steps, as follows:

(1) Our first claim is that the group morphism $SU_2 \to GL_4(\mathbb{R})$ constructed in Proposition 7.4 is in fact a morphism $SU_2 \to O_4$. In order to prove this, recall the following formula, valid for any $U \in SU_2$, from the proof of Proposition 7.4:

$$T_{U^*} = T_U^{-1}$$

We want to prove that the matrices $T_U \in GL_4(\mathbb{R})$ are orthogonal, and in view of the above formula, it is enough to prove that we have:

$$T_U^* = (T_U)^t$$

So, let us prove this. For any two matrices $M, N \in E$, we have:

$$\begin{aligned} < T_{U^*}(M), N > &= < U^*MU, N > \\ &= tr(U^*MUN) \\ &= tr(MUNU^*) \end{aligned}$$

On the other hand, we have as well the following formula:

$$\begin{aligned} <(T_U)^t(M), N> &= <M, T_U(N)> \\ &= <M, UNU^*> \\ &= tr(MUNU^*) \end{aligned}$$

Thus we have indeed $T_U^* = (T_U)^t$, which proves our $SU_2 \to O_4$ claim.

(2) In order now to finish, recall that we have by definition $c_1 = 1$, as a matrix. Thus, the action of SU_2 on the vector $c_1 \in E$ is given by:

$$T_U(c_1) = Uc_1U^* = UU^* = 1 = c_1$$

We conclude that $c_1 \in E$ is invariant under SU_2, and by orthogonality the following subspace of E must be invariant as well under the action of SU_2:

$$e_1^\perp = span_{\mathbb{R}}(c_2, c_3, c_4)$$

Now if we call this subspace F, and we identify $F \simeq \mathbb{R}^3$ by using the basis c_2, c_3, c_4, we obtain by restriction to F a morphism of groups as follows:

$$SU_2 \to O_3$$

But since this morphism is continuous and SU_2 is connected, its image must be connected too. Now since the target group decomposes as $O_3 = SO_3 \sqcup (-SO_3)$, and $1 \in SU_2$ gets mapped to $1 \in SO_3$, the whole image must lie inside SO_3, and we are done. □

The above result is quite interesting, because we will see in a moment that the morphism $SU_2 \to SO_3$ constructed there is surjective. Thus, we will have a way of parametrizing the elements $V \in SO_3$ by elements $U \in SO_2$, and so ultimately by parameters $(x, y, z, t) \in S^3_{\mathbb{R}}$. In order to work out all this, let us start with the following result, coming as a continuation of Proposition 7.4, independently of Theorem 7.5:

Proposition 7.6. *With respect to the standard basis c_1, c_2, c_3, c_4 of the vector space $\mathbb{R}^4 = span(c_1, c_2, c_3, c_4)$, the morphism $T : SU_2 \to GL_4(\mathbb{R})$ is given by:*

$$T_U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & x^2+y^2-z^2-t^2 & 2(yz-xt) & 2(xz+yt) \\ 0 & 2(xt+yz) & x^2+z^2-y^2-t^2 & 2(zt-xy) \\ 0 & 2(yt-xz) & 2(xy+zt) & x^2+t^2-y^2-z^2 \end{pmatrix}$$

Thus, when looking at T as a group morphism $SU_2 \to O_4$, what we have in fact is a group morphism $SU_2 \to O_3$, and even $SU_2 \to SO_3$.

Proof. With notations from Proposition 7.4 and its proof, let us first look at the action $L : SU_2 \curvearrowright \mathbb{R}^4$ by left multiplication, which is by definition given by:

$$L_U(M) = UM$$

In order to compute the matrix of this action, let us write, as usual:

$$U = xc_1 + yc_2 + zc_3 + tc_4$$

$$M = ac_1 + bc_2 + cc_3 + dc_4$$

By using the multiplication formulae in Proposition 7.3, we obtain:

$$\begin{aligned} UM &= (xc_1 + yc_2 + zc_3 + tc_4)(ac_1 + bc_2 + cc_3 + dc_4) \\ &= (xa - yb - zc - td)c_1 \\ &+ (xb + ya + zd - tc)c_2 \\ &+ (xc - yd + za + tb)c_3 \\ &+ (xd + yc - zb + ta)c_4 \end{aligned}$$

We conclude that the matrix of the left action considered above is:

$$L_U = \begin{pmatrix} x & -y & -z & -t \\ y & x & -t & z \\ z & t & x & -y \\ t & -z & y & x \end{pmatrix}$$

Similarly, let us look now at the action $R : SU_2 \curvearrowright \mathbb{R}^4$ by right multiplication, which is by definition given by the following formula:

$$R_U(M) = MU^*$$

In order to compute the matrix of this action, let us write, as before:

$$U = xc_1 + yc_2 + zc_3 + tc_4$$

$$M = ac_1 + bc_2 + cc_3 + dc_4$$

By using the multiplication formulae in Proposition 7.4, we obtain:

$$\begin{aligned} MU^* &= (ac_1 + bc_2 + cc_3 + dc_4)(xc_1 - yc_2 - zc_3 - tc_4) \\ &= (ax + by + cz + dt)c_1 \\ &+ (-ay + bx - ct + dz)c_2 \\ &+ (-az + bt + cx - dy)c_3 \\ &+ (-at - bz + cy + dx)c_4 \end{aligned}$$

We conclude that the matrix of the right action considered above is:

$$R_U = \begin{pmatrix} x & y & z & t \\ -y & x & -t & z \\ -z & t & x & -y \\ -t & -z & y & x \end{pmatrix}$$

Now by composing, the matrix of the adjoint matrix in the statement is:

$$\begin{aligned}
T_U &= R_UL_U\\
&= \begin{pmatrix}x&y&z&t\\-y&x&-t&z\\-z&t&x&-y\\-t&-z&y&x\end{pmatrix}\begin{pmatrix}x&-y&-z&-t\\y&x&-t&z\\z&t&x&-y\\t&-z&y&x\end{pmatrix}\\
&= \begin{pmatrix}1&0&0&0\\0&x^2+y^2-z^2-t^2&2(yz-xt)&2(xz+yt)\\0&2(xt+yz)&x^2+z^2-y^2-t^2&2(zt-xy)\\0&2(yt-xz)&2(xy+zt)&x^2+t^2-y^2-z^2\end{pmatrix}
\end{aligned}$$

Thus, we have indeed the formula in the statement. As for the remaining assertions, these are all clear either from this formula, or from Theorem 7.5. □

We can now formulate the Euler-Rodrigues result, as follows:

THEOREM 7.7. *We have a double cover map, obtained via the adjoint representation,*

$$SU_2\to SO_3$$

and this map produces the Euler-Rodrigues formula

$$U=\begin{pmatrix}x^2+y^2-z^2-t^2&2(yz-xt)&2(xz+yt)\\2(xt+yz)&x^2+z^2-y^2-t^2&2(zt-xy)\\2(yt-xz)&2(xy+zt)&x^2+t^2-y^2-z^2\end{pmatrix}$$

for the generic elements of SO_3.

PROOF. We know from the above that we have a group morphism $SU_2\to SO_3$, given by the formula in the statement, and the problem now is that of proving that this is a double cover map, in the sense that it is surjective, and with kernel $\{\pm1\}$.

(1) Regarding the kernel, this is elementary to compute, as follows:

$$\begin{aligned}
\ker(SU_2\to SO_3) &= \left\{U\in SU_2\Big|T_U(M)=M,\forall M\in E\right\}\\
&= \left\{U\in SU_2\Big|UM=MU,\forall M\in E\right\}\\
&= \left\{U\in SU_2\Big|Uc_i=c_iU,\forall i\right\}\\
&= \{\pm1\}
\end{aligned}$$

(2) Thus, we are done with this, and as a side remark here, this result shows that our morphism $SU_2\to SO_3$ is ultimately a morphism as follows:

$$PU_2\subset SO_3\quad,\quad PU_2=SU_2/\{\pm1\}$$

Here P stands for "projective", and it is possible to say more about the construction $G \to PG$, which can be performed for any subgroup $G \subset U_N$. But we will not get here into this, our next goal being that of proving that we have $PU_2 = SO_3$.

(3) We must prove now that the morphism $SU_2 \to SO_3$ is surjective. This is something non-trivial, and there are several advanced proofs for this, as follows:

– A first proof is by using Lie theory. To be more precise, the tangent spaces at 1 of both SU_2 and SO_3 can be explicitly computed, by doing some linear algebra, and the morphism $SU_2 \to SO_3$ follows to be surjective around 1, and then globally.

– Another proof is via representation theory. Indeed, the representations of SU_2 and SO_3 are subject to very similar formulae, called Clebsch-Gordan rules, and this shows that $SU_2 \to SO_3$ is surjective. We will discuss later, too.

– Yet another advanced proof, which is actually quite bordeline for what can be called "proof", is by using the ADE/McKay classification of the subgroups $G \subset SO_3$, which shows that there is no room strictly inside SO_3 for something as big as PU_2.

(4) In short, with some good knowledge of group theory, we are done. However, this is not our case, and we will present in what follows a more pedestrian proof, which was actually the original proof, based on the fact that any rotation $U \in SO_3$ has an axis.

(5) As a first computation, let us prove that any rotation $U \in Im(SU_2 \to SO_3)$ has an axis. We must look for fixed points of such rotations, and by linearity it is enough to look for fixed points belonging to the sphere $S^2_\mathbb{R} \subset \mathbb{R}^3$. Now recall that in our picture for the quotient map $SU_2 \to SO_3$, the space $\mathbb{R}^3$ appears as $F = span_\mathbb{R}(c_2, c_3, c_4)$, naturally embedded into the space $\mathbb{R}^4$ appearing as $E = span_\mathbb{R}(c_1, c_2, c_3, c_4)$. Thus, we must look for fixed points belonging to the sphere $S^3_\mathbb{R} \subset \mathbb{R}^4$ whose first coordinate vanishes. But, in our $\mathbb{R}^4 = E$ picture, this sphere $S^3_\mathbb{R}$ is the group SU_2. Thus, we must look for fixed points $V \in SU_2$ whose first coordinate with respect to c_1, c_2, c_3, c_4 vanishes, which amounts in saying that the diagonal entries of V must be purely imaginary numbers.

(6) Long story short, via our various identifications, we are led into solving the equation $UV = VU$ with $U, V \in SU_2$, and with V having a purely imaginary diagonal. So, with standard notations for SU_2, we must solve the following equation, with $p \in i\mathbb{R}$:

$$\begin{pmatrix}a&b\\-\bar{b}&\bar{a}\end{pmatrix}\begin{pmatrix}p&q\\-\bar{q}&\bar{p}\end{pmatrix}=\begin{pmatrix}p&q\\-\bar{q}&\bar{p}\end{pmatrix}\begin{pmatrix}a&b\\-\bar{b}&\bar{a}\end{pmatrix}$$

(7) But this is something which is routine. Indeed, by identifying coefficients we obtain the following equations, each appearing twice:

$$b\bar{q} = \bar{b}q \quad , \quad b(p - \bar{p}) = (a - \bar{a})q$$

In the case $b=0$ the only equation which is left is $q=0$, and reminding that we must have $p\in i\mathbb{R}$, we do have solutions, namely two of them, as follows:

$$V=\pm\begin{pmatrix}i&0\\0&i\end{pmatrix}$$

(8) In the remaining case $b\neq 0$, the first equation reads $b\bar{q}\in\mathbb{R}$, so we must have $q=\lambda b$ with $\lambda\in\mathbb{R}$. Now with this substitution made, the second equation reads $p-\bar{p}=\lambda(a-\bar{a})$, and since we must have $p\in i\mathbb{R}$, this gives $2p=\lambda(a-\bar{a})$. Thus, our equations are:

$$q=\lambda b\quad,\quad p=\lambda\cdot\frac{a-\bar{a}}{2}$$

Getting back now to our problem about finding fixed points, assuming $|a|^2+|b|^2=1$ we must find $\lambda\in\mathbb{R}$ such that the above numbers p,q satisfy $|p|^2+|q|^2=1$. But:

$$\begin{aligned}|p|^2+|q|^2&=|\lambda b|^2+\left|\lambda\cdot\frac{a-\bar{a}}{2}\right|^2\\&=\lambda^2(|b|^2+Im(a)^2)\\&=\lambda^2(1-Re(a)^2)\end{aligned}$$

Thus, we have again two solutions to our fixed point problem, given by:

$$\lambda=\pm\frac{1}{\sqrt{1-Re(a)^2}}$$

(9) Summarizing, we have proved that any rotation $U\in Im(SU_2\to SO_3)$ has an axis, and with the direction of this axis, corresponding to a pair of opposite points on the sphere $S^2_\mathbb{R}\subset\mathbb{R}^3$, being given by the above formulae, via $S^2_\mathbb{R}\subset S^3_\mathbb{R}=SU_2$.

(10) In order to finish, we must argue that any rotation $U\in SO_3$ has an axis. But this follows for instance from some topology, by using the induced map $S^2_\mathbb{R}\to S^2_\mathbb{R}$. Now since $U\in SO_3$ is uniquely determined by its rotation axis, which can be regarded as a point of $S^2_\mathbb{R}/\{\pm1\}$, plus its rotation angle $t\in[0,2\pi)$, by using $S^2_\mathbb{R}\subset S^3_\mathbb{R}=SU_2$ as in (9) we are led to the conclusion that U is uniquely determined by an element of $SU_2/\{\pm1\}$, and so appears indeed via the Euler-Rodrigues formula, as desired. □

Regarding now O_3, the extension from SO_3 is very simple, as follows:

THEOREM 7.8. *We have the Euler-Rodrigues formula*

$$U=\pm\begin{pmatrix}x^2+y^2-z^2-t^2&2(yz-xt)&2(xz+yt)\\2(xt+yz)&x^2+z^2-y^2-t^2&2(zt-xy)\\2(yt-xz)&2(xy+zt)&x^2+t^2-y^2-z^2\end{pmatrix}$$

for the generic elements of O_3.

Proof. This follows from Theorem 7.7, because the determinant of an orthogonal matrix $U \in O_3$ must satisfy $\det U = \pm 1$, and in the case $\det U = -1$, we have:

$$\det(-U) = (-1)^3 \det U = -\det U = 1$$

Thus, assuming $\det U = -1$, we can therefore rescale U into an element $-U \in SO_3$, and this leads to the conclusion in the statement. □

With the above small N examples worked out, let us discuss now the general theory, at arbitrary values of $N \in \mathbb{N}$. In the real case, we have the following result:

Proposition 7.9. *We have a decomposition as follows, with SO_N^{-1} consisting by definition of the orthogonal matrices having determinant -1:*

$$O_N = SO_N \cup SO_N^{-1}$$

Moreover, when N is odd the set SO_N^{-1} is simply given by $SO_N^{-1} = -SO_N$.

Proof. The first assertion is clear from definitions, because the determinant of an orthogonal matrix must be ± 1. The second assertion is clear too, and we have seen this already at $N = 3$, in the proof of Theorem 7.8. Finally, when N is even the situation is more complicated, and requires complex numbers. We will be back to this. □

In the complex case now, the result is simpler, as follows:

Proposition 7.10. *We have a decomposition as follows, with SU_N^d consisting by definition of the unitary matrices having determinant $d \in \mathbb{T}$:*

$$O_N = \bigcup_{d \in \mathbb{T}} SU_N^d$$

Moreover, the components are $SU_N^d = f \cdot SU_N$, where $f \in \mathbb{T}$ is such that $f^N = d$.

Proof. This is clear from definitions, and from the fact that the determinant of a unitary matrix belongs to $\mathbb{T}$, by extracting a suitable square root of the determinant. □

It is possible to use the decomposition in Proposition 7.10 in order to say more about what happens in the real case, in the context of Proposition 7.9, but we will not get into this. We will basically stop here with our study of O_N, U_N, and of their versions SO_N, SU_N. As a last result on the subject, however, let us record:

Theorem 7.11. *We have subgroups of O_N, U_N constructed via the condition*

$$(\det U)^d = 1$$

with $d \in \mathbb{N} \cup \{\infty\}$, which generalize both O_N, U_N and SO_N, SU_N.

