Jean-Baptiste Hiriart-Urruty

Mathematical Tapas

Volume 2
(From Undergraduate to Graduate Level)

Springer

Jean-Baptiste Hiriart-Urruty
Institute of Mathematics
Paul Sabatier University
Toulouse
France

ISSN 1615-2085 ISSN 2197-4144 (electronic)
Springer Undergraduate Mathematics Series
ISBN 978-3-319-68630-1 ISBN 978-3-319-68631-8 (eBook)
https://doi.org/10.1007/978-3-319-68631-8

Library of Congress Control Number: 2017957677

Mathematics Subject Classification (2010): 00-01, 00A07

Foreword

Mathematical tapas... but what are tapas? *Tapas* is a Spanish word (in the Basque country, one would also say *pintxos*) for small savory dishes typically served in bars, with drinks, shared with friends in a relaxed ambience. The offer is varied: it may be meat, fish, vegetables,... Each of the guests of the party selects the tapas he likes most at the moment. This is the spirit of the mathematical tapas that we present to the reader here.

The tapas that we offer are of the same character as in our previous volume, published with the same title, but which was aimed at the undergraduate level.[1] However, the tapas in this volume are more substantial; they therefore could be called *raciones* instead. Consisting of mathematical questions to answer – exercises (more than long problems, in their spirit) of various types – these raciones concern mathematics ranging from the undergraduate to the graduate level (roughly speaking, this corresponds to the end of the third year and the fourth year at university);[2] they do not cover the whole spectrum of mathematics, of course. Clearly, they reflect the mathematical interests and teaching experience of the author:

Metric, normed, Banach**, inner-product and** Hilbert **spaces**:
basic calculus with distances and norms, convergent or Cauchy sequences, oddities in the infinite-dimensional setting;
Differential calculus:
calculus rules, applications to unconstrained optimization problems in various settings;
Integration:
examples of effective calculations of integrals, nothing on the theoretical aspects;
Matrices:
especially symmetric and positive semidefinite, links with quadratic functions, interplays with geometry, convexity and optimization;
Convexity:
convex sets, convex functions, their use in optimization;
Optimization or "variational" problems:
arising in geometry (triangles or polyhedrons), unconstrained or constrained (mainly with equality constraints).

To reflect the variety of mathematics, there is no specific ordering of topics: the tapas are more or less "randomly" presented, even if some gatherings have been carried out (for example, on mean value theorems, on weakly converging sequences in Hilbert spaces, etc.)

[1] J.-B. Hiriart-Urruty, *Mathematical tapas*. Vol. 1 (for Undergraduates). Springer Undergraduate Mathematics Series (September 2016).

[2] Called "Licence 3" and "Master 1" in the European Higher Education system.

How have they been chosen?

- Firstly, because "I like them" and have tested them. In other words, each tapa *reveals something*: it could be an interesting inequality among integrals, a useful or surprising property of some mathematical objects (especially in the infinite-dimensional setting), or simply an elegant formula... I am just sensitive to the aesthetics of mathematics.

- Secondly, because they illustrate the following motto: *"if you solve it, you learn something".* During my career, I have taught hundreds of students and, therefore, posed thousands of exercises (in directed sessions of problem solving, for exams, etc.); but I have not included here (standard) questions whose objective is just to test the ability to calculate a gradient, sums of series or integrals, eigenvalues, etc. I have therefore limited my choice of tapas for this second volume to the (symbolic) number of **222**.

Where have they been chosen from?

I have observed that, year after year, some questions give rise to interest or surprise among students. These mathematical tapas are chosen from them, and also from my favorite journals posing such challenges: the *American Mathematical Monthly*, and the French mathematical journals *Revue de la Filière Mathématique* (formerly *Revue de Mathématiques Spéciales*) and *Quadrature.* From time to time, I have posed or solved questions posted in these journals. Some longer or more substantial tapas have been reconstituted from those already present, in a similar form, in the books that I have written in French in the past (see the references at the end). However, for many tapas, I must confess that I could not remember their origin or history.

How are they classified?

As in restaurant guides, each tapa has one, two or three stars (★):

- One star (★). Tapas of the first level, for students at the end of their undergraduate studies.

- Two stars (★★). Tapas of a more advanced level. That does not mean that solving them necessarily requires more expertise or wit than for one-starred tapas, but sometimes just more maturity (or prerequisite definitions) in mathematics.

- Three stars (★★★). The upper level in the proposed tapas, typically for students in the first years of their graduate studies. Some may be tough and need more chewing.

We admit that this classification is somewhat arbitrary for some of the problems, as it depends on the reader's background.

How are they presented?

Each tapa begins with a *statement*, of course. The statement may contain the answers to the posed questions; this is the case when the questions or proposals are formulated as "Show that... " or "Prove that".

There are no detailed solutions to all questions, as that would have inflated this booklet by a factor of three or four. Moreover, in mathematics, there is no uniform and unique way to write down answers. But, to help solve the posed challenges, I have proposed *hints*... They suggest a path to follow. A

question without any indication could be labelled "can't be done" or too time-consuming... ; the same question with "spoon-fed" steps could be considered too easy. We have tried to strike a balance between the two postures which reflects the variety of the tapas. Of course, an interested reader is asked to try to chew and swallow the tapas without having recourse to the hints.

When they are not integrated into the statements, we provide *answers* to questions, numerical results for example.

From time to time, we add some *comments*: on the context of the question, on its origin, on a possible extension.

Not all questions in mathematics have known answers... To illustrate this, at the end we have added 8 open problems or conjectures. This is our *open bar* section... the reader helps himself. Of course, there are neither hints nor answers for them since they are unknown... I have chosen these open problems to match the level and topics considered in this volume (numbers, real analysis, matrices, optimization,...): easy to understand, concerning various areas of mathematics, original for some of them. They are marked with the symbol (♣♣♣).

In spite of my efforts, some misprints or even mistakes may have slipped in; I just hope that they are not irreparable.

An essential characteristic of mathematics is to be universal and thus international. So, imagine a student or someone who has some knowledge in mathematics (say, undergraduate level) in seclusion for some time on an isolated island, or just put into jail... With a book like the one containing these tapas, he might even enjoy his time and savour some of them.

Bon appétit !

J.-B. Hiriart-Urruty (JBHU)
Toulouse and the Basque country (2015–2016).

Notations

All the notations, abbreviations and appellations we use are standard; however, here we make some of them more precise.

vs, abbreviation of *versus:* in opposition with, or faced with.
i.e., abbreviation of *id est:* that is to say.

$\mathbb{N}$: the set of natural integers, that is, $\{0, 1, 2...\}$.
x positive means $x > 0$; x nonnegative means $x \geqslant 0$. This varies from country to country, sometimes positive is used for nonnegative and strictly positive for positive. Here, we stand by the first appellations.
$f : I \to \mathbb{R}$ is increasing on the interval I means that $f(x) \leqslant f(y)$ whenever $x < y$ in I; to call such functions nondecreasing is a mistake from the logical viewpoint.
$f : I \to \mathbb{R}$ is strictly increasing on the interval I means that $f(x) < f(y)$ whenever $x < y$ in I.
$[a, b]$ denotes the closed interval of $\mathbb{R}$ with end-points a and b.
(a, b) is a somewhat ambiguous notation used to denote the open interval with end-points a and b; in some countries the (better) notation $]a, b[$ is used instead.
log or ln: used indifferently for the natural (or Napierian) logarithm.
$\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$; also denoted C_n^k in some countries.