Proof. This is indeed from definitions, and from the multiplicativity property of the determinant. We will be back to these groups, which are quite specialized, later on. □

7b. Lie algebras

In relation with differential geometry, let us start with the following definition:

Definition 7.12. *A Lie algebra is a vector space $\mathfrak{g}$ with an operation $(x,y)\to[x,y]$, called Lie bracket, subject to the following conditions:*

(1) $[x+y,z]=[x,z]+[y,z]$, $[x,y+z]=[x,y]+[x,z]$.
(2) $[\lambda x,y]=[x,\lambda y]=\lambda[x,y]$.
(3) $[x,x]=0$.
(4) $[[x,y],z]+[[y,z],x]+[[z,x],y]=0$.

As a basic example, consider a usual, associative algebra A. We can define then the Lie bracket on it as being the usual commutator, namely:

$$[x,y]=xy-yx$$

The above axioms (1,2,3) are then clearly satisfied, and in what regards axiom (4), called Jacobi identity, this is satisfied too, the verification being as follows:

$$\begin{aligned}
&[[x,y],z]+[[y,z],x]+[[z,x],y]\\
=\ &[xy-yx,z]+[yz-zy,x]+[zx-xz,y]\\
=\ &xyz-yxz-zxy+zyx+yzx-zyx-xyz+xzy+zxy-xzy-yzx+yxz\\
=\ &0
\end{aligned}$$

We will show in a moment that up to a certain abstract operation $\mathfrak{g}\to U\mathfrak{g}$, called enveloping Lie algebra construction, any Lie algebra appears in this way, with its Lie bracket being formally given by $[x,y]=xy-yx$. Before that, however, you might wonder where that Gothic letter $\mathfrak{g}$ comes from. That comes from the following fundamental result, making the connection with the theory of groups, denoted as usual by G:

Theorem 7.13. *Given a Lie group G, that is, a group which is a smooth manifold, with the group operations being smooth, the tangent space at the identity*

$$\mathfrak{g}=T_1(G)$$

is a Lie algebra, with its Lie bracket being basically a usual commutator.

Proof. This is something non-trivial, the idea being as follows:

(1) Let us first have a look at the orthogonal and unitary groups O_N,N_N. These are both Lie groups, and the corresponding Lie algebras $\mathfrak{o}_N,\mathfrak{u}_N$ can be computed by differentiating the equations defining O_N,U_N, with the conclusion being as follows:

$$\mathfrak{o}_N=\left\{A\in M_N(\mathbb{R})\Big|A^t=-A\right\}$$

$$\mathfrak{u}_N=\left\{B\in M_N(\mathbb{C})\Big|B^*=-B\right\}$$

This was for the correspondences $O_N \to \mathfrak{o}_N$ and $U_N \to \mathfrak{u}_N$. In the other sense, the correspondences $\mathfrak{o}_N \to O_N$ and $\mathfrak{u}_N \to U_N$ appear by exponentiation, the result here stating that, around 1, the orthogonal matrices can be written as $U = e^A$, with $A \in \mathfrak{o}_N$, and the unitary matrices can be written as $U = e^B$, with $B \in \mathfrak{u}_N$.

(2) Getting now to the Lie bracket, the first observation is that both $\mathfrak{o}_N, \mathfrak{u}_N$ are stable under the usual commutator of the $N \times N$ matrices. Indeed, assuming that $A, B \in M_N(\mathbb{R})$ satisfy $A^t = -A$, $B^t = -B$, their commutator satisfies $[A, B] \in M_N(\mathbb{R})$, and:

$$\begin{aligned} [A,B]^t &= (AB - BA)^t \\ &= B^tA^t - A^tB^t \\ &= BA - AB \\ &= -[A,B] \end{aligned}$$

Similarly, assuming that $A, B \in M_N(\mathbb{C})$ satisfy $A^* = -A$, $B^* = -B$, their commutator $[A, B] \in M_N(\mathbb{C})$ satisfies the condition $[A, B]^* = -[A, B]$.

(3) Thus, both tangent spaces $\mathfrak{o}_N, \mathfrak{u}_N$ are Lie algebras, with the Lie bracket being the usual commutator of the $N \times N$ matrices. It remains now to see what happens to the Lie bracket when exponentiating, and the formula here is as follows:

$$e^{[A,B]} = e^{AB-BA}$$

But the term on the right can be understood in terms of the differential geometry of O_N, U_N, and the situation is similar when dealing with an arbitrary Lie group G. □

With this understood, let us go back to the arbitrary Lie algebras, as axiomatized in Definition 7.12. There is an obvious analogy there with the axioms for the usual, associative algebras, and based on this analogy, we can build some abstract algebra theory for the Lie algebras. Let us record some basic results, along these lines:

PROPOSITION 7.14. *Let $\mathfrak{g}$ be a Lie algebra. If we define its ideals as being the vector spaces $\mathfrak{i} \subset \mathfrak{g}$ satisfying the condition*

$$x \in \mathfrak{i}, y \in \mathfrak{g} \implies [x, y] \in \mathfrak{i}$$

then the quotients $\mathfrak{g}/\mathfrak{i}$ are Lie algebras. Also, given a morphism of Lie algebras $f : \mathfrak{g} \to \mathfrak{h}$, its kernel $ker(f) \subset \mathfrak{g}$ is an ideal, and we have $\mathfrak{g}/\ker(f) = Im(f)$.

PROOF. All this is very standard, exactly as in the case of the associative algebras, and we will leave the various verifications here as an instructive exercise. □

Getting now to the point, remember our claim from the discussion after Definition 7.12, stating that up to a certain abstract operation $\mathfrak{g} \to U\mathfrak{g}$, called enveloping Lie algebra construction, any Lie algebra appears in fact from the "trivial" associative algebra construction, that is, with its Lie bracket being formally a usual commutator:

$$[x, y] = xy - yx$$

Time now to clarify this. The result here, making as well to the link with the various Lie group considerations from Theorem 7.13 and its proof, is as follows:

THEOREM 7.15. *Given a Lie algebra $\mathfrak{g}$, define its enveloping Lie algebra $U\mathfrak{g}$ as being the quotient of the tensor algebra of $\mathfrak{g}$, namely*

$$T(\mathfrak{g})=\bigoplus_{k=0}^\infty\mathfrak{g}^{\otimes k}$$

by the following associative algebra ideal, with x,y ranging over the elements of $\mathfrak{g}$:

$$I=< x\otimes y-y\otimes x-[x,y] >$$

Then $U\mathfrak{g}$ is an associative algebra, so it is a Lie algebra too, with bracket

$$[x,y]=xy-yx$$

and the standard embedding $\mathfrak{g}\subset U\mathfrak{g}$ is a Lie algebra embedding.

PROOF. This is something which is quite self-explanatory, and in what regards the examples, illustrations, and other things that can be said, for instance in relation with the Lie groups, we will leave some further reading here as an instructive exercise. □

Importantly, the above enveloping Lie algebra construction makes as well a link with Hopf algebra theory, and with quantum groups, via the following result:

THEOREM 7.16. *Given a Lie algebra $\mathfrak{g}$, its enveloping Lie algebra $U\mathfrak{g}$ is a cocommutative Hopf algebra, with comultiplication, counit and antipode given by*

$$\Delta:U\mathfrak{g}\to U(\mathfrak{g}\oplus\mathfrak{g})=U\mathfrak{g}\otimes U\mathfrak{g}\quad,\quad x\to x+x$$

$$\varepsilon:U\mathfrak{g}\to F\quad,\quad x\to 1$$

$$S:U\mathfrak{g}\to U\mathfrak{g}^{opp}=(U\mathfrak{g})^{opp}\quad,\quad x\to -x$$

via various standard identifications, for the various associative algebras involved.

PROOF. Again, this is something quite self-explanatory, and in what regards the examples, illustrations, and other things that can be said, for instance in relation with the Lie groups, we will leave some further reading here as an instructive exercise. □

7c. Symmetric spaces

Symmetric spaces.

7d. Haar integration

Let us begin with the compact group case, which is elementary. We first have:

DEFINITION 7.17. *Given a compact group G, and two of its representations,*

$$v : G \to U_N \quad , \quad w : G \to U_M$$

we define the space of intertwiners between these representations as being

$$Hom(v, w) = \left\{ T \in M_{M \times N}(\mathbb{C}) \Big| Tv_g = w_g T, \forall g \in G \right\}$$

and we use the following conventions:

(1) *We use the notations $Fix(v) = Hom(1, v)$, and $End(v) = Hom(v, v)$.*
(2) *We write $v \sim w$ when $Hom(v, w)$ contains an invertible element.*
(3) *We say that v is irreducible, and write $v \in Irr(G)$, when $End(v) = \mathbb{C}1$.*

The terminology here is standard, with Fix, Hom, End standing for fixed points, homomorphisms and endomorphisms. We will see later that irreducible means indecomposable, in a suitable sense. Here are now a few basic results, regarding these spaces:

PROPOSITION 7.18. *The spaces of intertwiners have the following properties, showing that these spaces form a tensor $*$-category:*

(1) $T \in Hom(v, w), S \in Hom(w, z) \implies ST \in Hom(v, z)$.
(2) $S \in Hom(v, w), T \in Hom(z, t) \implies S \otimes T \in Hom(v \otimes z, w \otimes t)$.
(3) $T \in Hom(v, w) \implies T^* \in Hom(w, v)$.

In particular, given $v : G \to U_N$, the linear space $End(v) \subset M_N(\mathbb{C})$ is a $*$-algebra.

PROOF. All the formulae in the statement are clear from definitions, via elementary computations. As for the last assertion, this is something coming from (1,3). □

We can now formulate our first Peter-Weyl type theorem, as follows:

THEOREM 7.19 (PW1). *Let $v : G \to U_N$ be a representation, consider its $*$-algebra of intertwiners $A = End(v)$, and write the unit of A as*

$$1 = p_1 + \ldots + p_k$$

with p_i being central minimal projections. We have then a decomposition

$$v = v_1 + \ldots + v_k$$

with each v_i being an irreducible representation, obtained by restricting v to $Im(p_i)$.

PROOF. This is something standard, the idea being as follows:

(1) We first associate to our representation $v : G \to U_N$ the corresponding action map on $\mathbb{C}^N$. If a linear subspace $W \subset \mathbb{C}^N$ is invariant, the restriction of the action map to W is an action map too, which must come from a subrepresentation $w \subset v$.

(2) Consider now a projection $p \in End(v)$. From $pv = vp$ we obtain that the linear space $W = Im(p)$ is invariant under v, and so this space must come from a subrepresentation $w \subset v$. It is routine to check that the operation $p \to w$ maps subprojections to subrepresentations, and minimal projections to irreducible representations.

(3) With these preliminaries in hand, let us decompose the algebra $End(v)$ as above, by using the decomposition $1 = p_1 + \ldots + p_k$ into central minimal projections. If we denote by $v_i \subset v$ the subrepresentation coming from the vector space $V_i = Im(p_i)$, then we obtain in this way a decomposition $v = v_1 + \ldots + v_k$, as in the statement. □

Here is now our second Peter-Weyl theorem, complementing Theorem 7.19:

THEOREM 7.20 (PW2). *Given a closed subgroup $G \subset_v U_N$, any of its irreducible smooth representations*

$$w : G \to U_M$$

appears inside a tensor product of the fundamental representation v and its adjoint $\bar{v}$.

PROOF. Given a representation $w : G \to U_M$, we define the space of coefficients $C_w \subset C(G)$ of this representation as being the following linear space:

$$C_w = span\Big[g \to w(g)_{ij}\Big]$$

With this notion in hand, the result can be deduced as follows:

(1) The construction $w \to C_w$ is functorial, in the sense that it maps subrepresentations into linear subspaces. This is indeed something which is routine to check.

(2) A closed subgroup $G \subset_v U_N$ is a Lie group, and a representation $w : G \to U_M$ is smooth when we have an inclusion $C_w \subset < C_v >$. This is indeed well-known.

(3) By definition of the Peter-Weyl representations, as arbitrary tensor products between the fundamental representation v and its conjugate $\bar{v}$, we have:

$$< C_v >= \sum_k C_{v^{\otimes k}}$$

(4) Now by putting together the above observations (2,3) we conclude that we must have an inclusion as follows, for certain exponents $k_1, \ldots, k_p$:

$$C_w \subset C_{v^{\otimes k_1} \oplus \ldots \oplus v^{\otimes k_p}}$$

(5) By using now (1), we deduce that we have an inclusion $w \subset v^{\otimes k_1} \oplus \ldots \oplus v^{\otimes k_p}$, and by applying Theorem 7.19, this leads to the conclusion in the statement. □

With this in hand, we can now talk about integration over G. It is convenient, for this purpose, to work with the integration functionals with respect to the various measures on G, instead of the measures themselves. We have the following key result:

THEOREM 7.21. *Any compact group G has a unique Haar integration, which can be constructed by starting with any faithful positive unital form $\psi\in C(G)^*$, and setting:*

$$\int_G=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n\psi^{*k}$$

Moreover, for any representation w we have the formula

$$\left(id\otimes\int_G\right)w=P$$

where P is the orthogonal projection onto $Fix(w)=\left\{\xi\in\mathbb C^n\Big|w\xi=\xi\right\}$.

PROOF. This is something very standard, the idea being as follows:

(1) Our first claim is that given a unital positive linear form $\psi:C(G)\to\mathbb C$, the limit in the statement exists, and the last formula holds too, with P being the orthogonal projection onto the 1-eigenspace of $M=(id\otimes\psi)w$. Indeed, we have:

$$(id\otimes\psi^{*k})w=M^k$$

It follows that our Cesàro limit is given by the following formula:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n\psi^{*k}(f)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n\tau(M^k)=\tau\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^nM^k\right)$$

Now since w is unitary we have $||w||=1$, and so $||M||\leq1$. Thus, the Cesàro limit on the right converges, and equals indeed the projection onto the 1-eigenspace of M:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^nM^k=P$$

(2) Our second claim is that when ψ is faithful, the limit in the statement is independent of ψ, and the last formula holds too. In order to prove this, we must show that the 1-eigenspace of $M=(id\otimes\psi)w$ equals the space $Fix(w)$, and with the inclusion $\supset$ being clear, we must establish the inclusion $\subset$, which amounts in proving that we have:

$$M\xi=\xi\implies w\xi=\xi$$

For this purpose, assume that we have $M\xi=\xi$, and consider the following function:

$$f=\sum_i\left(\sum_jw_{ij}\xi_j-\xi_i\right)\left(\sum_kw_{ik}\xi_k-\xi_i\right)^*$$

We must prove that we have $f = 0$. Since w is unitary, we have:

$$\begin{aligned} f &= \sum_i \left(\sum_j \left(w_{ij}\xi_j - \frac{1}{N}\xi_i \right) \right) \left(\sum_k \left(w_{ik}^*\bar{\xi}_k - \frac{1}{N}\bar{\xi}_i \right) \right) \\ &= \sum_{ijk} w_{ij}w_{ik}^*\xi_j\bar{\xi}_k - \frac{1}{N}w_{ij}\xi_j\bar{\xi}_i - \frac{1}{N}w_{ik}^*\xi_i\bar{\xi}_k + \frac{1}{N^2}\xi_i\bar{\xi}_i \\ &= \sum_j |\xi_j|^2 - \sum_{ij} w_{ij}\xi_j\bar{\xi}_i - \sum_{ik} w_{ik}^*\xi_i\bar{\xi}_k + \sum_i |\xi_i|^2 \\ &= ||\xi||^2 - < w\xi, \xi > - \overline{< w\xi, \xi >} + ||\xi||^2 \\ &= 2(||\xi||^2 - Re(< w\xi, \xi >)) \end{aligned}$$

By using now our assumption $M\xi = \xi$, we obtain from this that we have:

$$\begin{aligned} \psi(f) &= 2\psi(||\xi||^2 - Re(< w\xi, \xi >)) \\ &= 2(||\xi||^2 - Re(< M\xi, \xi >)) \\ &= 2(||\xi||^2 - ||\xi||^2) \\ &= 0 \end{aligned}$$

Now since ψ is faithful, this gives $f = 0$, and so $w\xi = \xi$, as claimed.