For two vectors $u = (u_1, ... , u_n)$ and $v = (v_1, ... , v_n)$ in $\mathbb{R}^n$, $u \leqslant v$ means componentwise inequality: $u_i \leqslant v_i$ for all $i = 1, ... , n$.
$\|\cdot\|$: unless otherwise specified, this denotes the usual Euclidean norm in $\mathbb{R}^n$.
tr$A$ or trace(A) stands for the trace of A.
det$A$ stands for the determinant of A.
$\mathcal{S}_n(\mathbb{R})$ is the set of $n \times n$ real symmetric matrices. Semidefinite and positive definite matrices are always symmetric; this is often recalled, and assumed if not.
$A \succcurlyeq 0$ (resp. $A \succ 0$) means that the (symmetric) matrix A is positive semidefinite (resp. positive definite).

$\langle \cdot, \cdot \rangle$ is the generic notation for an inner-product (or scalar product). However, for the usual inner product of two vectors u and v in $\mathbb{R}^d$, we also use the notation $u^T v$ for $\langle u, v \rangle$; for example, the quadratic form on $\mathbb{R}^d$ associated with $A \in \mathcal{S}_d(\mathbb{R})$ is denoted $\langle Ax, x \rangle$ or $x^T Ax$. Moreover, to make a distinction with the usual Euclidean space $\left(\mathbb{R}^d, \langle \cdot, \cdot \rangle\right)$, we make use of the notation $\langle\langle U, V \rangle\rangle = \text{tr}(U^T V)$ for the standard scalar product in $\mathcal{M}_{m,n}(\mathbb{R})$.
$f : H \to \mathbb{R}$ is a primitive or an anti-gradient of the mapping $g : H \to H$ if f is differentiable and $\nabla f = g$.

A point x is called critical (or stationary) for the differentiable function $f : H \to \mathbb{R}$ whenever $\nabla f(x) = 0$.

If $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at x, $\nabla^2 f(x)$ or $Hf(x)$ stands for the Hessian matrix of f at x, *i.e.*, the $n \times n$ symmetric matrix whose entries are the second-order partial derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$.

$\mathrm{co}(A)$: the convex hull of A, *i.e.,* the smallest convex set containing A.

$\overline{\mathrm{co}}(A)$: the closed convex hull of A, *i.e.,* the smallest closed convex set containing A, which is also the closure of $\mathrm{co}(A)$.

$\mathbf{1}_S$: the indicator function of $S \subset E$. By definition, $\mathbf{1}_S(x) = 1$ if $x \in S$, $\mathbf{1}_S(x) = 0$ if $x \notin S$.

Integrals: The notations $\int_a^b f(x)\,\mathrm{d}x$ (the RIEMANN integral of the continuous function f on $[a, b]$) and $\int_{[a,b]} f(x)\,\mathrm{d}\lambda(x)$, or even $\int_{[a,b]} f(x)\,\mathrm{d}x$ (the LEBESGUE integral of the measurable function f on $[a, b]$) cohabit in the text.

Classification by levels of difficulty

1–103: ★
104–189: ★★
190–222: ★★★
223–230: ♣♣♣

Classification by topics

Part 1. One-starred tapas

"What mathematics usually consists of is problems and solutions"
P. Halmos (1916–2006)
"The best way to learn mathematics is to solve problems"
J. Lelong-Ferrand (1918–2014)

1. ★ Pompeiu's *mean value theorem*

   Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). We suppose, moreover, that $0 < a < b$.

   1°) Show that there exists a $c \in (a, b)$ such that

   $$\frac{bf(a) - af(b)}{b - a} = f(c) - cf'(c). \tag{1}$$

   2°) Give a geometrical interpretation of the above result by considering the tangent line to the graph of f at the point $C = (c, f(c))$ and the line passing through the two points $A = (a, f(a))$ and $B = (b, f(b))$.

   ---

   *Hint.*1°) Apply the classical mean value theorem to the function

   $$F : u \in \left[\frac{1}{b}, \frac{1}{a}\right] \mapsto F(u) = uf\left(\frac{1}{u}\right).$$

   Answer. 2°) The tangent line at C intersects the Oy-axis at the same point as the line passing through A and B.

---

2. ★ *The mid-point mean value theorem*

   Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b).

   1°) Show that there exists a $c \in (a, b)$ such that

   $$f'(c)\left[f(c) - \frac{f(a) + f(b)}{2}\right] = -\left[c - \frac{a + b}{2}\right]. \tag{1}$$

   2°) Give a geometrical interpretation of the above result by considering the tangent line to the graph of f at $C = (c, f(c))$ and the line passing through $C = (c, f(c))$ and the mid-point $M = \left(\frac{a+b}{2}, \frac{f(a)+f(b)}{2}\right)$.

   ---

*Hint.*1°) Apply the classical mean value theorem to the function

$$g : x \in [a,b] \mapsto g(x) = (x-a)^2+(x-b)^2+[f(x)-f(a)]^2+[f(x)-f(b)]^2 .$$

Answer. 2°) Either $M = \left(\frac{a+b}{2}, \frac{f(a)+f(b)}{2}\right)$ lies on the graph of f, or the line passing through M and C is perpendicular to the tangent line at C.

---

3. ★ FLETT's *mean value theorem*

Let $f : I \to \mathbb{R}$ be differentiable on an open interval I and let $a < b$ be in I such that $f'(a) = f'(b)$.

1°) Show that there exists a $c \in (a,b)$ such that

$$\frac{f(c)-f(a)}{c-a} = f'(c). \tag{1}$$

2°) Give a geometrical interpretation of the above result by considering the tangent line to the graph of f at the point $C = (c, f(c))$.

Hints. 1°) We may suppose that $f'(a) = f'(b) = 0$, for if this is not the case we can work with the function $x \mapsto f(x) - xf'(a)$. Also, without loss of generality, one may assume that $a = 0, b = 1$ and $f(a) = 0$.

Looking at a graphical representation of f, one realizes that a point c in (1) should be a critical (or stationary) one for the "slope function" s :

$$\begin{cases} x \in [0,1] \mapsto s(x) = f(x)/x \text{ if } x \neq 0, \\ \qquad\qquad\quad s(0) = f'(0). \end{cases}$$

Since

$$s'(x) = \frac{xf'(x)-f(x)}{x^2},\ x \in (0,1],$$

it is sufficient to prove that there is some $c \in (0,1)$ such that $s'(c) = 0$. Then apply ROLLE's theorem by distinguishing three cases:

$$f(1) = 0,\ f(1) > 0,\ f(1) < 0.$$

Answer. 2°) The tangent line at C passes through the point $A = (a, f(a))$.

Comment. If one wishes to get rid of the condition $f'(a) = f'(b)$, the generalized result reads as follows: Let $f : I \to \mathbb{R}$ be differentiable on an open interval I and let $a < b$ be in I; there then exists a $c \in (a, b)$ such that

$$f(c) - f(a) = (c - a)f'(c) - \frac{1}{2}\frac{f'(b) - f'(a)}{b - a}(c - a)^2. \tag{2}$$

---

4. ★ *A mean value theorem for non-differentiable convex functions*

Let $f : I \to \mathbb{R}$ be convex on an open interval I and let $a < b$ be in I.

1°) First form. Show that there exists a $c \in (a, b)$ such that

$$\frac{f(b) - f(a)}{b - a} \in \left[f'_-(c), f'_+(c)\right], \tag{1}$$

where $f'_-(c)$ and $f'_+(c)$ denote the left and right derivatives of f at c.