(3) But with the above results in hand, all the assertions in the statement follow. □

Summarizing, we can integrate over G. In fact, we can say even more, as follows:

THEOREM 7.22. *The Haar integration over a closed subgroup $G \subset U_N$ is given by*

$$\int_G g_{i_1 j_1}^{e_1} \ldots g_{i_k j_k}^{e_k} \, dg = \sum_{\pi,\nu \in D(k)} \delta_\pi(i)\delta_\nu(j)W_k(\pi,\nu)$$

for any colored integer $k = e_1 \ldots e_k$ and any multi-indices i, j, where $D(k)$ is a linear basis of $Fix(v^{\otimes k})$, the associated generalized Kronecker symbols are given by

$$\delta_\pi(i) = < \pi, e_{i_1} \otimes \ldots \otimes e_{i_k} >$$

and $W_k = G_k^{-1}$ is the inverse of the Gram matrix, $G_k(\pi,\nu) = < \pi, \nu >$.

PROOF. We know from Theorem 7.21 that the integrals in the statement form altogether the orthogonal projection P^k onto the following space:

$$Fix(v^{\otimes k}) = span(D(k))$$

Consider now the following linear map, with $D(k) = \{\xi_k\}$ being as above:

$$E(x) = \sum_{\pi \in D(k)} < x, \xi_\pi > \xi_\pi$$

By a standard linear algebra computation, it follows that we have $P = WE$, where W is the inverse of the restriction of E to the following space:

$$K = span\left(T_\pi \Big| \pi \in D(k)\right)$$

But this restriction is the linear map given by the Gram matrix G_k, and so W is the linear map given by the Weingarten matrix $W_k = G_k^{-1}$, and this gives the result. □

Let us develop now some further Peter-Weyl theory. We first have:

THEOREM 7.23 (PW3). *The dense subalgebra $\mathcal{C}(G) \subset C(G)$ generated by the coefficients of the fundamental representation decomposes as a direct sum*

$$\mathcal{C}(G) = \bigoplus_{w \in Irr(G)} M_{\dim(w)}(\mathbb{C})$$

with the summands being pairwise orthogonal with respect to the scalar product

$$< f, g >= \int_G f\bar{g}$$

where $\int_G$ is the Haar integration over G.

PROOF. By combining the previous two Peter-Weyl results, we deduce that we have a linear space decomposition as follows:

$$\mathcal{C}(G) = \sum_{w \in Irr(G)} C_w = \sum_{w \in Irr(G)} M_{\dim(w)}(\mathbb{C})$$

Thus, in order to conclude, it is enough to prove that for any two irreducible representations $v, w \in Irr(G)$, the corresponding spaces of coefficients are orthogonal:

$$v \not\sim w \implies C_v \perp C_w$$

But this follows by Frobenius duality, by integrating. Let us set indeed:

$$P_{ia,jb} = \int_G v_{ij}\bar{w}_{ab}$$

Then P is the orthogonal projection onto the following vector space:

$$Fix(v \otimes \bar{w}) \simeq Hom(v, w) = \{0\}$$

Thus we have $P = 0$, and this gives the result. □

Finally, we have the following result, completing the Peter-Weyl theory:

THEOREM 7.24 (PW4). *The characters of irreducible representations belong to*

$$\mathcal{C}(G)_{central} = \left\{ f \in \mathcal{C}(G) \Big| f(gh) = f(hg), \forall g, h \in G \right\}$$

called algebra of central functions on G, and form an orthonormal basis of it.

Proof. Observe first that $\mathcal{C}(G)_{central}$ is indeed an algebra, which contains all the characters. Conversely, consider a function $f \in \mathcal{C}(G)$, written as follows:

$$f = \sum_{w \in Irr(G)} f_w$$

The condition $f \in \mathcal{C}(G)_{central}$ states then that for any $w \in Irr(G)$, we must have:

$$f_w \in \mathcal{C}(G)_{central}$$

But this means that f_w must be a scalar multiple of χ_w, so the characters form a basis of $\mathcal{C}(G)_{central}$, as stated. Also, the fact that we have an orthogonal basis follows from Theorem 7.23. As for the fact that the characters have norm 1, this follows from:

$$\int_G \chi_w \bar{\chi}_w = \sum_{ij} \int_G w_{ii} \bar{w}_{jj} = \sum_i \frac{1}{M} = 1$$

Here we have used the fact, coming from Frobenius duality, that the various integrals $\int_G w_{ij} \bar{w}_{kl}$ form the orthogonal projection onto $Fix(w \otimes \bar{w}) \simeq End(w) = \mathbb{C}1$. □

Moving ahead, we will need as well Tannakian duality. Let us start with:

Definition 7.25. *The Tannakian category associated to a closed subgroup $G \subset_v U_N$ is the collection $C_G = (C_G(k,l))$ of vector spaces*

$$C_G(k,l) = Hom(v^{\otimes k}, v^{\otimes l})$$

where the representations $v^{\otimes k}$ with $k = \circ \bullet \bullet \circ \ldots$ colored integer, defined by

$$v^{\otimes \emptyset} = 1 \quad , \quad v^{\otimes \circ} = v \quad , \quad v^{\otimes \bullet} = \bar{v}$$

and multiplicativity, $v^{\otimes kl} = v^{\otimes k} \otimes v^{\otimes l}$, are the Peter-Weyl representations.

Let us make a summary of what we have so far, regarding these spaces $C_G(k,l)$. In order to formulate our result, let us start with the following definition:

Definition 7.26. *Let H be a finite dimensional Hilbert space. A tensor category over H is a collection $C = (C(k,l))$ of linear spaces*

$$C(k,l) \subset \mathcal{L}(H^{\otimes k}, H^{\otimes l})$$

satisfying the following conditions:

1. *$S, T \in C$ implies $S \otimes T \in C$.*
2. *If $S, T \in C$ are composable, then $ST \in C$.*
3. *$T \in C$ implies $T^* \in C$.*
4. *$C(k,k)$ contains the identity operator.*
5. *$C(\emptyset, k)$ with $k = \circ\bullet, \bullet\circ$ contain the operator $R : 1 \to \sum_i e_i \otimes e_i$.*
6. *$C(kl, lk)$ with $k, l = \circ, \bullet$ contain the flip operator $\Sigma : a \otimes b \to b \otimes a$.*

With these conventions, we have the following result:

THEOREM 7.27. *The construction $G \to C_G$ given by the formula*

$$C_G(k,l) = Hom(v^{\otimes k}, v^{\otimes l})$$

and the construction $C \to G_C$ given by the formula

$$G_C = \left\{ g \in U_N \Big| Tg^{\otimes k} = g^{\otimes l}T\ ,\ \forall k,l, \forall T \in C(k,l) \right\}$$

are bijective, and inverse to each other:

PROOF. This is something quite technical, obtained by doing some abstract algebra, and for details here, we refer to the Tannakian duality literature. □

In order to reach now to more concrete things, let us formulate:

DEFINITION 7.28. *Let $P(k,l)$ be the set of partitions between an upper colored integer k, and a lower colored integer l. A collection of subsets*

$$D = \bigsqcup_{k,l} D(k,l)$$

with $D(k,l) \subset P(k,l)$ is called a category of partitions when it has the following properties:

1. *Stability under the horizontal concatenation, $(\pi,\sigma) \to [\pi\sigma]$.*
2. *Stability under vertical concatenation $(\pi,\sigma) \to [^\sigma_\pi]$, with matching middle symbols.*
3. *Stability under the upside-down turning $*$, with switching of colors, $\circ \leftrightarrow \bullet$.*
4. *Each set $P(k,k)$ contains the identity partition $||\ldots||$.*
5. *The sets $P(\emptyset, \circ\bullet)$ and $P(\emptyset, \bullet\circ)$ both contain the semicircle $\cap$.*
6. *The sets $P(k,\bar{k})$ with $|k| = 2$ contain the crossing partition $\slash\!\!\!\backslash$.*

Let us formulate as well the following definition:

DEFINITION 7.29. *Given a partition $\pi \in P(k,l)$ and an integer $N \in \mathbb{N}$, we can construct a linear map between tensor powers of $\mathbb{C}^N$,*

$$T_\pi : (\mathbb{C}^N)^{\otimes k} \to (\mathbb{C}^N)^{\otimes l}$$

by the following formula, with $e_1, \ldots, e_N$ being the standard basis of $\mathbb{C}^N$,

$$T_\pi(e_{i_1} \otimes \ldots \otimes e_{i_k}) = \sum_{j_1 \ldots j_l} \delta_\pi \begin{pmatrix} i_1 & \ldots & i_k \\ j_1 & \ldots & j_l \end{pmatrix} e_{j_1} \otimes \ldots \otimes e_{j_l}$$

and with the coefficients on the right being Kronecker type symbols,

$$\delta_\pi \begin{pmatrix} i_1 & \ldots & i_k \\ j_1 & \ldots & j_l \end{pmatrix} \in \{0,1\}$$

whose values depend on whether the indices fit or not.

To be more precise, we put the indices of i,j on the legs of π, in the obvious way. In case all the blocks of π contain equal indices of i,j, we set $\delta_\pi\binom{i}{j} = 1$. Otherwise, we set $\delta_\pi\binom{i}{j} = 0$. The relation with the Tannakian categories comes from:

THEOREM 7.30. *Each category of partitions $D = (D(k,l))$ produces a family of compact groups $G = (G_N)$, with $G_N \subset_v U_N$, via the formula*

$$Hom(v^{\otimes k}, v^{\otimes l}) = span\left(T_\pi\Big|\pi \in D(k,l)\right)$$

and the Tannakian duality correspondence.

PROOF. Given an integer $N \in \mathbb{N}$, consider the correspondence $\pi \to T_\pi$ constructed in Definition 7.29, and then the collection of linear spaces in the statement, namely:

$$C(k,l) = span\left(T_\pi\Big|\pi \in D(k,l)\right)$$

According to Definition 7.28, this collection of spaces $C = (C(k,l))$ satisfies the axioms from Definition 7.26. Thus Tannakian duality applies, and gives the result. □

We call "easy" the closed subgroups $G \subset_v U_N$ appearing as above, from categories of partitions D. As a basic result here, due to Brauer, we have:

THEOREM 7.31. *We have the following results:*

(1) *The unitary group U_N is easy, coming from the category $\mathcal{P}_2$.*
(2) *The orthogonal group O_N is easy too, coming from the category P_2.*

PROOF. This is something very standard, the idea being as follows:

(1) The group U_N being defined via the relations $v^* = v^{-1}$, $v^t = \bar{v}^{-1}$, the associated Tannakian category is $C = span(T_\pi|\pi \in D)$, with:

$$D = < \overset{\cap}{\circ\bullet}\ ,\ \overset{\cap}{\bullet\circ} > = \mathcal{P}_2$$

(2) The group $O_N \subset U_N$ being defined by imposing the relations $v_{ij} = \bar{v}_{ij}$, the associated Tannakian category is $C = span(T_\pi|\pi \in D)$, with:

$$D = < \mathcal{P}_2, \overset{\circ}{\underset{\bullet}{|}}, \overset{\bullet}{\underset{\circ}{|}} > = P_2$$

Thus, we are led to the conclusion in the statement. □

There are many other easy groups, and as a basic example here, we have:

THEOREM 7.32. *The symmetric group S_N, regarded as group of unitary matrices,*

$$S_N \subset O_N \subset U_N$$

via the permutation matrices, is easy, coming from the category of all partitions P.

PROOF. Consider the easy group $G \subset O_N$ coming from the category of all partitions P. Since P is generated by the one-block partition $Y \in P(2,1)$, we have:

$$C(G) = C(O_N)\Big/\Big\langle T_Y \in Hom(v^{\otimes 2}, v)\Big\rangle$$

Since we have $T_Y(e_i \otimes e_j) = \delta_{ij} e_i$, the above relations read:

$$T_Y \in Hom(v^{\otimes 2}, v) \iff v_{ij} v_{ik} = \delta_{jk} v_{ij}, \forall i, j, k$$

In other words, the elements v_{ij} must be projections, and these projections must be pairwise orthogonal on the rows of $v = (v_{ij})$. We conclude that $G \subset O_N$ is the subgroup of matrices $g \in O_N$ having the property $g_{ij} \in \{0, 1\}$. Thus we have $G = S_N$, as claimed. □

Now back to the general easy group case, we have the following result:

THEOREM 7.33. *For an easy group $G \subset U_N$, coming from a category of partitions $D = (D(k,l))$, we have the Weingarten formula*

$$\int_G g_{i_1 j_1}^{e_1} \ldots g_{i_k j_k}^{e_k} \, dg = \sum_{\pi, \nu \in D(k)} \delta_\pi(i) \delta_\nu(j) W_{kN}(\pi, \nu)$$

for any $k = e_1 \ldots e_k$ and any i, j, where $D(k) = D(\emptyset, k)$, δ are usual Kronecker type symbols, checking whether the indices match, and $W_{kN} = G_{kN}^{-1}$, with

$$G_{kN}(\pi, \nu) = N^{|\pi \vee \nu|}$$

where $|.|$ is the number of blocks.

PROOF. This follows from the abstract Weingarten formula, from Theorem 7.22. Indeed, in the easy group case the Kronecker type symbols there are then the usual ones, and the Gram matrix being as well the correct one, we obtain the result. □

There are many applications of this formula. We will be back to this.

7e. Exercises

This was a quite standard chapter on classical Lie groups and symmetric spaces, and integration on them, and as exercises on all this, we have:

EXERCISE 7.34. *Clarify the details, for Lie algebras of O_N, U_N, Sp_N.*

EXERCISE 7.35. *Clarify as well the details, for the Lie algebras in general.*

EXERCISE 7.36. *Lean more about enveloping Lie algebras, and their properties.*

EXERCISE 7.37. *Learn as well about deformations of these enveloping Lie algebras.*

EXERCISE 7.38. *Clarify what we said above, in relation with symmetric spaces.*

EXERCISE 7.39. *Clarify all the details, in the proof of the Peter-Weyl theorem.*

EXERCISE 7.40. *Learn, with full details, the proof of Tannakian duality.*

EXERCISE 7.41. *Experiment a bit with Weingarten calculus, over various groups.*

As bonus exercise, have a look at advanced probability theory, such as random matrices. This is where the integration technology developed above really shines.

CHAPTER 8

Stokes, applications

8a. Differential forms

Differential forms.

8b. Contour integrals

We would like to discuss now the integration over curves, surfaces, and other smooth manifolds. Let us start with something very basic, regarding the curves, namely:

THEOREM 8.1. *The length of a curve* $\gamma:[a,b]\to\mathbb R^N$ *is given by*

$$L(\gamma)=\int_a^b||\gamma'(t)||dt$$

with $||.||$ *being the usual norm of* $\mathbb R^N$.

PROOF. This is something quite intuitive, that can even stand as a definition for the length of the curves, and that we will not prove in detail here, the idea being as follows:

(1) To start with, what is the length of a curve? Good question, and in answer, a physicist would say that this is the quantity obtained by integrating the magnitude of the velocity vector over the curve, with respect to time. But this velocity vector is $\gamma'(t)$, having magnitude $||\gamma'(t)||$, so we are led to the formula in the statement.

(2) Regarding now mathematicians, these would say that the length of a curve is the following quantity, with $(t_1=a,t_2,\ldots,t_{n-1},t_n=b)$ being a uniform division of (a,b):

$$L(\gamma)=\lim_{n\to\infty}\sum_{i=1}^n||\gamma(t_i)-\gamma(t_{i-1})||$$

But, by using the fundamental theorem of calculus, we can write this as follows:

$$L(\gamma)=\lim_{n\to\infty}\sum_{i=1}^n\left|\left|\int_{t_{i-1}}^{t_i}\gamma'(t)dt\right|\right|$$

And the point now is that, by doing some standard analysis, that we will leave here as an instructive exercise, we are led to the formula in the statement.

(3) So, we have our definition-theorem for the length of the curves, and for the discussion to be complete, we just need an illustration for this. But here, we have the following standard computation, for the length of the ellipses $(x/a)^2 + (y/b)^2 = 1$:

$$\begin{aligned} L &= 4\int_0^{\pi/2}\sqrt{\left(\frac{da\cos t}{dt}\right)^2+\left(\frac{db\sin t}{dt}\right)^2}\,dt \\ &= 4\int_0^{\pi/2}\sqrt{a^2\sin^2 t+b^2\cos^2 t}\,dt \end{aligned}$$

There are of course many other applications of the formula in the statement, and we will leave some thinking and computations here as an instructive exercise. □

Next, let us talk about the contour integrals over curves. We have here:

THEOREM 8.2. *Given a path $\gamma\subset\mathbb R^3$, we can talk about integrals of type*

$$I=\int_\gamma f(x)dx_1+g(x)dx_2+h(x)dx_3$$

with $f,g,h:\mathbb R^3\to\mathbb R$, which are independent on the chosen parametrization of the path.

PROOF. This is something quite straightforward, the idea being as follows:

(1) Regarding the statement itself, assume indeed that we have a path in $\mathbb R^3$, which can be best thought of as corresponding to a function as follows:

$$\gamma:[a,b]\to\mathbb R^3$$

Observe that this function γ is not exactly the path itself, for instance because the following functions produce the same path, parametrized differently:

$$\begin{aligned} \delta:[0,b-a]\to\mathbb R^3 &\quad,\quad \delta(t)=\gamma(t+a) \\ \varepsilon:[0,1]\to\mathbb R^3 &\quad,\quad \varepsilon(t)=\delta((b-a)t) \\ \varphi:[0,1]\to\mathbb R^3 &\quad,\quad \varphi(t)=\varepsilon(t^2) \\ \psi:[0,1]\to\mathbb R^3 &\quad,\quad \psi(t)=\varepsilon(1-t) \\ &\vdots \end{aligned}$$

Our claim, however, is that we can talk about integrals as follows, with $f,g,h:\mathbb R^3\to\mathbb R$, which are independent on the chosen parametrization of our path:

$$I=\int_\gamma f(x)dx_1+g(x)dx_2+h(x)dx_3$$

(2) In order to prove this, let us choose a parametrization $\gamma:[a,b]\to\mathbb R^3$ as above. This parametrization has as components three functions $\gamma_1,\gamma_2,\gamma_3$, given by:

$$\gamma=(\gamma_1,\gamma_2,\gamma_3):[a,b]\to\mathbb R^3$$

In order to construct the integral I, it is quite clear, by suitably cutting our path into pieces, that we can restrict the attention to the case where all three components $\gamma_1, \gamma_2, \gamma_3 : [a, b] \to \mathbb{R}$ are increasing, or decreasing. Thus, we can assume that these three components are as follows, increasing or decreasing, and bijective on their images:

$$\gamma_i : [a, b] \to [a_i, b_i]$$

(3) Moreover, by using the obvious symmetry between the coordinates x_1, x_2, x_3, in order to construct I, we just need to construct integrals of the following type:

$$I_1 = \int_\gamma f(x)dx_1$$

(4) So, let us construct this latter integral I_1, under the assumptions in (2). The simplest case is when the first path, $\gamma_1 : [a, b] \to [a_1, b_1]$, is the identity:

$$\gamma_1 : [a, b] \to [a, b] \quad , \quad \gamma_1(x) = x$$

In other words, the simplest case is when our path is of the following form, with $\gamma_2, \gamma_3 : [a, b] \to \mathbb{R}$ being certain functions, that should be increasing or decreasing, as per our conventions (2) above, but in what follows we will not need this assumption:

$$\gamma(x_1) = (x_1, \gamma_2(x_1), \gamma_3(x_1))$$

But now, with this convention made, we can define our contour integral, or rather its first component, as explained above, as a usual one-variable integral, as follows:

$$I_1 = \int_a^b f(x_1, \gamma_2(x_1), \gamma_3(x_1))dx_1$$

(5) With this understood, let us examine now the general case, where the first path, $\gamma_1 : [a, b] \to [a_1, b_1]$, is arbitrary, increasing or decreasing, and bijective on its image. In this case we can reparametrize our curve, as to have it as in (4) above, as follows:

$$\tilde{\gamma} = (id, \gamma_2\gamma_1^{-1}, \gamma_3\gamma_1^{-1}) : [a_1, b_1] \to \mathbb{R}^3$$

Now since we want our integral $I_1 = \int_\gamma f(x)dx_1$ to be independent of the parametrization, we are led to the following formula for it, coming from the formula in (4):

$$\begin{aligned} I_1 &= \int_{\tilde{\gamma}} f(x)dx_1 \\ &= \int_{a_1}^{b_1} f(x_1, \gamma_2\gamma_1^{-1}(x_1), \gamma_3\gamma_1^{-1}(x_1))dx_1 \\ &= \int_a^b f(\gamma_1(y_1), \gamma_2(y_1), \gamma_3(y_1))\gamma_1'(y_1)dy_1 \end{aligned}$$

Here we have used at the end the change of variable formula, with $x_1 = \gamma_1(y_1)$.

(6) Thus, job done, we have our definition for the contour integrals, with the formula being as follows, obtained by using (5) for all three coordinates x_1, x_2, x_3:

$$\begin{aligned} I &= \int_a^b f(\gamma_1(y_1), \gamma_2(y_1), \gamma_3(y_1))\gamma_1'(y_1)dy_1 \\ &+ \int_a^b g(\gamma_1(y_2), \gamma_2(y_2), \gamma_3(y_2))\gamma_2'(y_2)dy_2 \\ &+ \int_a^b h(\gamma_1(y_3), \gamma_2(y_3), \gamma_3(y_3))\gamma_3'(y_3)dy_3 \end{aligned}$$

And with this, we are led to the conclusion in the statement. □

Let us record as well the following more compact form of Theorem 8.2:

THEOREM 8.3. *The contour integrals over a curve $\gamma : [a, b] \to \mathbb{R}^3$ are given by*

$$\int_\gamma < F(x), dx > = \int_a^b < F(\gamma(y)), \gamma'(y)dy >$$

valid for any $F : \mathbb{R}^3 \to \mathbb{R}^3$, where on the right $\gamma'(y)dy = (\gamma_i'(y_i)dy_i)_i$.

PROOF. This is a fancy reformulation of what we did in Theorem 8.2 and its proof. Indeed, with the notation $F = (F_1, F_2, F_3) = (f, g, h)$, the integral computed there is:

$$\int_\gamma < F(x), dx > = \int_\gamma F_1(x)dx_1 + F_2(x)dx_2 + F_3(x)dx_3$$

As for the value of this integral, according to the proof of Theorem 8.2, this is:

$$\begin{aligned} \int_\gamma < F(x), dx > &= \int_a^b F_1(\gamma_1(y_1), \gamma_2(y_1), \gamma_3(y_1))\gamma_1'(y_1)dy_1 \\ &+ \int_a^b F_2(\gamma_1(y_2), \gamma_2(y_2), \gamma_3(y_2))\gamma_2'(y_2)dy_2 \\ &+ \int_a^b F_3(\gamma_1(y_3), \gamma_2(y_3), \gamma_3(y_3))\gamma_3'(y_3)dy_3 \end{aligned}$$

Now observe that we can write this in a more compact way, as follows:

$$\begin{aligned} \int_\gamma < F(x), dx > &= \int_a^b F_1(\gamma(y_1))\gamma'(y_1)dy_1 \\ &+ \int_a^b F_2(\gamma(y_2))\gamma'(y_2)dy_2 \\ &+ \int_a^b F_3(\gamma(y_3))\gamma'(y_3)dy_3 \end{aligned}$$

And we can do even better. Indeed, we have only one integral here, $\int_a^b$, and in order to best express the integrand, consider the formal vector in the statement, namely:

$$\gamma'(y)dy = \begin{pmatrix}\gamma_1'(y_1)dy_1\\ \gamma_2'(y_2)dy_2\\ \gamma_3'(y_3)dy_3\end{pmatrix}$$

Our integrand appears then as the scalar product of $F(\gamma(y))$ with this latter vector $\gamma'(y)dy$, so our formula above for the contour integral takes the following form:

$$\int_\gamma < F(x), dx >= \int_a^b < F(\gamma(y)), \gamma'(y)dy >$$

Thus, we are led to the conclusion in the statement. □

More concretely now, let us temporarily forget about the paths γ, and have a look at the quantities which are to be integrated, namely:

$$\alpha = F_1(x)dx_1 + F_2(x)dx_2 + F_3(x)dx_3$$

Obviously, these are something rather mathematical, and many things can be said here. However, we can have some physical intuition on them. Assume indeed that we are given a function as follows, that you can think for instance as corresponding to an external force, with $F(x) \in \mathbb{R}^3$ being the force vector applied at a given point $x \in \mathbb{R}^3$:

$$F : \mathbb{R}^3 \to \mathbb{R}^3$$

By writing $F = (F_1, F_2, F_3)$, we can then consider the following quantity, and when $F : \mathbb{R}^3 \to \mathbb{R}^3$ varies, we obtain exactly the abstract quantities α considered above:

$$< F(x), dx >= F_1(x)dx_1 + F_2(x)dx_2 + F_3(x)dx_3$$

Thus, all in all, what we have done in the above with our construction of contour integrals, was to define quantities as follows, with γ being a path in $\mathbb{R}^3$, and with $F : \mathbb{R}^3 \to \mathbb{R}^3$ being a certain function, that we can think of, if we want, as being a force:

$$I = \int_\gamma < F(x), dx >$$

Which brings us into physics. Indeed, by assuming now that $F : \mathbb{R}^3 \to \mathbb{R}^3$ does correspond to a force, we can formulate the following definition:

Definition 8.4. *The work done by a force $F = F(x)$ for moving a particle from point $p \in \mathbb{R}^3$ to point $q \in \mathbb{R}^3$ via a given path $\gamma : p \to q$ is the following quantity:*

$$W(\gamma) = \int_\gamma < F(x), dx >$$

We say that F is conservative if this work quantity $W(\gamma)$ does not depend on the chosen path $\gamma : p \to q$, and in this case we denote this quantity by $W(p, q)$.

This definition is something quite subtle, and as a first comment, assume that we have two paths $\gamma : p \to q$ and $\delta : p \to q$. We can then consider the path $\circ : p \to p$ obtained by going along $\gamma : p \to q$, and then along δ reversed, $\delta^{-1} : q \to p$, and we have:

$$W(\circ) = W(\gamma) - W(\delta)$$

Thus F is conservative precisely when, for any loop $\circ : p \to p$, we have:

$$W(\circ) = 0$$

Intuitively, this means that F is some sort of "clean", ideal force, with no dirty things like friction involved. As we will soon see, gravity is such a clean force, with a simple example coming from throwing a rock up in the sky. That rock will travel on a loop $p \to q \to p$, and will come back here to p unchanged, save for the fact that its speed vector is reversed. Thus, and assuming now that work has something to do with energy, which is intuitive, there has been no overall work of gravity on this loop, $W(\circ) = 0$.

As a first result now, regarding the conservative forces, we have:

THEOREM 8.5. *The work done by a conservative force F on a mass m object is*

$$W(p, q) = T(q) - T(p)$$

with $T = m||v||^2/2$ standing as usual for the kinetic energy of the object.

PROOF. Assuming that F is conservative, and acts via the usual formula $F = ma$ on our object of mass m, we have the following computation, as desired:

$$\begin{aligned} W(p,q) &= \int_p^q < F(x), dx > \\ &= m \int_p^q < a(x), dx > \\ &= m \int_p^q \left\langle \frac{dv(x)}{dt}, v(x)dt \right\rangle \\ &= \frac{m}{2} \int_p^q \frac{d < v(x), v(x) >}{dt}\, dt \\ &= \frac{m}{2} \int_p^q \frac{d||v(x)||^2}{dt}\, dt \\ &= \frac{m}{2} \left(||v(q)||^2 - ||v(p)||^2 \right) \\ &= T(q) - T(p) \end{aligned}$$

Here we have used in the middle the fact that the time derivative of a scalar product of functions $< v, w >$ consists of two terms, which are equal when $v = w$. $\square$

In order to formulate our next result, observe that we have the following computation for the contour integrals of gradients, which is independent on the chosen path:

$$\begin{aligned}\int_p^q <\nabla V(x), dx> &= \int_p^q \frac{dV}{dx_1}\cdot dx_1 + \frac{dV}{dx_2}\cdot dx_2 + \frac{dV}{dx_3}\cdot dx_3\\ &= \int_p^q dV\\ &= V(q) - V(p)\end{aligned}$$

To be more precise, this computation certainly works when V is a function of just one variable, x_1, x_2 or x_3, thanks to the fundamental theorem of calculus, and the general case follows from this, by using the chain rule for derivatives.

Now with this formula in hand, we can formulate the following result, which is quite conceptual, and which includes some basic gravitation physics too, at the end:

THEOREM 8.6. *A force F is conservative precisely when it is of the form*

$$F = -\nabla V$$

for a certain function V, and in this case the work done by it is given by:

$$W(p, q) = V(p) - V(q)$$

Also, the gravitation force is conservative, coming from $V = -C/||x||$, with $C > 0$.

PROOF. This is something quite tricky, the idea being as follows:

(1) In one sense, assume that F is conservative. Since the work $W(p, q) = W(\gamma)$ is independent of the chosen path $\gamma : p \to q$, we can find a function V such that:

$$W(p, q) = V(p) - V(q)$$

Observe that this function V is well-defined up to an additive constant. Now with this formula in hand, we further obtain, as desired:

$$\begin{aligned}W(p,q) = V(p) - V(q) &\implies \int_p^q < F(x), dx >= -\int_p^q <\nabla V(x), dx>\\ &\implies < F(x), dx >= - <\nabla V(x), dx>\\ &\implies F_i(x)dx_i = -\frac{dV}{dx_i}\cdot dx_i\\ &\implies F_i(x) = -\frac{dV}{dx_i}\\ &\implies F = -\nabla V\end{aligned}$$

(2) In the other sense now, assuming $F = -\nabla V$, we have the following computation, valid for any loop $\circ : p \to p$, which shows that F is indeed conservative:

$$W(\circ) = -\int_\circ \nabla V = 0$$

More generally, regarding the work done by such a force $F = -\nabla V$, along a path $\gamma : p \to q$, which is independent on this path γ, this is given by:

$$W(p,q) = -\int_p^q \nabla V = V(p) - V(q)$$

(3) Finally, regarding the last assertion, we will leave this as an exercise. □

We can now put everything together, and we have the following result:

Theorem 8.7. *Given a conservative force F, appearing as follows, with V being uniquely determined up to an additive constant,*

$$F = -\nabla V$$

the movements of a particle under F preserve the total energy, given by

$$E = T + V$$

with $T = m||v||^2/2$ being the kinetic energy, and with V being called potential energy.

Proof. By using Theorem 8.5 and Theorem 8.7, we have:

$$W(p,q) = T(q) - T(p) \quad , \quad W(p,q) = V(p) - V(q)$$

Now observe that these equalities give the following formula:

$$T(p) + V(p) = T(q) + V(q)$$

Thus, the total energy $E = T + V$ is conserved, as claimed. □

8c. Green and Stokes

At the level of mathematics now, in relation with the above, a useful result is:

Theorem 8.8 (Green). *Given a plane curve $C \subset \mathbb{R}^2$, we have the formula*

$$\int_C Pdx + Qdy = \int_D \left(\frac{dQ}{dx} - \frac{dP}{dy}\right) dxdy$$

where $D \subset \mathbb{R}^2$ is the domain enclosed by C.