2°) Second form. Prove that there exist c_1 and c_2 in (a, b) and nonnegative real numbers λ_1 and λ_2, such that:

$$\begin{cases} f \text{ is differentiable at } c_1 \text{ and } c_2; \\ \lambda_1 + \lambda_2 = 1; \\ \frac{f(b)-f(a)}{b-a} = \lambda_1 f'(c_1) + \lambda_2 f'(c_2). \end{cases} \tag{2}$$

Hints. 1°) If $\overline{x}$ minimizes a convex function $\varphi : I \to \mathbb{R}$, then

$$0 \in \left[\varphi'_-(\overline{x}), \varphi'_+(\overline{x})\right].$$

2°) As consequences of the convexity of f :

- The derivative function f' exists almost everywhere; it is increasing whenever it exists.

- At a point x where f is not differentiable, $f'_-(x) \leqslant f'_+(x)$ and the line-segment $\left[f'_-(x), f'_+(x)\right]$ is a substitute for the derivative of f at x; it plays the same role.

---

5. ★ *A mean value theorem for functions having right derivatives*

Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$; we suppose that f has a right derivative $f'_+(x)$ at each point x in (a, b).

1°) Preliminaries. Suppose that $f(a) = f(b) = 0$.

(a) Show that there exists a $c \in (a, b)$ such that $f'_+(c) \leqslant 0$.

(b) Show that there exists a $d \in (a, b)$ such that $f'_+(d) \geqslant 0$.

2°) Prove that there exist two points c and d in (a, b) such that

$$\frac{f(b) - f(a)}{b - a} \in \left[f'_-(c), f'_+(d)\right]. \tag{1}$$

Hints. 1°) (a). Assuming that f is not identically zero, consider the two cases where f achieves either a positive maximum value or a negative minimum value on $[a, b]$.

1°) (b). The assumptions made on f remain in force if one passes to $-f$. Thus, apply here the result of 1°) (a) to $-f$.

2°) As usual in the context of proving mean value theorems, apply the results of 1°) to the "put right" function

$$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a). \tag{2}$$

Comment. As a consequence of the mean value theorem above, we have the following:

Let f be a continuous real-valued function on an interval I. If the right derivative of f exists and equals 0 at each interior point of I, then f is constant on I.

This is a particular case of the kind of results proved by U. Dini (1878); *cf.* the note:

J.-B. Hiriart-Urruty, *Le théorème de* Dini *sur les taux de variation d'une fonction.* Revue de Mathématiques Spéciales, n°9 (1984), 369–370.

---

6. ★ *Two identities involving the 4-th powers of numbers*

All the variables involved in the identities below are real numbers.

1°) (a) Check that

$$a^4 + b^4 = (a^2 + ab\sqrt{2} + b^2) \times (a^2 - ab\sqrt{2} + b^2),$$

which has as a special case

$$a^4 + 4b^4 = \left[(a + b)^2 + b^2\right] \times \left[(a - b)^2 + b^2\right]. \tag{1}$$

(b) Use the identity (1) above to prove that, for any integer $n \geqslant 2$, the integer $n^4 + 4$ cannot be a prime number.

2°) Check that

$$(a+b)^4 - (a-b)^4 = 8ab \times (a^2 + b^2). \tag{2}$$

Hint. The two questions 1°) and 2°) are independent.

Answer. To answer the second part of the first question, we may use the following factorization:

$$n^4 + 4 = (n^2 + 2n + 2) \times (n^2 - 2n + 2).$$

Comment. (1) is known as S. GERMAIN's identity; (2) is known as A.-M. LEGENDRE's identity.

---

7. ★ *Limit of integrals of mean values*

Let $f : [0,1] \to \mathbb{R}$ be a continuous function.

1°) Let n be a positive integer, and let I_n be the following multiple integral:

$$I_n = \int_{[0,1]^n} f\left(\frac{x_1 + x_2 + \ldots x_n}{n}\right) \, \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n.$$

1°) Prove that I_n has a limit when $n \to +\infty$.

2°) We modify I_n slightly by substituting in its formulation the geometric mean value in place of the arithmetic mean value:

$$J_n = \int_{[0,1]^n} f\left(\sqrt[n]{x_1 x_2 \ldots x_n}\right) \, \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n.$$

Prove that J_n also has a limit as $n \to +\infty$.

Hint. Begin by considering $f(x) = x^p$, then polynomial functions f; then use the fact that f can be uniformly approximated on $[0,1]$ by polynomial functions (WEIERSTRASS' theorem).

The difficulty here is that the domain of integration in the proposed integrals moves with n too.

Answers. 1°) Consider firstly $f(x) = x^p$, where p is a nonnegative integer. By developing the integral I_n for such a function we obtain:

$$I_n = \frac{1}{n^p} \sum_{1 \leqslant i_1, \ldots, i_p \leqslant n} \int_{[0,1]^n} x_{i_1} \ldots x_{i_p} \, \mathrm{d}x_1 \ldots \mathrm{d}x_n. \tag{1}$$

There are two parts to be distinguished in the sum in (1) above: one, S_n, is the sum ranging over pairwise distinct indices $i_1, \ldots, i_p$, and a second component, T_n, is the sum ranging over indices with at least two equal among them. Then, we firstly have

$$\begin{aligned} S_n &= p! \sum_{1 \leqslant i_1, \ldots, i_p \leqslant n} \int_{[0,1]^n} x_{i_1} \ldots x_{i_p} \, \mathrm{d}x_1 \ldots \mathrm{d}x_n \\ \text{(by FUBINI's theorem)} &= p! \binom{n}{p} \left(\frac{1}{2}\right)^p, \end{aligned}$$

so that $S_n / n^p \to (1/2)^p = f(1/2)$ as $n \to +\infty$.

Secondly, each term in T_n is bounded from above by 1, and the number of terms in this sum is negligible when compared to n^p. Thus, $T_n / n^p \to 0$ as $n \to +\infty$. Altogether,

$$I_n = \frac{S_n}{n^p} + \frac{T_n}{n^p} \to f\left(\frac{1}{2}\right) \text{ as } n \to +\infty.$$

The result is easily extended to any polynomial function (due to the linearity of the integral), and then to any continuous function via the WEIERSTRASS approximation theorem.

2°) Here also, we begin with $f(x) = x^p$. For such a function f, we obtain

$$J_n = \prod_{i=1}^{n} \left[\int_0^1 x_i^{p/n} \, \mathrm{d}x_i \right] \int_0^1 x_i^{p/n} \, \mathrm{d}x_i = \left(1 + \frac{p}{n}\right)^{-n}. \tag{2}$$

When $n \to +\infty$, J_n converges to $e^{-p} = \left(\frac{1}{e}\right)^p = f(\frac{1}{e})$.

We proceed as in 1°). Finally, for any continuous function f,

$$J_n \to f\left(\frac{1}{e}\right) \text{ as } n \to +\infty.$$

Comment. For those readers who have some knowledge of Probability, the results in this Tapa could be explained (and even proved) via the so-called (strong) law of large numbers; consider independent random variables X_i, each of them uniformly distributed on $[0,1]$, $\frac{X_1+\ldots+X_n}{n}$ which converges almost surely towards $\frac{1}{2}$, and the expectation of $f\left(\frac{X_1+\ldots+X_n}{n}\right)$, which tends to $f\left(\frac{1}{2}\right)$ as $n \to +\infty$.

---

8. ★ *A simple homeomorphism acting on the graph of a continuous mapping*

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a continuous mapping and let G_f be its graph, that is to say:

$$G_f = \{(x, f(x)) \mid x \in \mathbb{R}^n\}.$$

1°) Recall why G_f is a closed connected subset of $\mathbb{R}^n \times \mathbb{R}^m$.