Proof. Assume indeed that we have a plane curve $C \subset \mathbb{R}^2$, without self-intersections, which is piecewise C^1, and is assumed to be counterclockwise oriented. In order to prove the formula regarding P, we can parametrize the enclosed domain D as follows:

$$D = \left\{(x,y)\Big| a \leq x \leq b, f(x) \leq y \leq g(x)\right\}$$

We have then the following computation, which gives the result for P:

$$\begin{aligned}\int_D -\frac{dP}{dy}\,dxdy &= \int_a^b\left(\int_{f(x)}^{g(x)} -\frac{dP}{dy}(x,y)dy\right)dx\\ &= \int_a^b P(x,f(x))-P(x,g(x))dx\\ &= \int_a^b P(x,f(x))dx+\int_b^a P(x,g(x))dx\\ &= \int_C Pdx\end{aligned}$$

As for the result for the Q term, the computation here is similar. ☐

Moving forward to 3D, let us start with a standard definition, as follows:

Definition 8.9. *The vector product of two vectors in $\mathbb{R}^3$ is given by*

$$x\times y=||x||\cdot||y||\cdot\sin\theta\cdot n$$

where $n\in\mathbb{R}^3$ with $n\perp x,y$ and $||n||=1$ is constructed using the right-hand rule:

$$\begin{matrix} & \uparrow_{x\times y}\\ \leftarrow_x & \\ \swarrow_y & \end{matrix}$$

Alternatively, in usual vertical linear algebra notation for all vectors,

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}\times\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}=\begin{pmatrix}x_2y_3-x_3y_2\\x_3y_1-x_1y_3\\x_1y_2-x_2y_1\end{pmatrix}$$

the rule being that of computing 2×2 determinants, and adding a middle sign.

In practice now, in order to get familiar with the vector products, nothing better than doing some classical mechanics. We have here the following key result:

Theorem 8.10. *In the gravitational 2-body problem, the angular momentum*

$$J=x\times p$$

with $p=mv$ being the usual momentum, is conserved.

Proof. There are several things to be said here, the idea being as follows:

(1) First of all the usual momentum, $p=mv$, is not conserved, because the simplest solution is the circular motion, where the moment gets turned around. But this suggests precisely that, in order to fix the lack of conservation of the momentum p, what we have to do is to make a vector product with the position x. Leading to J, as above.

(2) Regarding now the proof, consider indeed a particle m moving under the gravitational force of a particle M, assumed, as usual, to be fixed at 0. By using the fact that for two proportional vectors, $p \sim q$, we have $p \times q = 0$, we obtain:

$$\begin{aligned}\dot{J} &= \dot{x} \times p + x \times \dot{p}\\ &= v \times mv + x \times ma\\ &= m(v \times v + x \times a)\\ &= m(0+0)\\ &= 0\end{aligned}$$

Now since the derivative of J vanishes, this quantity is constant, as stated. □

At the mathematical level now, we have the following key result:

THEOREM 8.11. *The area of a surface $S \subset \mathbb{R}^3$, parametrized as $S = r(D)$ with $r : D \to \mathbb{R}^3$ and $D \subset \mathbb{R}^2$, is given by*

$$A(S) = \int_D ||r_x \times r_y||\, dxdy$$

with $r_x, r_y : D \to \mathbb{R}^3$ being the partial derivatives of r, and $||.||$ being the norm of $\mathbb{R}^3$.

PROOF. Thus is something quite similar to Theorem 8.2, and many things can be said here, at the theoretical level, in analogy with those for the curves. Among others, let us mention that we can talk, more generally, about surface integrals, defined as follows:

$$\int_S f(s)ds = \int_D f(r(x,y))||r_x \times r_y||\, dxdy$$

As for the basic illustrations of the formula in the statement, consider a surface of type $z = f(x,y)$. Here we have $r(x,y) = (x, y, f(x,y))$, and we obtain:

$$\begin{aligned}A(S) &= \int_D \left|\left|\begin{pmatrix}1\\0\\f_x\end{pmatrix} \times \begin{pmatrix}0\\1\\f_y\end{pmatrix}\right|\right| dxdy\\ &= \int_D \left|\left|\begin{pmatrix}-f_x\\-f_y\\1\end{pmatrix}\right|\right| dxdy\\ &= \int_D \sqrt{f_x^2 + f_y^2 + 1}\, dxdy\end{aligned}$$

There are of course many other applications of the formula in the statement, and we will leave some thinking and computations here as an instructive exercise. □

Many other things can be said, as a continuation of the above, but let us not deviate too much, from what we wanted to do here. Next, comes the following definition:

Definition 8.12. *We can talk about the divergence of $\varphi:\mathbb{R}^3\to\mathbb{R}^3$, as being*

$$<\nabla,\varphi>=\left\langle\begin{pmatrix}\frac{d}{dx}\\ \frac{d}{dy}\\ \frac{d}{dz}\end{pmatrix},\begin{pmatrix}\varphi_x\\ \varphi_y\\ \varphi_z\end{pmatrix}\right\rangle=\frac{d\varphi_x}{dx}+\frac{d\varphi_y}{dy}+\frac{d\varphi_z}{dz}$$

as well as about the curl of the same function $\varphi:\mathbb{R}^3\to\mathbb{R}^3$, as being

$$\nabla\times\varphi=\begin{vmatrix}u_x&\frac{d}{dx}&\varphi_x\\ u_y&\frac{d}{dy}&\varphi_y\\ u_z&\frac{d}{dz}&\varphi_z\end{vmatrix}=\begin{pmatrix}\frac{d\varphi_z}{dy}-\frac{d\varphi_y}{dz}\\ \frac{d\varphi_x}{dz}-\frac{d\varphi_z}{dx}\\ \frac{d\varphi_y}{dx}-\frac{d\varphi_x}{dy}\end{pmatrix}$$

where u_x,u_y,u_z are the unit vectors along the coordinate directions x,y,z.

Getting back now to calculus tools, in 3 dimensions, we have the following result:

Theorem 8.13 (Stokes). *Given a smooth oriented surface $S\subset\mathbb{R}^3$, with boundary $C\subset\mathbb{R}^3$, and a vector field F, we have the following formula:*

$$\int_S<(\nabla\times F)(x),n(x)>dx=\int_C<F(x),dx>$$

In other words, the line integral of a vector field F over a loop C equals the surface integral of the curl of the vector field φ, over the enclosed surface S.

Proof. This basically follows from the Green theorem, the idea being as follows:

(1) Let us first parametrize our surface S, and its boundary C. We can assume that we are in the situation where we have a closed oriented curve $\gamma:[a,b]\to\mathbb{R}^2$, with interior $D\subset\mathbb{R}^2$, and where the surface appears as $S=\psi(D)$, with $\psi:D\to\mathbb{R}^3$. In this case, the function $\delta=\psi\circ\gamma$ parametrizes the boundary of our surface, $C=\delta[a,b]$.

(2) Let us first look at the integral on the right in the statement. We have the following formula, with $J_y(\psi)$ standing for the Jacobian of ψ at the point $y=\gamma(t)$:

$$\begin{aligned}\int_C<F(x),dx>&=\int_\gamma<F(\psi(\gamma)),d\psi(\gamma)>\\&=\int_\gamma<F(\psi(y)),J_y(\psi)d\gamma>\end{aligned}$$

In order to further process this formula, let us introduce the following function:

$$P(u,v)=\left\langle F(\psi(u,v)),\frac{d\psi}{du}(u,v)\right\rangle e_u+\left\langle F(\psi(u,v)),\frac{d\psi}{dv}(u,v)\right\rangle e_v$$

In terms of this function, we have the following formula for our line integral:

$$\int_C<F(x),dx>=\int_\gamma<P(y),dy>$$

(3) In order to compute now the other integral in the statement, we first have:

$$\begin{aligned}
&\frac{dP_v}{du}-\frac{dP_u}{dv}\\
=&\left\langle\frac{d(F\psi)}{du},\frac{d\psi}{dv}\right\rangle+\left\langle F\psi,\frac{d^2\psi}{dudv}\right\rangle-\left\langle\frac{d(F\psi)}{dv},\frac{d\psi}{du}\right\rangle-\left\langle F\psi,\frac{d^2\psi}{dvdu}\right\rangle\\
=&\left\langle\frac{d(F\psi)}{du},\frac{d\psi}{dv}\right\rangle-\left\langle\frac{d(F\psi)}{dv},\frac{d\psi}{du}\right\rangle\\
=&\left\langle\frac{d\psi}{dv},\left(J_{\psi(u,v)}F-(J_{\psi(u,v)}F)^t\right)\frac{d\psi}{du}\right\rangle\\
=&\left\langle\frac{d\psi}{dv},(\nabla\times F)\times\frac{d\psi}{du}\right\rangle\\
=&\left\langle\nabla\times F,\frac{d\psi}{du}\times\frac{d\psi}{dv}\right\rangle
\end{aligned}$$

We conclude that the integral on the left in the statement is given by:

$$\begin{aligned}
&\int_S<(\nabla\times F)(x),n(x)>dx\\
=&\int_D\left\langle(\nabla\times F)(\psi(u,v)),\frac{d\psi}{du}(u,v)\times\frac{d\psi}{dv}(u,v)\right\rangle dudv\\
=&\int_D\left(\frac{dP_v}{du}-\frac{dP_u}{dv}\right)dudv
\end{aligned}$$

(4) But with this, we are done, because the integrals computed in (2) and (3) are indeed equal, due to the Green theorem. Thus, the Stokes formula holds indeed. □

As a conclusion to what we did so far, we have the following statement:

THEOREM 8.14. *The following results hold, in 3 dimensions:*

(1) *Fundamental theorem for gradients, namely*

$$\int_a^b<\nabla f,dx>=f(b)-f(a)$$

(2) *Fundamental theorem for divergences, or Gauss or Green formula,*

$$\int_B<\nabla,\varphi>=\int_S<\varphi(x),n(x)>dx$$

(3) *Fundamental theorem for curls, or Stokes formula,*

$$\int_A<(\nabla\times\varphi)(x),n(x)>dx=\int_P<\varphi(x),dx>$$

where S is the boundary of the body B, and P is the boundary of the area A.

PROOF. This follows indeed from the various formulae established above. □

8d. Gauss, Maxwell

We would like to discuss now, following Gauss and others, some applications of the above to electrostatics, and why not to electrodynamics too. Let us start with:

FACT 8.15 (Coulomb law). *Any pair of charges $q_1, q_2 \in \mathbb{R}$ is subject to a force as follows, which is attractive if $q_1 q_2 < 0$ and repulsive if $q_1 q_2 > 0$,*

$$||F|| = K \cdot \frac{|q_1 q_2|}{d^2}$$

where $d > 0$ is the distance between the charges, and $K > 0$ is a certain constant.

Observe the amazing similarity with the Newton law for gravity. However, as we will soon see, passed a few simple facts, things will be more complicated here.

In analogy with our study of gravity, let us start with:

DEFINITION 8.16. *Given charges $q_1, \ldots, q_k \in \mathbb{R}$ located at positions $x_1, \ldots, x_k \in \mathbb{R}^3$, we define their electric field to be the vector function*

$$E(x) = K \sum_i \frac{q_i (x - x_i)}{||x - x_i||^3}$$

so that their force applied to a charge $Q \in \mathbb{R}$ positioned at $x \in \mathbb{R}^3$ is given by $F = QE$.

Observe the analogy with gravity, save for the fact that instead of masses $m > 0$ we have now charges $q \in \mathbb{R}$, and that at the level of constants, G gets replaced by K.

More generally, we will be interested in electric fields of various non-discrete charge configurations, such as charged curves, surfaces and solid bodies. So, let us formulate:

DEFINITION 8.17. *The electric field of a charge configuration $L \subset \mathbb{R}^3$, with charge density function $\rho : L \to \mathbb{R}$, is the vector function*

$$E(x) = K \int_L \frac{\rho(z)(x - z)}{||x - z||^3} \, dz$$

so that the force of L applied to a charge Q positioned at x is given by $F = QE$.

With the above definitions in hand, it is most convenient now to forget about the charges, and focus on the study of the corresponding electric fields E.

These fields are by definition vector functions $E : \mathbb{R}^3 \to \mathbb{R}^3$, with the convention that they take $\pm\infty$ values at the places where the charges are located, and intuitively, are best represented by their field lines, which are constructed as follows:

Definition 8.18. *The field lines of an electric field $E:\mathbb{R}^3\to\mathbb{R}^3$ are the oriented curves $\gamma\subset\mathbb{R}^3$ pointing at every point $x\in\mathbb{R}^3$ at the direction of the field, $E(x)\in\mathbb{R}^3$.*

As a basic example here, for one charge the field lines are the half-lines emanating from its position, oriented according to the sign of the charge:

$$\begin{matrix}\nwarrow&\uparrow&\nearrow\\ \leftarrow&\oplus&\rightarrow\\ \swarrow&\downarrow&\searrow\end{matrix}\qquad\qquad\qquad\begin{matrix}\searrow&\downarrow&\swarrow\\ \rightarrow&\ominus&\leftarrow\\ \nearrow&\uparrow&\nwarrow\end{matrix}$$

For two charges now, if these are of opposite signs, $+$ and $-$, you get a picture that you are very familiar with, namely that of the field lines of a bar magnet:

$$\begin{matrix}\nearrow&\nearrow&\rightarrow&\rightarrow&\rightarrow&\rightarrow&\searrow&\searrow\\ \nwarrow&\uparrow&\nearrow&\rightarrow&\rightarrow&\searrow&\downarrow&\swarrow\\ \leftarrow&\oplus&\rightarrow&\rightarrow&\rightarrow&\rightarrow&\ominus&\leftarrow\\ \swarrow&\downarrow&\searrow&\rightarrow&\rightarrow&\nearrow&\uparrow&\nwarrow\\ \searrow&\searrow&\rightarrow&\rightarrow&\rightarrow&\rightarrow&\nearrow&\nearrow\end{matrix}$$

If the charges are $+,+$ or $-,-$, you get something of similar type, but repulsive this time, with the field lines emanating from the charges being no longer shared:

$$\begin{matrix}\leftarrow&\nwarrow&\nwarrow\\ &\uparrow&\nearrow\\ \leftarrow&\oplus&\\ &\downarrow&\searrow\\ \leftarrow&\swarrow&\swarrow\end{matrix}\qquad\qquad\qquad\begin{matrix}\nearrow&\nearrow&\rightarrow\\ \nwarrow&\uparrow&\\ &\oplus&\rightarrow\\ \swarrow&\downarrow&\\ \searrow&\searrow&\rightarrow\end{matrix}$$

These pictures, and notably the last one, with $+,+$ charges, are quite interesting, because the repulsion situation does not appear in the context of gravity. Thus, we can only expect our geometry here to be far more complicated than that of gravity.

In general now, the first thing that can be said about the field lines is that, by definition, they do not cross. Thus, what we have here is some sort of oriented 1D foliation of $\mathbb{R}^3$, in the sense that $\mathbb{R}^3$ is smoothly decomposed into oriented curves $\gamma\subset\mathbb{R}^3$.

The field lines, as constructed in Definition 8.18, obviously do not encapsulate the whole information about the field, with the direction of each vector $E(x)\in\mathbb{R}^3$ being there, but with the magnitude $||E(x)||\geq 0$ of this vector missing. However, say when drawing, when picking up uniformly radially spaced field lines around each charge, and with the number of these lines proportional to the magnitude of the charge, and then completing the picture, the density of the field lines around each point $x\in\mathbb{R}$ will give you the magnitude $||E(x)||\geq 0$ of the field there, up to a scalar. So, let us formulate:

Proposition 8.19. *Given an electric field $E:\mathbb{R}^3\to\mathbb{R}^3$, the knowledge of its field lines is the same as the knowledge of the composition*

$$nE:\mathbb{R}^3\to\mathbb{R}^3\to S$$

where $S\subset\mathbb{R}^3$ is the unit sphere, and $n:\mathbb{R}^3\to S$ is the rescaling map, namely:

$$n(x)=\frac{x}{||x||}$$

However, in practice, when the field lines are accurately drawn, the density of the field lines gives you the magnitude of the field, up to a scalar.

Proof. We have two assertions here, the idea being as follows:

(1) The first assertion is clear from definitions, with of course our usual convention that the electric field and its problematics take place outside the locations of the charges, which makes everything in the statement to be indeed well-defined.

(2) Regarding now the last assertion, which is of course a bit informal, this follows from the above discussion. It is possible to be a bit more mathematical here, with a definition, formula and everything, but we will not need this, in what follows. □

Let us introduce now a key definition, as follows:

Definition 8.20. *The flux of an electric field $E:\mathbb{R}^3\to\mathbb{R}^3$ through a surface $S\subset\mathbb{R}^3$, assumed to be oriented, is the quantity*

$$\Phi_E(S)=\int_S < E(x),n(x) > dx$$

with $n(x)$ being unit vectors orthogonal to S, following the orientation of S. Intuitively, the flux measures the signed number of field lines crossing S.

Here by orientation of S we mean precisely the choice of unit vectors $n(x)$ as above, orthogonal to S, which must vary continuously with x. For instance a sphere has two possible orientations, one with all these vectors $n(x)$ pointing inside, and one with all these vectors $n(x)$ pointing outside. More generally, any surface has locally two possible orientations, so if it is connected, it has two possible orientations. In what follows the convention is that the closed surfaces are oriented with each $n(x)$ pointing outside.

Regarding the last sentence of Definition 8.20, this is of course something informal, meant to help, coming from the interpretation of the field lines from Proposition 8.19. However, we will see later that this simple interpretation can be of great use.

As a first illustration, let us do a basic computation, as follows:

PROPOSITION 8.21. *For a point charge $q \in \mathbb{R}$ at the center of a sphere S,*

$$\Phi_E(S) = \frac{q}{\varepsilon_0}$$

where the constant is $\varepsilon_0 = 1/(4\pi K)$, independently of the radius of S.

PROOF. Assuming that S has radius r, we have the following computation:

$$\begin{aligned}
\Phi_E(S) &= \int_S < E(x), n(x) > dx \\
&= \int_S \left\langle \frac{Kqx}{r^3}, \frac{x}{r} \right\rangle dx \\
&= \int_S \frac{Kq}{r^2} \, dx \\
&= \frac{Kq}{r^2} \times 4\pi r^2 \\
&= 4\pi K q
\end{aligned}$$

Thus with $\varepsilon_0 = 1/(4\pi K)$ as above, we obtain the result. □

More generally now, we have the following key result, due to Gauss, which is the foundation of advanced electrostatics, and of everything following from it, namely electrodynamics, and then quantum mechanics, and particle physics:

THEOREM 8.22 (Gauss law). *The flux of a field E through a surface S is given by*

$$\Phi_E(S) = \frac{Q_{enc}}{\varepsilon_0}$$

where Q_{enc} is the total charge enclosed by S, and $\varepsilon_0 = 1/(4\pi K)$.

PROOF. This basically follows from Proposition 8.21, a bit modified, by adding to the computation there a number of standard ingredients. We refer here for instance to Feynman [32], but we will be back to this right next, with a more advanced proof. □

In relation now with our previous mathematics, we have the following result:

THEOREM 8.23. *Given an electric potential E, its divergence is given by*

$$< \nabla, E > = \frac{\rho}{\varepsilon_0}$$

where ρ denotes as usual the charge distribution. Also, we have

$$\nabla \times E = 0$$

meaning that the curl of E vanishes.

PROOF. We have several assertions here, the idea being as follows:

(1) The first formula, called Gauss law in differential form, follows from:

$$\begin{aligned}\int_B < \nabla, E > &= \int_S < E(x), n(x) > dx\\ &= \Phi_E(S)\\ &= \frac{Q_{enc}}{\varepsilon_0}\\ &= \int_B \frac{\rho}{\varepsilon_0}\end{aligned}$$

Now since this must hold for any B, this gives the formula in the statement.

(2) Regarding the curl, by discretizing and linearity we can assume that we are dealing with a single charge q, positioned at 0. We have, by using spherical coordinates r,s,t:

$$\begin{aligned}\int_a^b < E(x), dx > &= \int_a^b \left\langle \frac{Kqx}{||x||^3}, dx\right\rangle\\ &= \int_a^b \left\langle \frac{Kq}{r^2}\cdot\frac{x}{||x||}, dx\right\rangle\\ &= \int_a^b \frac{Kq}{r^2}\,dr\\ &= \left[-\frac{Kq}{r}\right]_a^b\\ &= Kq\left(\frac{1}{r_a}-\frac{1}{r_b}\right)\end{aligned}$$

In particular the integral of E over any closed loop vanishes, and by using now the Stokes formula, we conclude that the curl of E vanishes, as stated. □

So long for electrostatics, which provide a good motivation and illustration for our mathematics. When upgrading to electrodynamics, things become even more interesting, because our technology can be used in order to understand the Maxwell equations:

THEOREM 8.24. *Electrodynamics is governed by the formulae*

$$< \nabla, E > = \frac{\rho}{\varepsilon_0} \quad , \quad < \nabla, B > = 0$$

$$\nabla \times E = -\dot{B} \quad , \quad \nabla \times B = \mu_0 J + \mu_0\varepsilon_0 \dot{E}$$

called Maxwell equations.

Proof. This is something fundamental, coming as a tricky mixture of physics and mathematics. To be more precise, the first formula is the Gauss law, ρ being the charge, and ε_0 being a constant, and with this Gauss law more or less replacing the Coulomb law from electrostatics. The second formula is something basic, and anonymous. The third formula is the Faraday law. As for the fourth formula, this is the Ampère law, as modified by Maxwell, with J being the volume current density, and μ_0 being a constant. □

However, the above is not all. Quite remarkably, the constants μ_0, ε_0 are related by the following formula, due to Biot-Savart, with c being the speed of light:

$$\mu_0\varepsilon_0 = \frac{1}{c^2}$$

So, what has light to do with all this? The idea is that accelerating or decelerating charges produce electromagnetic radiation, of various wavelengths, called light, of various colors, and with all this coming from the mathematics of the Maxwell equations.

8e. Exercises

We had an exciting chapter here, and as exercises on this, we have:

Exercise 8.25. *Compute the lengths of some curves, of your choice.*

Exercise 8.26. *Learn more about conservative forces, and their properties.*

Exercise 8.27. *Compute the areas of some surfaces, of your choice.*

Exercise 8.28. *Learn more about vectors products, curl, gradient and divergence.*

Exercise 8.29. *Learn about the other basic applications of the Green formula.*

Exercise 8.30. *Clarify all the details in the proof of the Stokes formula.*

Exercise 8.31. *Try to compute the field lines, for simple charge configurations.*

Exercise 8.32. *Learn more about the Gauss law, and the Maxwell equations.*

As bonus exercise, read more electrodynamics. It is all about it.

Part III

Riemannian manifolds

CHAPTER 9

Length, area, volume

9a. Computing lengths

Computing lengths.

9b. Computing areas

Computing areas.

9c. Computing volumes

Computing volumes.

9d. Further computations

Further computations.

9e. Exercises

Exercises:

Exercise 9.1.

Exercise 9.2.

Exercise 9.3.

Exercise 9.4.

Exercise 9.5.

Exercise 9.6.

Exercise 9.7.

Exercise 9.8.

Bonus exercise.

CHAPTER 10

Riemannian manifolds

10a. Riemannian manifolds

10b.

10c.

10d.

10e. Exercises

Exercises:

Exercise 10.1.

Exercise 10.2.

Exercise 10.3.

Exercise 10.4.

Exercise 10.5.

Exercise 10.6.

Exercise 10.7.

Exercise 10.8.

Bonus exercise.

CHAPTER 11

Dirac operator

11a. Dirac operator

11b.

11c.

11d.

11e. Exercises

Exercises:

EXERCISE 11.1.

EXERCISE 11.2.

EXERCISE 11.3.

EXERCISE 11.4.

EXERCISE 11.5.

EXERCISE 11.6.

EXERCISE 11.7.

EXERCISE 11.8.

Bonus exercise.

CHAPTER 12

Nash embedding

12a. Nash embedding

12b.

12c.

12d.

12e. Exercises

Exercises:

EXERCISE 12.1.

EXERCISE 12.2.

EXERCISE 12.3.

EXERCISE 12.4.

EXERCISE 12.5.

EXERCISE 12.6.

EXERCISE 12.7.

EXERCISE 12.8.

Bonus exercise.

Part IV

Lorentz manifolds

CHAPTER 13

Relativity theory

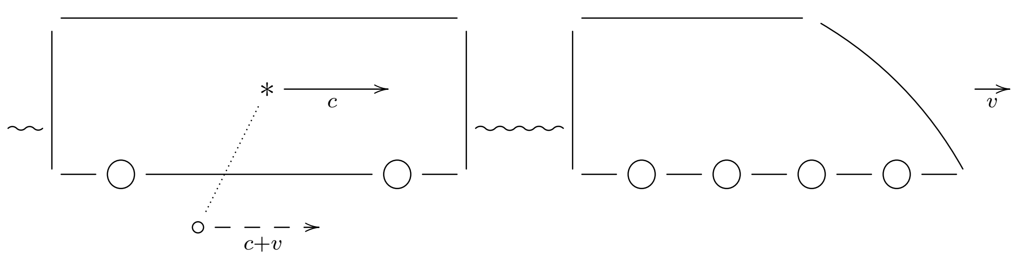
13a. Relativity theory

We would like to first discuss some basic theoretical physics, asking for more geometry. Based on experiments by Fizeau, then Michelson-Morley and others, and some theoretical physics by Maxwell and Lorentz too, Einstein came upon the following principles:

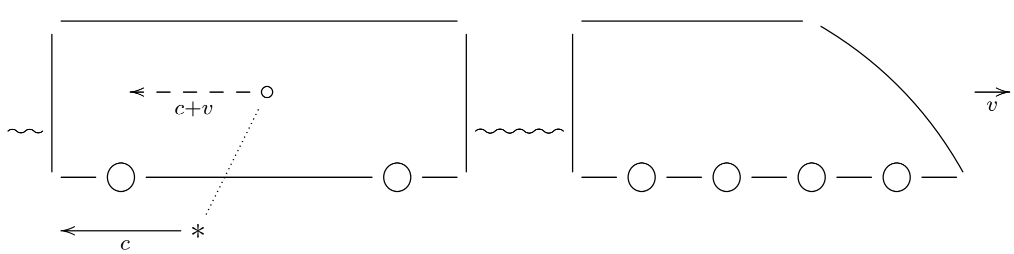
FACT 13.1 (Einstein principles). *The following happen:*

(1) *Light travels in vacuum at a finite speed, $c < \infty$.*
(2) *This speed c is the same for all inertial observers.*
(3) *In non-vacuum, the light speed is lower, $v < c$.*
(4) *Nothing can travel faster than light, $v \not> c$.*

The point now is that, obviously, something is wrong here. Indeed, assuming for instance that we have a train, running in vacuum at speed $v > 0$, and someone on board lights a flashlight $*$ towards the locomotive, then an observer $\circ$ on the ground will see the light traveling at speed $c + v > c$, which is a contradiction:



Equivalently, with the same train running, in vacuum at speed $v > 0$, if the observer on the ground lights a flashlight $*$ towards the back of the train, then viewed from the train, that light will travel at speed $c + v > c$, which is a contradiction again:



Summarizing, Fact 13.1 implies $c + v = c$, so contradicts classical mechanics, which therefore needs a fix. By dividing all speeds by c, as to have $c = 1$, and by restricting the attention to the 1D case, to start with, we are led to the following puzzle:

PUZZLE 13.2. *How to define speed addition on the space of* 1D *speeds, which is*

$$I = [-1, 1]$$

with our $c = 1$ *convention, as to have* $1 + c = 1$*, as required by physics?*

In view of our basic geometric knowledge, a natural idea here would be that of wrapping $[-1, 1]$ into a circle, and then stereographically projecting on $\mathbb{R}$. Indeed, we can then "import" to $[-, 1, 1]$ the usual addition on $\mathbb{R}$, via the inverse of this map.

So, let us see where all this leads us. First, the formula of our map is as follows:

PROPOSITION 13.3. *The map wrapping* $[-1, 1]$ *into the unit circle, and then stereographically projecting on* $\mathbb{R}$ *is given by the formula*

$$\varphi(u) = \tan\left(\frac{\pi u}{2}\right)$$

with the convention that our wrapping is the most straightforward one, making correspond $\pm 1 \to i$*, with negatives on the left, and positives on the right.*

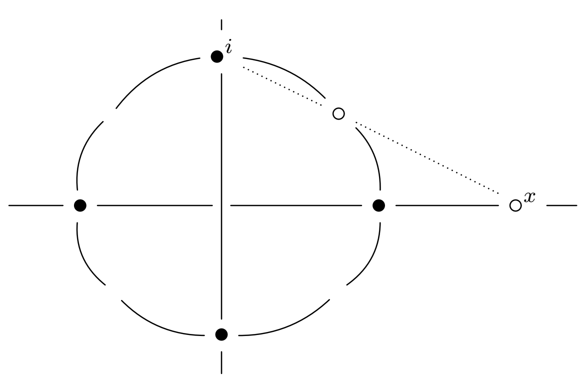
PROOF. Regarding the wrapping, as indicated, this is given by:

$$u \to e^{it} \quad , \quad t = \pi u - \frac{\pi}{2}$$

Indeed, this correspondence wraps $[-1, 1]$ as above, the basic instances of our correspondence being as follows, and with everything being fine modulo 2π:

$$-1 \to \frac{\pi}{2} \quad , \quad -\frac{1}{2} \to -\pi \quad , \quad 0 \to -\frac{\pi}{2} \quad , \quad \frac{1}{2} \to 0 \quad , \quad 1 \to \frac{\pi}{2}$$

Regarding now the stereographic projection, the picture here is as follows:



Thus, by Thales, the formula of the stereographic projection is as follows:

$$\frac{\cos t}{x}=\frac{1-\sin t}{1}\implies x=\frac{\cos t}{1-\sin t}$$

Now if we compose our wrapping operation above with the stereographic projection, what we get is, via the above Thales formula, and some trigonometry:

$$\begin{aligned}
x &= \frac{\cos t}{1-\sin t}\\
&= \frac{\cos\left(\pi u-\frac{\pi}{2}\right)}{1-\sin\left(\pi u-\frac{\pi}{2}\right)}\\
&= \frac{\cos\left(\frac{\pi}{2}-\pi u\right)}{1+\sin\left(\frac{\pi}{2}-\pi u\right)}\\
&= \frac{\sin(\pi u)}{1+\cos(\pi u)}\\
&= \frac{2\sin\left(\frac{\pi u}{2}\right)\cos\left(\frac{\pi u}{2}\right)}{2\cos^2\left(\frac{\pi u}{2}\right)}\\
&= \tan\left(\frac{\pi u}{2}\right)
\end{aligned}$$

Thus, we are led to the conclusion in the statement. □

The above result is very nice, but when it comes to physics, things do not work, for instance because of the wrong slope of the function $\varphi(u)=\tan\left(\frac{\pi u}{2}\right)$ at the origin, which makes our summing on $[-1,1]$ not compatible with the Galileo addition, at low speeds.

So, what to do? Obviously, trash Proposition 13.3, and start all over again. Getting back now to Puzzle 13.2, this has in fact a simpler solution, based this time on algebra, and which in addition is the good, physically correct solution, as follows:

THEOREM 13.4. *If we sum the speeds according to the Einstein formula*

$$u+_e v=\frac{u+v}{1+uv}$$

then the Galileo formula still holds, approximately, for low speeds

$$u+_e v\simeq u+v$$

and if we have $u=1$ or $v=1$, the resulting sum is $u+_e v=1$.