2°) Show simply that G_f is homeomorphic to $\mathbb{R}^n$ (*i.e.,* there exists a bijection θ from $\mathbb{R}^n$ onto G_f such that both θ and θ^{-1} are continuous).

Hints. 1°) G_f is the image of $\mathbb{R}^n$ under the continuous mapping $x \in \mathbb{R}^n \mapsto (x, f(x)) \in \mathbb{R}^n \times \mathbb{R}^m$; it can also be described as

$$\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \mid y - f(x) = 0\}.$$

2°) Use the function $\theta(x) = (x, f(x))$. Clearly, θ is bijective from $\mathbb{R}^n$ onto G_f and continuous, while $\theta^{-1} : (x, f(x)) \in G_f \mapsto x \in \mathbb{R}^n$ is also continuous.

---

9. ★ *When* $\mathrm{tr}(AB) = 0$ *implies* $AB = 0$

Let A and B be two symmetric $n \times n$ real matrices. We suppose that A and B are either positive semidefinite or negative semidefinite.

Prove that $AB = 0$ whenever $\mathrm{tr}(AB) = 0$.

Warm-up: Prove that the only positive semidefinite matrix whose trace is zero is the null matrix.

Hints. The real vector space $\mathcal{M}_n(\mathbb{R})$, equipped with the scalar product $\langle\langle U, V\rangle\rangle = \mathrm{tr}(U^T V)$, is a Euclidean space. The norm derived from $\langle\langle \cdot, \cdot \rangle\rangle$ is denoted by $\|\cdot\|$.

For a positive semidefinite matrix U, there exists a (unique) positive semidefinite matrix, denoted by $U^{1/2}$ and called the "square root of U", such that $U^{1/2}U^{1/2} = U$.

There is no loss of generality in assuming that both A and B are positive semidefinite. It is suggested here to prove that $A^{1/2}B^{1/2} = 0$.

Answer. Assume without loss of generality that both A and B are positive semidefinite. We denote by $A^{1/2}$ (resp. $B^{1/2}$) the square root of A (resp. of B). By assumption, we have that

$$\mathrm{tr}(A^{1/2}A^{1/2}B^{1/2}B^{1/2}) = 0.$$

Using the properties of the trace function, we infer from the above that

$$0 = \mathrm{tr}(B^{1/2}A^{1/2}A^{1/2}B^{1/2}) = \left\|A^{1/2}B^{1/2}\right\|^2,$$

whence $A^{1/2}B^{1/2} = 0$. Thus, $AB = A^{1/2}(A^{1/2}B^{1/2})B^{1/2} = 0$.

Comments. - It is quite surprising that a weak assumption like $\mathrm{tr}(AB) = 0$ implies a strong property like $AB = 0$.

- Let $\mathcal{K}$ be the closed convex cone (in $\mathcal{S}_n(\mathbb{R})$) of positive semidefinite matrices. With the techniques used in this Tapa, one can easily prove the following:

$$\langle\langle A, A'\rangle\rangle \geqslant 0 \text{ when both } A \text{ and } A' \text{ lie in } \mathcal{K} \text{ (or in } -\mathcal{K});$$
$$-\mathcal{K} = \{B \in \mathcal{S}_n(\mathbb{R}) : \langle\langle A, B\rangle\rangle \leqslant 0 \text{ for all } A \in \mathcal{K}\}.$$

- $\mathcal{S}_n(\mathbb{R})$ is a subspace (of $\mathcal{M}_n(\mathbb{R})$) of the utmost importance; among its various properties, it is of dimension $\frac{n(n+1)}{2}$ and contains exclusively diagonalizable matrices. An interesting result asserts that, in some sense, one cannot do better; here it is: Let V be a vector subspace of $\mathcal{M}_n(\mathbb{R})$ containing only diagonalizable matrices; then the dimension of V is bounded from above by $\frac{n(n+1)}{2}$.

---

10. ★ *Diagonal and off-diagonal entries of a positive definite matrix*

Let $A = [a_{ij}] \in \mathcal{S}_n(\mathbb{R})$ be positive definite.

1°) What can be said about the diagonal entries a_{ii} of A?

2°) Check that the $n(n-1)/2$ off-diagonal entries $a_{ij}, i < j$, of A can be arbitrary real numbers.

Answers. 1°) Not much... except that they are positive (because $e_i^T A e_i = a_{ii}$).

Indeed, if $r_1, r_2, ..., r_n$ is any collection of positive real numbers, the diagonal matrix with $r_1, r_2, ..., r_n$ on its diagonal is positive definite.

2°) Let S be an arbitrary $n \times n$ symmetric matrix with zero diagonal entries. For $r > 0$, the matrix $A_r = S + rI_n$ has eigenvalues $\lambda_i + r$, where the λ_i's are eigenvalues of S (this is easy to check). Hence, for r large enough, all the eigenvalues of A_r are positive, and A_r is therefore positive definite. But the off-diagonal entries of A_r are those of S.

---

11. ★ *Is the product of two symmetric invertible matrices symmetric and invertible?*

If $A \in \mathcal{M}_n(\mathbb{R})$ is invertible, we know that so is its transpose and $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$; the notation A^{-T} can therefore be used without any ambiguity for either the inverse of the transpose of A, or for the transpose of the inverse of A.

Inverting and transposing matrices obey similar rules:

$$(AB)^{-1} = B^{-1}A^{-1}; (AB)^T = B^T A^T.$$

For a matrix $S \in \mathcal{M}_n(\mathbb{R})$, one easily checks that:

$$(S \text{ is symmetric and invertible}) \iff \left(S^{-T}S = I_n\right). \qquad (1)$$

Let now S_1 and S_2 be two symmetric invertible matrices, and consider their product $S = S_1 S_2$. We have

$$S^{-T}S = (S_2^{-T}S_1^{-T})(S_1 S_2) = S_2^{-T}(S_1^{-T}S_1)S_2 = I_n, \qquad (2)$$

whence (with (1)), S is symmetric and invertible...

Oops... this is grossly wrong: $S = S_1 S_2$ is indeed invertible... but not necessarily symmetric. So, where is the flaw in the above proof?

Answer. $(S_1 S_2)^{-T}$ is not $S_2^{-T}S_1^{-T}$ but $S_1^{-T}S_2^{-T}$... Hence, in the development (2), we have

$$S^{-T}S = (S_1^{-T}S_2^{-T})(S_1 S_2),$$

and we cannot go further...

Comment. At first glance, one easily falls into this trap...

---

12. ★ *Elliptic sets of minimal perimeter among those of prescribed area*

Let $\mathcal{E}$ be an elliptic set in the plane, *i.e.,* a compact convex set whose boundary is an ellipse Γ with parametric equation:

$$\begin{cases} x(t) = a\cos(t), \\ y(t) = b\sin(t), \end{cases} \quad t \in [0, 2\pi]\,.$$

We suppose that the area of $\mathcal{E}$ is $S > 0$. We denote by P the perimeter of $\mathcal{E}$, *i.e.* the length of Γ.

1°) Show that

$$P \geqslant 2\sqrt{\pi S}. \tag{1}$$

2°) What are the elliptic sets $\mathcal{E}$ for which $P = 2\sqrt{\pi S}$?

Hint. 1°) $S = \pi ab$, while $P = \int_0^{2\pi} \sqrt{a^2\sin^2(t) + b^2\cos^2(t)}\,\mathrm{d}t$.

Due to the concavity of the function $x \mapsto \sqrt{x}$, we have

$$\sqrt{a^2\sin^2(t) + b^2\cos^2(t)} \geqslant a\sin^2(t) + b\cos^2(t) \text{ for all } t \in [0, 2\pi]\,. \tag{2}$$

Answers. 1°) With the help of the inequality (2), we get that

$$P \geqslant \int_0^{2\pi} \left[a\sin^2(t) + b\cos^2(t)\right]\,\mathrm{d}t = \pi(a+b).$$

Now, the evaluation $S = \pi ab$ and the familiar inequality $(a+b)/2 \geqslant \sqrt{ab}$ directly lead to the announced inequality (1).

2°) The elliptic sets $\mathcal{E}$ for which $P = 2\sqrt{\pi S}$ are disks of radius $R = \sqrt{S/\pi}$.

---

13. ★ *A glimpse of the* Legendre *transform*

Let $f : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^\infty$ satisfying the following property: there exists a $c_0 > 0$ such that

$$f''(x) \geqslant c_0 \text{ for all } x \in \mathbb{R}. \tag{1}$$

Such a function is therefore (strictly) convex.

1°) Show that f' is a strictly increasing homeomorphism from $\mathbb{R}$ onto $\mathbb{R}$. We set $g = (f')^{-1}$.

2°) Check that g is of class $\mathcal{C}^\infty$ and provide the expressions of $g'(y)$ and $g''(y)$ in terms of f'', f''' and g.

3°) Let $f^* : \mathbb{R} \to \mathbb{R}$ be defined as

$$f^*(y) = yx - f(x), \text{ where } x \text{ is the unique solution of } f'(x) = y. \quad (2)$$

Prove that f^* is of class $\mathcal{C}^\infty$ and determine $(f^*)'(y)$ and $(f^*)''(y)$ in terms of f'' and g.

4°) We suppose, moreover, that there exists a $b_0 > 0$ such that

$$f''(x) \leqslant b_0 \text{ for all } x \in \mathbb{R}. \quad (1')$$

Check that one is quite entitled to define $(f^*)^*$, and that we have: $(f^*)^* = f$.

5°) Verify that

$$f^*(y) = \sup_{x \in \mathbb{R}} (yx - f(x)). \quad (3)$$

6°) Let f_1 and f_2 satisfy all the assumptions imposed above on f, let $\varphi = f_1 - f_2$ and $\psi = f_2^* - f_1^*$. Prove that:

$$\begin{cases} \inf_{x \in \mathbb{R}} \varphi = \inf_{x \in \mathbb{R}} \psi; \\ \sup_{x \in \mathbb{R}} \varphi = \sup_{x \in \mathbb{R}} \psi. \end{cases} \quad (4)$$

Hints. 1°) Use mean value inequalities like

$$f'(y) - f'(x) \geqslant c_0(y - x) \text{ whenever } y > x$$

to prove that f' is injective and that

$$f'(x) \to +\infty \text{ when } x \to +\infty, \quad f'(x) \to -\infty \text{ when } x \to -\infty.$$

5°) The differentiable concave function $x \mapsto yx - f(x)$ is maximized on $\mathbb{R}$ at the unique point $x = (f')^{-1}(y) = g(y)$.

6°) Just use the definitions of an infimum and of a supremum on $\mathbb{R}$ of real-valued functions.

Answers. 2°) According to the inverse function theorem, $g = (f')^{-1}$ is of class $\mathcal{C}^1$ on $\mathbb{R}$, with:

$$g'(y) = \frac{1}{f''[g(y)]} \text{ for all } y \in \mathbb{R}. \quad (5)$$

The function g' itself is of class $\mathcal{C}^1$ on $\mathbb{R}$, and:

$$g''(y) = \frac{f'''\left[g(y)\right]}{\left\{f''\left[g(y)\right]\right\}^3}. \tag{6}$$

By induction, one easily verifies that g is of class $\mathcal{C}^\infty$.

3°) We have that

$$f^*(y) = yg(y) - f\left[g(y)\right].$$

Consequently, f^* is of class $\mathcal{C}^\infty$, and keeping in mind that $f'\left[g(y)\right] = y$,

$$\left(f^*\right)'(y) = g(y), \quad \left(f^*\right)''(y) = \frac{1}{f''\left[g(y)\right]}. \tag{7}$$

4°) The additional assumption (1') implies that $\left(f^*\right)''(y) \geqslant \frac{1}{b_0}$ for all $y \in \mathbb{R}$. So, it is possible to apply to f^* the process displayed above for f, and thus define $\left(f^*\right)^*$. Then, it suffices to permute the roles of x and y in the basic relation $f(x) + f^*(y) = xy$ to obtain $\left(f^*\right)^* = f$.

Comments. The transformation $f \mapsto f^*$ expressed in (2) (by inverting f') is called the LEGENDRE transformation. The same transformation, as expressed in (3) (resulting from a maximization problem, where the differentiability of f is not assumed), is called the LEGENDRE–FENCHEL transformation. It plays a key-role in modern Convex analysis and Optimization.

---

14. ★ HLAWKA's *inequality or the parallelepiped inequality*

Let $(E, \|\cdot\|)$ be a normed vector space and let u, v, w be three elements in E. Show that

$$\|u+v\| + \|v+w\| + \|w+u\| \leqslant \|u\| + \|v\| + \|w\| + \|u+v+w\|\,. \tag{1}$$

Hint. Develop the product

$$\begin{aligned}
\Delta &= \Big[\, \|u\| + \|v\| + \|w\| + \|u+v+w\| - \|u+v\| \\
&\quad - \|v+w\| - \|w+u\| \,\Big] \times \Big[\, \|u\| + \|v\| + \|w\| + \|u+v+w\| \,\Big] \\
&= \Big(\|u\| + \|v\| - \|u+v\| \Big) \times \Big(\|w\| - \|u+v\| + \|u+v+w\| \Big) \\
&\quad + 2 \text{ similar expressions;}
\end{aligned}$$

then make use of the triangle inequality, several times, to obtain $\Delta \geqslant 0$.

Comment. The inequality (1) is named after the Austrian mathematician E. Hlawka (1916–2009).

---

15. ★ *The* Massera–Schäffer *inequality*

Let $(E, \|\cdot\|)$ be a normed vector space and let u, v be two non-null elements in E.

1°) Show that

$$\left\| \frac{u}{\|u\|} - \frac{v}{\|v\|} \right\| \leqslant \frac{2}{\max(\|u\|, \|v\|)} \|u - v\| . \tag{1}$$

2°) Verify that one cannot replace the 2 in inequality (1) by a constant $k < 2$.

3°) (*a*) Check that the mapping $x \neq 0 \mapsto f(x) = \frac{x}{\|x\|}$ is Lipschitz on the open set $\Omega = \{x : \|x\| > 1\}$, with a Lipschitz constant $L = 2$.

(*b*) Provide a counterexample showing that the (best) Lipschitz constant of f on Ω cannot be lowered to $L < 2$.

Hints. 1°) For symmetry reasons, it suffices to consider the case where $\|v\| \leqslant \|u\|$.

2°) Proof by contradiction. Choose u of norm 1, $v = tu$ with $t > 0$ tending to 0.

Answer. 3°) (*b*). Let $E = \mathbb{R}^2$ be equipped with the norm $\|(x, y)\| = \max(|x|, |y|)$. Then consider $u = (1, 1)$ and $v = (1 - \varepsilon, 1 + \varepsilon)$ with $0 < \varepsilon < 1$. We have

$$\left\| \frac{u}{\|u\|} - \frac{v}{\|v\|} \right\| = \frac{2}{1 + \varepsilon} \times \varepsilon \text{ and } \|u - v\| = \varepsilon.$$

Comment. The inequality (1) is due to Massera and Schäffer (1958).