PROOF. All this is self-explanatory, and clear from definitions, and with the Einstein formula of $u+_e v$ itself being just an obvious solution to Puzzle 13.2, provided that, importantly, we know 0 geometry, and rely on very basic algebra only. □

So, very nice, problem solved, at least in 1D. But, shall we give up with geometry, and the stereographic projection? Certainly not, let us try to recycle that material. In order to do this, let us recall that the usual trigonometric functions are given by:

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i} \quad , \quad \cos x = \frac{e^{ix} + e^{-ix}}{2} \quad , \quad \tan x = \frac{e^{ix} - e^{-ix}}{i(e^{ix} + e^{-ix})}$$

The point now is that, and you might know this from calculus, the above functions have some natural "hyperbolic" or "imaginary" analogues, constructed as follows:

$$\sinh x = \frac{e^{x} - e^{-x}}{2} \quad , \quad \cosh x = \frac{e^{x} + e^{-x}}{2} \quad , \quad \tanh x = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

But the function on the right, tanh, starts reminding the formula of Einstein addition, from Theorem 13.4. So, we have our idea, and we are led to the following result:

THEOREM 13.5. *The Einstein speed summation in 1D is given by*

$$\tanh x +_e \tanh y = \tanh(x+y)$$

with $\tanh : [-\infty, \infty] \to [-1, 1]$ *being the hyperbolic tangent function.*

PROOF. This follows by putting together our various formulae above, but it is perhaps better, for clarity, to prove this directly. Our claim is that we have:

$$\tanh(x+y) = \frac{\tanh x + \tanh y}{1 + \tanh x \tanh y}$$

But this can be checked via direct computation, from the definitions, as follows:

$$\begin{aligned}
& \frac{\tanh x + \tanh y}{1 + \tanh x \tanh y} \\
= \ & \left(\frac{e^x - e^{-x}}{e^x + e^{-x}} + \frac{e^y - e^{-y}}{e^y + e^{-y}} \right) \Big/ \left(1 + \frac{e^x - e^{-x}}{e^x + e^{-x}} \cdot \frac{e^y - e^{-y}}{e^y + e^{-y}} \right) \\
= \ & \frac{(e^x - e^{-x})(e^y + e^{-y}) + (e^x + e^{-x})(e^y - e^{-y})}{(e^x + e^{-x})(e^y + e^{-y}) + (e^x - e^{-x})(e^y + e^{-y})} \\
= \ & \frac{2(e^{x+y} - e^{-x-y})}{2(e^{x+y} + e^{-x-y})} \\
= \ & \tanh(x+y)
\end{aligned}$$

Thus, we are led to the conclusion in the statement. ☐

Very nice all this, hope you agree. As a conclusion, passing from the Riemann stereographic projection sum to the Einstein summation basically amounts in replacing:

$$\tan \to \tanh$$

Let us formulate as well this finding more philosophically, as follows:

CONCLUSION 13.6. *The Einstein speed summation in 1D is the imaginary analogue of the summation on $[-1,1]$ obtained via Riemann's stereographic projection.*

Getting now to several dimensions, we have an obvious analogue of Puzzle 13.2 here, and after doing the math, we are led to the following conclusion:

THEOREM 13.7. *When defining the Einstein speed summation in* 3D *as*

$$u+_e v=\frac{1}{1+<u,v>}\left(u+v+\frac{u\times(u\times v)}{1+\sqrt{1-||u||^2}}\right)$$

in $c=1$ units, the following happen:

(1) *When $u\sim v$, we recover the previous* 1D *formula.*
(2) *We have $||u||,||v||<1\implies ||u+_e v||<1$.*
(3) *When $||u||=1$, we have $u+_e v=u$.*
(4) *When $||v||=1$, we have $||u+_e v||=1$.*
(5) *However, $||v||=1$ does not imply $u+_e v=v$.*
(6) *Also, the formula $u+_e v=v+_e u$ fails.*

In addition, the above formula is physically correct, agreeing with experiments.

PROOF. This is something quite tricky, with the key physics claim at the end being indeed true, and with the idea with the mathematical part being as follows:

(1) This is something which follows from definitions.

(2) In order to simplify notation, let us set $\delta=\sqrt{1-||u||^2}$, which is the inverse of the quantity $\gamma=1/\sqrt{1-||u||^2}$. With this convention, we have:

$$\begin{aligned}u+_e v&=\frac{1}{1+<u,v>}\left(u+v+\frac{<u,v>u-||u||^2v}{1+\delta}\right)\\&=\frac{(1+\delta+<u,v>)u+(1+\delta-||u||^2)v}{(1+<u,v>)(1+\delta)}\end{aligned}$$

Taking now the squared norm and computing gives the following formula:

$$||u+_e v||^2=\frac{(1+\delta)^2||u+v||^2+(||u||^2-2(1+\delta))(||u||^2||v||^2-<u,v>^2)}{(1+<u,v>)^2(1+\delta)^2}$$

But this formula can be further processed by using $\delta=\sqrt{1-||u||^2}$, and by navigating through the various quantities which appear, we obtain, as a final product:

$$||u+_e v||^2=\frac{||u+v||^2-||u||^2||v||^2+<u,v>^2}{(1+<u,v>)^2}$$

But this type of formula is exactly what we need, for what we want to do. Indeed, by assuming $||u||,||v||<1$, we have the following estimate:

$$\begin{aligned}||u+_e v||^2<1 &\iff ||u+v||^2-||u||^2||v||^2+<u,v>^2<(1+<u,v>)^2\\ &\iff ||u+v||^2-||u||^2||v||^2<1+2<u,v>\\ &\iff ||u||^2+||v||^2-||u||^2||v||^2<1\\ &\iff (1-||u||^2)(1-||v||^2)>0\end{aligned}$$

Thus, we are led to the conclusion in the statement.

(3) This is something elementary, coming from definitions.

(4) This comes from the squared norm formula established in the proof of (2) above, because when assuming $||v||=1$, we obtain:

$$\begin{aligned}||u+_e v||^2 &= \frac{||u+v||^2-||u||^2+<u,v>^2}{(1+<u,v>)^2}\\ &= \frac{||u||^2+1+2<u,v>-||u||^2+<u,v>^2}{(1+<u,v>)^2}\\ &= \frac{1+2<u,v>+<u,v>^2}{(1+<u,v>)^2}\\ &= 1\end{aligned}$$

(5) This is clear, from the obvious lack of symmetry of our formula.

(6) This is again clear, from the obvious lack of symmetry of our formula. □

Time now to draw some concrete conclusions, from the above speed computations. Since speed $v=d/t$ is distance over time, we must fine-tune distance d, or time t, or both. Let us first discuss, following as usual Einstein, what happens to time t. Here the result, which might seem quite surprising, at a first glance, is as follows:

THEOREM 13.8. *Relativistic time is subject to Lorentz dilation*

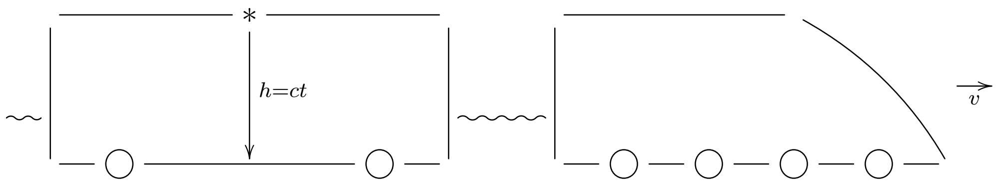
$$t\to\gamma t$$

where the number $\gamma\geq 1$, called Lorentz factor, is given by the formula

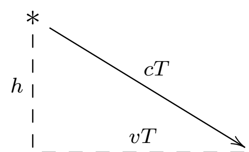
$$\gamma=\frac{1}{\sqrt{1-v^2/c^2}}$$

with v being the moving speed, at which time is measured.

PROOF. Assume indeed that we have a train, moving to the right with speed v, through vacuum. In order to compute the height h of the train, the passenger onboard switches on the ceiling light bulb, measures the time t that the light needs to hit the floor, by traveling at speed c, and concludes that the train height is $h = ct$:



On the other hand, an observer on the ground will see here something different, namely a right triangle, with on the vertical the height of the train h, on the horizontal the distance vT that the train has traveled, and on the hypotenuse the distance cT that light has traveled, with T being the duration of the event, according to his watch:



Now by Pythagoras applied to this triangle, we have the following formula:

$$h^2 + (vT)^2 = (cT)^2$$

Thus, the observer on the ground will reach to the following formula for h:

$$h = \sqrt{c^2 - v^2} \cdot T$$

But h must be the same for both observers, so we have the following formula:

$$\sqrt{c^2 - v^2} \cdot T = ct$$

It follows that the two times t and T are indeed not equal, and are related by:

$$T = \frac{ct}{\sqrt{c^2 - v^2}} = \frac{t}{\sqrt{1 - v^2/c^2}} = \gamma t$$

Thus, we are led to the formula in the statement. □

Let us discuss now what happens to length. We have here the following result:

THEOREM 13.9. *Relativistic length is subject to Lorentz contraction*

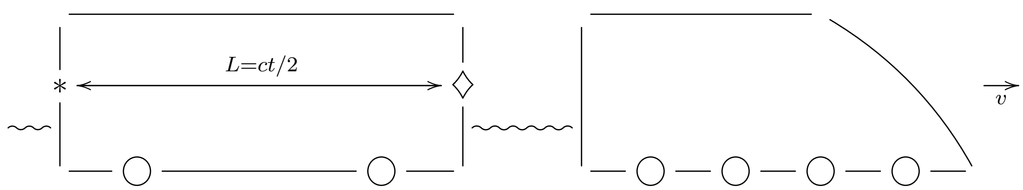
$$L \to L/\gamma$$

where the number $\gamma \geq 1$*, called Lorentz factor, is given by the usual formula*

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}$$

with v *being the moving speed, at which length is measured.*

Proof. As before in the proof of Theorem 13.8, meaning in the same train traveling at speed v, in vacuum, imagine now that the passenger wants to measure the length L of the car. For this purpose he switches on the light bulb, now at the rear of the car, and measures the time t needed for the light to reach the front of the car, and get reflected back by a mirror installed there, according to the following scheme:



He concludes that, as marked above, the length L of the car is given by:

$$L = \frac{ct}{2}$$

Now viewed from the ground, the duration of the event is $T = T_1 + T_2$, where $T_1 > T_2$ are respectively the time needed for the light to travel forward, among others for beating v, and the time for the light to travel back, helped this time by v. More precisely, if l denotes the length of the train car viewed from the ground, the formula of T is:

$$T = T_1 + T_2 = \frac{l}{c-v} + \frac{l}{c+v} = \frac{2lc}{c^2 - v^2}$$

With this data, the formula $T = \gamma t$ of time dilation established before reads:

$$\frac{2lc}{c^2 - v^2} = \gamma t = \frac{2\gamma L}{c}$$

Thus, the two lengths L and l are indeed not equal, and related by:

$$l = \frac{\gamma L(c^2 - v^2)}{c^2} = \gamma L\left(1 - \frac{v^2}{c^2}\right) = \frac{\gamma L}{\gamma^2} = \frac{L}{\gamma}$$

Thus, we are led to the conclusion in the statement. □

With the above discussed, time now to get into the real thing, namely happens to our usual $\mathbb{R}^4$. The result here, which is something quite tricky, is as follows:

Theorem 13.10. *In the context of a relativistic object moving with speed v along the x axis, the frame change is given by the Lorentz transformation*

$$x' = \gamma(x - vt)$$

$$y' = y$$

$$z' = z$$

$$t' = \gamma(t - vx/c^2)$$

with $\gamma = 1/\sqrt{1 - v^2/c^2}$ being as usual the Lorentz factor.

PROOF. We know that, with respect to the non-relativistic formulae, x is subject to the Lorentz dilation by γ, and we obtain as desired:

$$x' = \gamma(x - vt)$$

Regarding y, z, these are obviously unchanged, so done with these too. Finally, for t we must use the reverse Lorentz transformation, given by the following formulae:

$$x = \gamma(x' + vt')$$
$$y = y'$$
$$z = z'$$

By using the formula of x' we can compute t', and we obtain the following formula:

$$\begin{aligned} t' &= \frac{x - \gamma x'}{\gamma v} \\ &= \frac{x - \gamma^2(x - vt)}{\gamma v} \\ &= \frac{\gamma^2 vt + (1 - \gamma^2)x}{\gamma v} \end{aligned}$$

On the other hand, we have the following computation:

$$\gamma^2 = \frac{c^2}{c^2 - v^2} \implies \gamma^2(c^2 - v^2) = c^2 \implies (\gamma^2 - 1)c^2 = \gamma^2 v^2$$

Thus we can finish the computation of t' as follows:

$$\begin{aligned} t' &= \frac{\gamma^2 vt + (1 - \gamma^2)x}{\gamma v} \\ &= \frac{\gamma^2 vt - \gamma^2 v^2 x/c^2}{\gamma v} \\ &= \gamma\left(t - \frac{vx}{c^2}\right) \end{aligned}$$

We are therefore led to the conclusion in the statement. □

Now since y, z are irrelevant, we can put them at the end, and put the time t first, as to be close to x. By multiplying as well the time equation by c, our system becomes:

$$ct' = \gamma(ct - vx/c)$$
$$x' = \gamma(x - vt)$$
$$y' = y$$
$$z' = z$$

In linear algebra terms, the result is as follows:

THEOREM 13.11. *The Lorentz transformation is given by*

$$\begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix}$$

where $\gamma = 1/\sqrt{1 - v^2/c^2}$ *as usual, and where* $\beta = v/c$.

PROOF. In terms of $\beta = v/c$, replacing v, the system looks as follows:

$$ct' = \gamma(ct - \beta x)$$
$$x' = \gamma(x - \beta ct)$$
$$y' = y$$
$$z' = z$$

But this gives the formula in the statement. □

As a nice and basic theoretical application of the Lorentz transform, this brings a new viewpoint on the Einstein speed addition formula, the result being follows:

THEOREM 13.12. *The speed addition formula in* 3*D relativity is*

$$u +_e v = \frac{1}{1 + < u, v >}\left(u + v + \frac{u \times (u \times v)}{1 + \sqrt{1 - ||u||^2}}\right)$$

in $c = 1$ *units.*

PROOF. We already know this, but the point is that we can derive this as well from the formula of the Lorentz transform, by computing some derivatives, as follows:

(1) The idea will be that of differentiating x, y, z, t in the formulae for the inverse Lorentz transform, which are as follows:

$$x = \gamma(x' + ut')$$
$$y = y'$$
$$z = z'$$
$$t = \gamma(t' + ux'/c^2)$$

(2) Indeed, by differentiating these equalities, we obtain the following formulae:

$$dx = \gamma(dx' + udt')$$
$$dy = dy'$$
$$dz = dz'$$
$$dt = \gamma(dt' + udx'/c^2)$$

(3) Now by dividing the first three formulae by the fourth one, we obtain:

$$\frac{dx}{dt} = \frac{dx' + udt'}{dt' + udx'/c^2}$$

$$\frac{dy}{dt} = \frac{dy'}{\gamma(dt' + udx'/c^2)}$$

$$\frac{dz}{dt} = \frac{dz'}{\gamma(dt' + udx'/c^2)}$$

(4) We can make these look better by dividing everywhere by dt', and we get:

$$\frac{dx}{dt} = \frac{dx'/dt' + u}{1 + u/c^2 \cdot dx'/dt'}$$

$$\frac{dy}{dt} = \frac{dy'/dt'}{\gamma(1 + u/c^2 \cdot dx'/dt')}$$

$$\frac{dz}{dt} = \frac{dz'/dt'}{\gamma(1 + u/c^2 \cdot dx'/dt')}$$

(5) In terms of speeds now, this means that we have, with $w = u +_e v$:

$$w_x = \frac{v_x + u}{1 + u/c^2 \cdot v_x}$$

$$w_y = \frac{v_y}{\gamma(1 + u/c^2 \cdot v_x)}$$

$$w_z = \frac{v_z}{\gamma(1 + u/c^2 \cdot v_x)}$$

(6) Now in $c = 1$ units, these formulae are as follows, with $w = u +_e v$:

$$w_x = \frac{v_x + u}{1 + uv_x}$$

$$w_y = \frac{v_y}{\gamma(1 + uv_x)}$$

$$w_z = \frac{v_z}{\gamma(1 + uv_x)}$$

(7) In vector notation now, the above formulae show that we have:

$$u +_e v = \frac{1}{1 + < u, v >} \left(u + \begin{pmatrix} v_x \\ v_y/\gamma \\ v_z/\gamma \end{pmatrix} \right)$$

(8) On the other hand, we have the following formula, coming from definitions:

$$u \times (u \times v) = \begin{pmatrix} 0 \\ -u^2 v_y \\ -u^2 v_z \end{pmatrix}$$

(9) We deduce from this that we have the following formula:

$$v+\frac{u\times(u\times v)}{1+\sqrt{1-u^2}}=\begin{pmatrix}v_x\\ v_y/\gamma\\ v_z/\gamma\end{pmatrix}$$

(10) Summarizing, we have recovered the formula for speed addition in relativity, from before, in our present configuration, with u assumed to be along the Ox axis. But with this, the result in general follows too, either by decomposing one speed vector with respect to the other, or simply by arguing that everything is rotation invariant. □

Recall now that in the non-relativistic setting two events are separated by space Δx and time Δt, with these two separation variables being independent. In relativistic physics this is no longer true, and the correct analogue of this comes from:

THEOREM 13.13. *The following quantity, called relativistic spacetime separation*

$$\Delta s^2=c^2\Delta t^2-(\Delta x^2+\Delta y^2+\Delta z^2)$$

is invariant under relativistic frame changes.