---

16. ★ *The* Dunkl–Williams *inequality*

Let E be a vector space equipped with an inner product $\langle \cdot, \cdot \rangle$. We denote by $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$ the norm derived from the inner product.

1°) Let u, v be two non-null elements in E. Show that

$$\left\| \frac{u}{\|u\|} - \frac{v}{\|v\|} \right\| \leqslant \frac{2}{\|u\| + \|v\|} \|u - v\| . \tag{1}$$

2°) Prove that equality holds in (1) if and only if:

$$\|u\| = \|v\| \text{ or } \frac{u}{\|u\|} = -\frac{v}{\|v\|}. \tag{2}$$

Hints. 1°) The key is the calculus rule $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle$, and its scalar version $(\|x\| + \|y\|)^2 = \|x\|^2 + \|y\|^2 + 2\|x\| \times \|y\|$.

Comments. - A consequence of (1) is that the mapping $x \neq 0 \mapsto f(x) = \frac{x}{\|x\|}$ is Lipschitz on the open set $\Omega = \{x : \|x\| > 1\}$, with Lipschitz constant $L = 1$.

- The inequality (1) is due to Dunkl and Williams (1964). It has been proved furthermore that the Dunkl–Williams inequality characterizes the norms derived from inner products.

For more on inequalities involving norms, see:

J.-B. Hiriart-Urruty, *La lutte pour les inégalités.* Revue de la Filière Mathématiques RMS (ex-Revue Mathématiques Spéciales), Vol 122, n° 2, 12–18 (2011–2012).

---

17. ★ *Examples of norms on the space of polynomial functions*

Let E be the vector space of real polynomial functions of a real variable. For $a < b$, one defines $N_{a,b} : E \to \mathbb{R}$ as follows:

$$N_{a,b}(P) = \max_{x \in [a,b]} |P(x)| , \ P \in E.$$

Check that $N_{a,b}$ is a norm on E.

Hint. All the required properties for $N_{a,b}$ to be a norm are easy to prove; only one is often overlooked:

$$(N_{a,b}(P) = 0) \Rightarrow (P = 0) .$$

When the polynomial function P satisfies $P(x) = 0$ for all $x \in [a, b]$, P is null everywhere on $\mathbb{R}$ (because a non-null polynomial of degree n has at most n roots).

---

18. ★ *Comparing the norms* $\|\cdot\|_1$ *and* $\|\cdot\|_\infty$ *on* $\mathcal{C}([-1,1],\mathbb{R})$

For a positive integer n, let $f_n : [-1,1] \to \mathbb{R}$ be defined as follows:

$$f_n(x) \quad = \quad \begin{cases} 1 & \text{if } x \in [-1,0]\,, \\ 1+x^n & \text{if } x \in [0,1]\,. \end{cases}$$

1°) Show that the sequence of functions (f_n) converges in $(\mathcal{C}([-1,1],\mathbb{R}), \|\cdot\|_1)$ towards the function g which constantly takes the value 1 on $[-1,1]$.

2°) (a) Does the sequence of functions (f_n) converge in $(\mathcal{C}([-1,1],\mathbb{R}), \|\cdot\|_\infty)$ towards the function g?

(b) Does the sequence of functions (f_n) converge pointwise towards the function g?

3°) What conclusion on the norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ can be drawn from the results above?

Answers. 1°) We have: $\|f_n - g\|_1 = \frac{1}{n+1} \to 0$ as $n \to +\infty$.

2°) (a). No, since $\|f_n - g\|_\infty \geqslant |f_n(1) - g(1)| = 1$.

(b) No. The real sequence $(f_n(1) = 2)$ is constant, its limit is different from $g(1) = 1$. Actually, the sequence of functions (f_n) converges pointwise towards a function $\widetilde{g}$ which is not in $\mathcal{C}([-1,1],\mathbb{R})$.

3°) The norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are not equivalent on $\mathcal{C}([-1,1],\mathbb{R})$.

---

19. ★ *Two examples of metric spaces*

1°) Let d be defined on $(0,+\infty) \times (0,+\infty)$ by:

$$x > 0, y > 0 \mapsto d(x,y) = \left|\log\left(\frac{x}{y}\right)\right|. \tag{1}$$

Check that d is a distance on $(0,+\infty)$.

2°) Let d be defined on $\mathbb{R}^n \times \mathbb{R}^n$ by:

$$x = (x_1, ..., x_n), y = (y_1, ..., y_n) \mapsto d(x,y) = \operatorname{card}\{i : x_i \neq y_i\}. \tag{2}$$

(a) Check that d is a distance on $\mathbb{R}^n$.

(b) Let N be defined on $\mathbb{R}^n$ by:

$$x = (x_1, ..., x_n) \mapsto N(x) = \text{card}\,\{i : x_i \neq 0\}\,. \tag{3}$$

- Is N a norm on $\mathbb{R}^n$?
- Show that $\left(\|x\|_p\right)^p$ has a limit as $p > 0$ tends to 0 and compare it with $N(x)$.

Answers. 2°) (b). No, N is not a norm: the property $N(\lambda x) = |\lambda|\, N(x)$ is not satisfied for all $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}^n$.

However, $d(x, y) = N(x - y)$ (*cf.* (2)) defines a distance on $\mathbb{R}^n$.

We have:

$$\lim_{p \to 0} \left(\|x\|_p\right)^p = N(x) \text{ for all } x \in \mathbb{R}^n.$$

Comments. The distance d defined in (2) is called the HAMMING distance; it is used in Computer science. The function N defined in (3) measures the so-called sparsity of $x = (x_1, ..., x_n) \in \mathbb{R}^n$; it is much used in the area of Compressed sensing.

---

20. ★ *The* LIPSCHITZ *property of the distance function to a set*

Let (E, d) be a metric space. For a nonempty set $S \subset E$, one defines the *distance function* to S as:

$$x \in E \mapsto d_S(x) = \inf_{s \in S} d(x, s). \tag{1}$$

Prove that

$$|d_S(x) - d_S(x)| \leqslant d(x, y) \text{ for all } x, y \text{ in } E. \tag{2}$$

Hint. Properly use the definition of $\mu = \inf A$ when $A \subset \mathbb{R}^+$, and the triangle inequality with the distance d.

Comments. - It is somewhat surprising that one obtains the LIPSCHITZ property of d_S on the whole of E (property (2)) without any assumption on the set S.

- The distance function does not detect any difference between S and its closure $\overline{S}$; indeed, $d_S = d_{\overline{S}}$.

- We have:

$$\{x \in E : d_S(x) \leqslant 0\} = \{x \in E : d_S(x) = 0\} = \overline{S} = (\mathrm{bd}S) \cup (\mathrm{int}S).$$

Hence, the distance function d_S is unable to distinguish the boundary of S from the interior of S. This will be the privilege of the signed distance function Δ_S (see the next proposal).

---

21. ★ *The signed distance function to a set*

Let (E, d) be a metric space. Let $S \subset E$ be nonempty and different from the whole space E. One defines the (so-called) *signed distance function* to S as:

$$x \in E \mapsto \Delta_S(x) = d_S(x) - d_{S^c}(x),$$

where S^c denotes the complementary set of S in E (S^c is sometimes denoted $E \setminus S$). Thus,

$$\Delta_S(x) \quad = \quad \begin{cases} d_S(x) & \text{if } x \notin S, \\ -d_{S^c}(x) & \text{if } x \in S. \end{cases}$$

1°) Show that:

$$\{x \in E : \Delta_S(x) < 0\} = \mathrm{int}S; \tag{1}$$
$$\{x \in E : \Delta_S(x) = 0\} = \mathrm{bd}S; \tag{2}$$
$$\{x \in E : \Delta_S(x) > 0\} = \mathrm{ext}S. \tag{3}$$

(ext$S$ is the exterior of S, that is, the interior of S^c.)

Comments. - Some examples in the plane will help to familiarize the reader with Δ_S.

- Clearly, $\Delta_{S^c} = -\Delta_S$.

- The three sets int$S$, bd$S$, ext$S$ form a partition of E. According to the formulas in (1)–(3) above, the Δ_S function is able to distinguish between them.

- If S_1 and S_2 are two closed sets, then:

$$(S_1 \subset S_2) \Leftrightarrow (\Delta_{S_1} \leqslant \Delta_{S_2}). \tag{4}$$

---

22. ★ *The distance functions and convexity*

Let $(E, \|\cdot\|)$ be a normed vector space. Let $S \subset E$ be nonempty, different from the whole space E, and closed. We denote by d_S (resp. Δ_S) the distance function (resp. the signed distance function) to S.

1°) Prove the equivalence of the following assertions:

(i) S is convex;
(ii) d_S is convex;
(iii) Δ_S is convex.

2°) Prove that

$$|\Delta_S(x) - \Delta_S(x)| \leqslant \|x - y\| \text{ for all } x, y \text{ in } E. \tag{1}$$

Hints. 1°) The convexity of the function $f : E \to \mathbb{R}$ implies the convexity of the sublevel-sets $\{x \in E : f(x) \leqslant r\}$, $r \in \mathbb{R}$.

2°) The functions d_S and d_{S^c} are LIPSCHITZ on the whole of E, with 1 as a LIPSCHITZ constant (see Tapa 20). Moreover, if $x \in \text{int}S$ and $y \in \text{ext}S$, there is a point z on the line-segment $[x, y]$ which lies on the boundary of S; consequently, $\|y - x\| = \|y - z\| + \|z - x\|$.

Comment. To the equivalences displayed in 1°) one could add the convexity of the function d_S^2 ... but not that of Δ_S^2.

---

23. ★ *The k-*LIPSCHITZ *envelope of a function*

Let (E, d) be a metric space and $f : E \to \mathbb{R}$. For our approach to designing the k-LIPSCHITZ (lower-)envelope of f ($k \geqslant 0$), we make the following assumption:

$$\left\{ \begin{array}{c} \text{There exists a } \varphi : E \to \mathbb{R},\ k\text{-LIPSCHITZ on } E, \text{ such that} \\ \varphi(x) \leqslant f(x) \text{ for all } x \in E. \end{array} \right. \tag{1}$$

1°) Verify that assumption (1) is equivalent to the following:

$$\left\{ \begin{array}{c} \text{There exists } \overline{x} \ \in E \text{ and } r \in \mathbb{R} \text{ such that} \\ -kd(x, \overline{x}) + r \leqslant f(x) \text{ for all } x \in E. \end{array} \right. \tag{2}$$

For reasons that one can easily imagine, such functions $\psi : x \mapsto \psi(x) = -kd(x, \overline{x}) + r$ are called "hat-like"; there are examples of k- LIPSCHITZ functions on E.

Under the assumption (1), one defines $f_k : E \to \mathbb{R}$ as follows:

$$x \in E \mapsto f_k(x) = \inf_{y \in E} \left[f(y) + k d(x,y) \right]. \tag{3}$$

2°) Prove that f_k is the largest k-LIPSCHITZ function φ on E satisfying $\varphi \leqslant f$. Thus, f_k is called the k-LIPSCHITZ (lower-)envelope of f.

3°) (*a*) Do f and f_k coincide at some $x \in E$?

(*b*) Suppose that f is k-LIPSCHITZ on a subset $S \subset E$. Do f and f_k coincide on S?

4°) Differentiability. Suppose that $(E, \|\cdot\|)$ is a normed vector space and that f is differentiable at $\overline{x} \in E$. Prove that if $\|Df(\overline{x})\| > k$, then f_k and f do not coincide at $\overline{x}$ (that is to say, $f_k(\overline{x}) < f(\overline{x})$). Here $\|Df(\overline{x})\|$ stands for the norm of $Df(\overline{x})$ as a continuous linear form on E: $\|Df(\overline{x})\| = \sup_{\|d\| \leqslant 1} |Df(\overline{x})d|$.

5°) Convexity. Suppose that $(E, \|\cdot\|)$ is a normed vector space and that f is convex on E.

(*a*) Show that the k-LIPSCHITZ envelope f_k of f is also convex on E.

(*b*) Show that f and f_k coincide at a point $\overline{x}$ where f is differentiable if and only if $\|Df(\overline{x})\| \leqslant k$.

6°) Lower-semicontinuous envelope of f. Suppose that f satisfies assumption (1). Prove that $g = \sup_{l \geqslant k} f_l$ is the lower-semicontinuous envelope of f, *i.e.* the largest lower-semicontinuous function g such that $g \leqslant f$.

Hints. 2°) Check firstly that f_k is finite everywhere, then prove that it is k-LIPSCHITZ on E.

4°) Proof by contradiction.

5°) (*b*). In view of the result in 4°), it remains to prove that $f_k(\overline{x}) = f(\overline{x})$ whenever $\|Df(\overline{x})\| \leqslant k$. For that, use the convex inequality:

$$f(y) \geqslant f(\overline{x}) + Df(\overline{x})(y - \overline{x}) \text{ for all } y \in E.$$

6°) An analytical definition of the lower-semicontinuous envelope g of f is:

$$g(x) = \liminf_{x' \to x} f(x').$$

From the geometrical viewpoint, $\{(x,r) \in E \times \mathbb{R},\ g(x) \leqslant r\}$ is exactly the closure of $\{(x,r) \in E \times \mathbb{R},\ f(x) \leqslant r\}$.

Answers. 3°) (*a*)–(*b*). The answer is No in both cases. Counterexamples can be built up with functions f of a real variable, for which $f_k(x) < f(x)$ for all $x \in \mathbb{R}$.

Comments. - Under the assumption

$$\left\{ \begin{array}{c} \text{There exists a } \varphi : E \to \mathbb{R},\ k\text{-Lipschitz on } E, \text{ such that} \\ \varphi(x) \geqslant f(x) \text{ for all } x \in E, \end{array} \right. \tag{4}$$

the smallest k-Lipschitz function φ on E satisfying $\varphi \geqslant f$ is

$$x \in E \mapsto f^k(x) = \sup_{y \in E} \left[f(y) - kd(x,y) \right]. \tag{5}$$

f^k is called the k-Lipschitz (upper-)envelope of f.

- Definitions (3) and (5) are global, in the sense that the values of f on the whole of E are of importance.

Reference:

J.-B. Hiriart-Urruty and M. Volle, *Enveloppe k-Lipschitzienne d'une fonction.* Revue de Mathématiques Spéciales n°9–10, (1996), 785–793.

---

24. ★ *Extensions of* Lipschitz *functions*

Let (E,d) be a metric space, let $S \subset E$ be nonempty, and let $f : S \to \mathbb{R}$ be a Lipschitz function on S with k as a Lipschitz constant. We intend to determine the "best" Lipschitz extensions of f to the whole of E with the same Lipschitz constant k. For that, we define:

$$f_{S,k}(x) = \inf_{y \in S} \left[f(y) + kd(x,y) \right], \tag{1}$$

$$f^{S,k}(x) = \sup_{y \in S} \left[f(y) - kd(x,y) \right]. \tag{2}$$

1°) Prove that $f_{S,k}$ is Lipschitz on the whole of E with k as a Lipschitz constant, and that it coincides with f on S.