PROOF. This is something important, and as before with such things, we will take our time, and carefully understand how this result works:

(1) Let us first examine the case of the standard configuration. We must prove that the quantity $K=c^2t^2-x^2-y^2-z^2$ is invariant under Lorentz transformations, in the standard configuration. For this purpose, observe that we have:

$$K=\left\langle\begin{pmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&-1\end{pmatrix}\begin{pmatrix}ct\\x\\y\\z\end{pmatrix},\begin{pmatrix}ct\\x\\y\\z\end{pmatrix}\right\rangle$$

Now recall that the Lorentz transformation is given in standard configuration by the following formula, where $\gamma=1/\sqrt{1-v^2/c^2}$ as usual, and where $\beta=v/c$:

$$\begin{pmatrix}ct'\\x'\\y'\\z'\end{pmatrix}=\begin{pmatrix}\gamma&-\beta\gamma&0&0\\-\beta\gamma&\gamma&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\begin{pmatrix}ct\\x\\y\\z\end{pmatrix}$$

Thus, if we denote by L the matrix of the Lorentz transformation, and by E the matrix found before, we must prove that for any vector ξ we have:

$$< E\xi,\xi >=< EL\xi,L\xi >$$

Since the matrix L is symmetric, we have the following formula:

$$< EL\xi,L\xi >=< LEL\xi,\xi >$$

Thus, we must prove that we have $E = LEL$. But this is the same as proving that we have $L^{-1}E = EL$. Moreover, by using the fact that the passage $L \to L^{-1}$ is given by $\beta \to -\beta$, what we eventually want to prove is that:

$$L_{-\beta}E = EL_\beta$$

So, let us prove this. As usual we can restrict the attention to the upper left corner, call that NW corner, and here we have the following computation:

$$(L_{-\beta}E)_{NW} = \begin{pmatrix}\gamma & \beta\gamma\\ \beta\gamma & \gamma\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} = \begin{pmatrix}\gamma & -\beta\gamma\\ \beta\gamma & -\gamma\end{pmatrix}$$

On the other hand, we have as well the following computation:

$$(EL_\beta)_{NW} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}\gamma & -\beta\gamma\\ -\beta\gamma & \gamma\end{pmatrix} = \begin{pmatrix}\gamma & -\beta\gamma\\ \beta\gamma & -\gamma\end{pmatrix}$$

The matrices on the right being equal, this gives the result.

(2) Now let us prove the invariance in general. The matrix of the Lorentz transformation being a bit complicated, in this case, as explained in the above, the best is to use for this the raw formulae of x', t', that we found in the above, namely:

$$x' = x + (\gamma - 1)\frac{< v, x > v}{||v||^2} - \gamma t v$$

$$t' = \gamma\left(t - \frac{< v, x >}{c^2}\right)$$

With these formulae in hand, we have the following computation:

$$\begin{aligned}
& (ct')^2 - ||x'||^2 \\
= \ & \gamma^2\left(ct - \frac{< v, x >}{c}\right)^2 \\
& -||x||^2 - (\gamma - 1)^2\frac{< v, x >^2}{||v||^2} - \gamma^2 t^2 ||v||^2 \\
& -2(\gamma - 1)\frac{< v, x >^2}{||v||^2} + 2\gamma t < v, x > +2\gamma(\gamma - 1)t < v, x > \\
= \ & \gamma^2 t^2 (c^2 - ||v||^2) - ||x||^2 \\
& + < v, x > (-2\gamma^2 t + 2\gamma t + 2\gamma(\gamma - 1)t) \\
& + < v, x >^2 \left(\frac{\gamma^2}{c^2} - \frac{(\gamma - 1)^2}{||v||^2} - \frac{2(\gamma - 1)}{||v||^2}\right) \\
= \ & c^2 t^2 - ||x||^2 + < v, x >^2 \left(\frac{\gamma^2}{c^2} - \frac{\gamma^2 - 1}{||v||^2}\right) \\
= \ & c^2 t^2 - ||x||^2
\end{aligned}$$

Here we have used the following trivial formula, for the coefficient of t^2:

$$\gamma^2(c^2-||v||^2)=\frac{c^2-||v||^2}{1-||v||^2/c^2}=c^2$$

Also, we have used the following formula, for the coefficient of $< v,x >^2$:

$$\begin{aligned}\frac{\gamma^2}{c^2}-\frac{\gamma^2-1}{||v||^2} &= \gamma^2\left(\frac{1}{c^2}-\frac{1}{||v||^2}\right)+\frac{1}{||v||^2}\\ &= \frac{1}{1-||v||^2/c^2}\cdot\frac{||v||^2-c^2}{c^2||v||^2}+\frac{1}{||v||^2}\\ &= \frac{c^2}{c^2-||v||^2}\cdot\frac{||v||^2-c^2}{c^2||v||^2}+\frac{1}{||v||^2}\\ &= -\frac{1}{||v||^2}+\frac{1}{||v||^2}\\ &= 0\end{aligned}$$

Thus, we are led to the conclusion in the statement. □

13b.

13c.

13d.

13e. Exercises

Exercises:

Exercise 13.14.

Exercise 13.15.

Exercise 13.16.

Exercise 13.17.

Exercise 13.18.

Exercise 13.19.

Exercise 13.20.

Exercise 13.21.

Bonus exercise.

CHAPTER 14

Lorentz manifolds

14a. Lorentz manifolds

14b.

14c.

14d.

14e. Exercises

Exercises:

Exercise 14.1.

Exercise 14.2.

Exercise 14.3.

Exercise 14.4.

Exercise 14.5.

Exercise 14.6.

Exercise 14.7.

Exercise 14.8.

Bonus exercise.

CHAPTER 15

Curved spacetime

15a. Curved spacetime

15b.

15c.

15d.

15e. Exercises

Exercises:

Exercise 15.1.

Exercise 15.2.

Exercise 15.3.

Exercise 15.4.

Exercise 15.5.

Exercise 15.6.

Exercise 15.7.

Exercise 15.8.

Bonus exercise.

CHAPTER 16

Advanced aspects

16a. Advanced aspects

16b.

16c.

16d.

16e. Exercises

Congratulations for having read this book, and no exercises for this final chapter.

Bibliography

[1] V.I. Arnold, Ordinary differential equations, Springer (1973).

[2] V.I. Arnold, Mathematical methods of classical mechanics, Springer (1974).

[3] V.I. Arnold, Lectures on partial differential equations, Springer (1997).

[4] V.I. Arnold, Catastrophe theory, Springer (1974).

[5] V.I. Arnold and B.A. Khesin, Topological methods in hydrodynamics, Springer (1998).

[6] M.F. Atiyah, K-theory, CRC Press (1964).

[7] M.F. Atiyah, The geometry and physics of knots, Cambridge Univ. Press (1990).

[8] M.F. Atiyah and I.G. MacDonald, Introduction to commutative algebra, Addison-Wesley (1969).

[9] T. Banica, Calculus and applications (2024).

[10] T. Banica, Linear algebra and group theory (2024).

[11] T. Banica, Introduction to modern physics (2025).

[12] R.J. Baxter, Exactly solved models in statistical mechanics, Academic Press (1982).

[13] N. Berline, E. Getzler and M. Vergne, Heat kernels and Dirac operators, Springer (2004).

[14] B. Blackadar, K-theory for operator algebras, Cambridge Univ. Press (1986).

[15] S.J. Blundell and K.M. Blundell, Concepts in thermal physics, Oxford Univ. Press (2006).

[16] S.M. Carroll, Spacetime and geometry, Cambridge Univ. Press (2004).

[17] A.R. Choudhuri, Astrophysics for physicists, Cambridge Univ. Press (2012).

[18] A. Connes, Noncommutative geometry, Academic Press (1994).

[19] A. Connes and M. Marcolli, Noncommutative geometry, quantum fields and motives, AMS (2008).

[20] W.N. Cottingham and D.A. Greenwood, An introduction to the standard model of particle physics, Cambridge Univ. Press (2012).

[21] P.A. Davidson, Introduction to magnetohydrodynamics, Cambridge Univ. Press (2001).

[22] P.A.M. Dirac, Principles of quantum mechanics, Oxford Univ. Press (1930).

[23] M.P. do Carmo, Differential geometry of curves and surfaces, Dover (1976).

[24] M.P. do Carmo, Riemannian geometry, Birkhäuser (1992).

[25] S. Dodelson, Modern cosmology, Academic Press (2003).

[26] S.K. Donaldson, Riemann surfaces, Oxford Univ. Press (2004).

[27] R. Durrett, Probability: theory and examples, Cambridge Univ. Press (1990).

[28] A. Einstein, Relativity: the special and the general theory, Dover (1916).

[29] L.C. Evans, Partial differential equations, AMS (1998).

[30] W. Feller, An introduction to probability theory and its applications, Wiley (1950).

[31] E. Fermi, Thermodynamics, Dover (1937).

[32] R.P. Feynman, R.B. Leighton and M. Sands, The Feynman lectures on physics, Caltech (1963).

[33] R.P. Feynman and A.R. Hibbs, Quantum mechanics and path integrals, Dover (1965).

[34] P. Flajolet and R. Sedgewick, Analytic combinatorics, Cambridge Univ. Press (2009).

[35] A.P. French, Special relativity, Taylor and Francis (1968).

[36] W. Fulton, Algebraic topology, Springer (1995).

[37] W. Fulton and J. Harris, Representation theory, Springer (1991).

[38] C. Godsil and G. Royle, Algebraic graph theory, Springer (2001).

[39] H. Goldstein, C. Safko and J. Poole, Classical mechanics, Addison-Wesley (1980).

[40] M.B. Green, J.H. Schwarz and E. Witten, Superstring theory, Cambridge Univ. Press (2012).

[41] D.J. Griffiths, Introduction to electrodynamics, Cambridge Univ. Press (2017).

[42] D.J. Griffiths and D.F. Schroeter, Introduction to quantum mechanics, Cambridge Univ. Press (2018).

[43] D.J. Griffiths, Introduction to elementary particles, Wiley (2020).

[44] P. Griffiths and J. Harris, Principles of algebraic geometry, Wiley (1994).

[45] A. Grothendieck and J. Dieudonné, Éléments de géométrie algébrique, IHES (1967).

[46] A. Grothendieck et al., Séminaire de géométrie algébrique, IHES (1972).

[47] G.H. Hardy and E.M. Wright, An introduction to the theory of numbers, Oxford Univ. Press (1938).

[48] J. Harris, Algebraic geometry, Springer (1992).

[49] R. Hartshorne, Algebraic geometry, Springer (1977).

[50] A. Hatcher, Algebraic topology, Cambridge Univ. Press (2002).

[51] H. Hofer and E. Zehnder, Symplectic invariants and Hamiltonian dynamics, Birkhäuser (1994).

[52] L. Hörmander, The analysis of linear partial differential operators, Springer (1983).

[53] R.A. Horn and C.R. Johnson, Matrix analysis, Cambridge Univ. Press (1985).

[54] K. Huang, Introduction to statistical physics, CRC Press (2001).

[55] J.E. Humphreys, Introduction to Lie algebras and representation theory, Springer (1972).

[56] J.E. Humphreys, Linear algebraic groups, Springer (1975).

[57] K. Ireland and M. Rosen, A classical introduction to modern number theory, Springer (1982).

[58] N. Jacobson, Basic algebra, Dover (1974).

[59] V.F.R. Jones, Index for subfactors, *Invent. Math.* **72** (1983), 1–25.

[60] V.F.R. Jones, A polynomial invariant for knots via von Neumann algebras, *Bull. Amer. Math. Soc.* **12** (1985), 103–111.

[61] V.F.R. Jones, Hecke algebra representations of braid groups and link polynomials, *Ann. of Math.* **126** (1987), 335–388.

[62] V.F.R. Jones, On knot invariants related to some statistical mechanical models, *Pacific J. Math.* **137** (1989), 311–334.

[63] V.F.R. Jones, Planar algebras I (1999).

[64] M. Karoubi, K-theory: an introduction, Springer (1978).

[65] T. Kibble and F.H. Berkshire, Classical mechanics, Imperial College Press (1966).

[66] T. Lancaster and K.M. Blundell, Quantum field theory for the gifted amateur, Oxford Univ. Press (2014).

[67] L.D. Landau and E.M. Lifshitz, Course of theoretical physics, Pergamon Press (1960).

[68] S. Lang, Algebra, Addison-Wesley (1993).

[69] S. Lang, Abelian varieties, Dover (1959).

[70] P. Lax, Linear algebra and its applications, Wiley (2007).

[71] P. Lax, Functional analysis, Wiley (2002).

[72] J.M. Lee, Introduction to topological manifolds, Springer (2011).

[73] J.M. Lee, Introduction to smooth manifolds, Springer (2012).

[74] J.M. Lee, Introduction to Riemannian manifolds, Springer (2019).

[75] D. McDuff and D. Salamon, Introduction to symplectic topology, Oxford Univ. Press (2017).

[76] P. Petersen, Linear algebra, Springer (2012).

[77] P. Petersen, Riemannian geometry, Springer (2006).

[78] W. Rudin, Principles of mathematical analysis, McGraw-Hill (1964).

[79] W. Rudin, Real and complex analysis, McGraw-Hill (1966).

[80] W. Rudin, Fourier analysis on groups, Dover (1974).

[81] B. Ryden, Introduction to cosmology, Cambridge Univ. Press (2002).

[82] B. Ryden and B.M. Peterson, Foundations of astrophysics, Cambridge Univ. Press (2010).

[83] W. Schlag, A course in complex analysis and Riemann surfaces, AMS (2014).

[84] D.V. Schroeder, An introduction to thermal physics, Oxford Univ. Press (1999).

[85] J.P. Serre, A course in arithmetic, Springer (1973).

[86] J.P. Serre, Linear representations of finite groups, Springer (1977).

[87] I.R. Shafarevich, Basic algebraic geometry, Springer (1974).

[88] J.H. Silverman, The arithmetic of elliptic curves, Springer (1986).

[89] J.H. Silverman and J.T. Tate, Rational points on elliptic curves, Springer (2015).

[90] B. Singh, Basic commutative algebra, World Scientific (2011).

[91] C.H. Taubes, Differential geometry, Oxford Univ. Press (2011).

[92] J.R. Taylor, Classical mechanics, Univ. Science Books (2003).

[93] J. von Neumann, Mathematical foundations of quantum mechanics, Princeton Univ. Press (1955).

[94] S. Weinberg, Foundations of modern physics, Cambridge Univ. Press (2011).

[95] S. Weinberg, Lectures on quantum mechanics, Cambridge Univ. Press (2012).

[96] S. Weinberg, Lectures on astrophysics, Cambridge Univ. Press (2019).

[97] H. Weyl, The theory of groups and quantum mechanics, Princeton Univ. Press (1931).

[98] H. Weyl, The classical groups: their invariants and representations, Princeton Univ. Press (1939).

[99] H. Weyl, Space, time, matter, Princeton Univ. Press (1918).

[100] B. Zwiebach, A first course in string theory, Cambridge Univ. Press (2004).

Index