2°) Check that $f^{S,k} = -\left(-f\right)_{S,k}$.

3°) Let $g : E \to \mathbb{R}$ be a k-Lipschitz extension of f, that is to say, a Lipschitz function on E, with k as a Lipschitz constant, and coinciding with f on S. Show that

$$f^{S,k} \leqslant g \leqslant f_{S,k}. \tag{3}$$

Comments. - To give a simple example, if f is the null function on S and $k \geqslant 0$, the resulting extensions $f_{S,k}$ and $f^{S,k}$ are kd_S and $-kd_S$, respectively.

- If $(E, \|\cdot\|)$ is a normed vector space, if $S \subset E$ is convex and if $f : S \to \mathbb{R}$ is convex, then the extension $f_{S,k}$ is also convex.

- The formulations (1) and (2) go back to E.J. McShane (1934). They are in a certain sense "localized" (to S) versions of definitions (3) and (5) in the previous Tapa.

---

25. ★ *A further useful distance function*

Let $(E, \|\cdot\|)$ be a normed vector space and let $S \subset E$ be assumed nonempty, different from the whole space E. We define the function $\mu_S : E \to \mathbb{R}$ as follows:

$$\mu_S(x) = -\, d_{S^c}(x),$$

where S^c denotes the complementary set of S in E, and d_{S^c} the distance function to this set.

1°) Show that μ_S is convex on S whenever S is convex.

2°) The function μ_S is Lipschitz on E with 1 as a Lipschitz constant (as a distance function to a set). We consider the "maximal" Lipschitz extension of $\mu_S = -\, d_{S^c}$ from S to the whole of E with the same Lipschitz constant 1, that is (see Tapa 24):

$$f_S(x) = \inf_{y \in S} \left[\mu_S(y) + \|x - y\|\right]. \tag{1}$$

Prove that f_S is nothing else than the signed distance function Δ_S to the set S (see Tapa 21 for its definition).

---

26. ★ Cauchy *sequences with cluster points*

Let (u_n) be a Cauchy sequence in a metric space (E, d). We suppose that (u_n) has a cluster point $a \in E$.

Show that (u_n) is a convergent sequence with $\lim_{n \to +\infty} u_n = a$.

Hint. Use the definitions of a Cauchy sequence and of a cluster point; then apply the triangle inequality with the distance d.

---

27. ★ *Sequences in a compact set with unique cluster points*

Let (E,d) be a metric space and let K be a compact subset of E. We consider a sequence $(u_n)_{n\geqslant 1}$ of elements in K possessing only one cluster point a.

1°) Show that the whole sequence $(u_n)_{n\geqslant 1}$ is convergent and that its limit is a.

2°) Show with the help of a counterexample in $(\mathbb{R},|\cdot|)$ that the result above may fail if K is not assumed to be compact.

Hint. 1°) Proof by contradiction: suppose there is an $\varepsilon>0$ and a subsequence $(u_{n_k})_{k\geqslant 1}$ extracted from $(u_n)_{n\geqslant 1}$ such that $d(u_{n_k},a)\geqslant\varepsilon$ for all k. Since $(u_{n_k})_{k\geqslant 1}$ is a sequence in the compact set K, a new convergent subsequence can be extracted.

Answer. 2°) Consider the real sequence $(u_n)_{n\geqslant 1}$ defined as: $u_n=1$ if n is even, $u_n=1+n$ if n is odd. This non-convergent sequence has just one cluster point $a=1$.

---

28. ★ CAUCHY *sequences vs convergent sequences or series*

Let (E,d) be a metric space and let $(u_n)_{n\geqslant 1}$ be a sequence of elements in E.

1°) Suppose that:

$$\text{For all positive integer } p,\ d(u_{n+p},u_n)\to 0 \text{ when } n\to+\infty. \qquad (1)$$

Is (u_n) a CAUCHY sequence?

2°) (a) Suppose that the series with general term $d(u_{n+1},u_n)$ is convergent (that is: $\sum_{n=1}^{+\infty} d(u_{n+1},u_n)<+\infty$). Is (u_n) a CAUCHY sequence?

(b) Conversely, if (u_n) is a CAUCHY sequence, is the series with general term $d(u_{n+1},u_n)$ convergent?

Answers. 1°) No... although the condition (1) looks pretty much like the CAUCHY property for the sequence (u_n). Counterexamples on $E=\mathbb{R}$ are:

$$u_n = \ln(n), \text{ whence } |u_{n+p}-u_n|=\ln\left(1+\frac{p}{n}\right);$$

$$u_n = 1+\frac{1}{2}+\ldots+\frac{1}{n}, \text{ whence } |u_{n+p}-u_n|\leqslant\frac{p}{n+1}.$$

Both real sequences do satisfy (1) but are not convergent, hence are not CAUCHY sequences.

2°) (*a*) Yes it is, since, by using the triangle inequality:

$$\text{For } p < q,\ d(u_p, u_q) \leqslant \sum_{i=p}^{q-1} d(u_i, u_{i+1}), \tag{2}$$

and the convergence of the series with general term $d(u_{n+1}, u_n)$ ensures that $\sum_{i=p}^{q-1} d(u_i, u_{i+1})$ tends to 0 when $p, q \to +\infty$.

(*b*) No. The real sequence $\left(\frac{(-1)^n}{n}\right)_{n\geqslant 1}$ is a CAUCHY sequence, while $|u_{n+1} - u_n| \geqslant \frac{1}{n}$ for all $n \geqslant 1$, so that the series with general term $|u_{n+1} - u_n|$ is divergent.

---

29. ★ *A characterization of complete normed vector spaces (*BANACH *spaces)*

Let $(E, \|\cdot\|)$ be a normed vector space. We intend to characterize the completeness of $(E, \|\cdot\|)$ in terms of series in E.

1°) Easy part. Assuming that E is complete, check that the series with general term v_n converges in E whenever the (real) series with general term $\|v_n\|$ converges.

2°) Converse. Suppose that $(E, \|\cdot\|)$ satisfies the property used above: a series with general term v_n converges in E whenever the (real) series with general term $\|v_n\|$ converges. Prove that E is complete.

Hint. 2°) Start with a CAUCHY sequence (u_n) in E. Therefore, for all positive integers k, there exists an N_k such that

$$\|u_m - u_n\| \leqslant \frac{1}{2^k} \text{ whenever } m \geqslant n \geqslant N_k.$$

We may assume that $N_1 < N_2 < ... < N_k < ...$ Then consider the series with general term $v_n = u_{N_n} - u_{N_{n-1}}$. This has been done so that

$$\begin{aligned} v_1 + ... + v_n &= -u_{N_1} + u_{N_n}; \\ \sum_{n=1}^{+\infty} \|v_n\| &< +\infty. \end{aligned}$$

Make use of the assumption implying the convergence of the series with general term v_n; as a consequence, the CAUCHY sequence (u_n) has a cluster point $v \in E$; it therefore converges towards v (see Tapa 26).

Comment. The result of question 1°) is the usual way of taking advantage of the completeness of E. The converse, *i.e.*, the result of question 2°), is an unusual way of proving the completeness of E.

---

30. ★ *A natural attempt at norming the sum of two normed vector spaces*

Let V_1 and V_2 be two vector subspaces of the same vector space E. We suppose that V_1 is normed with a norm denoted by $\|\cdot\|_1$ and that V_2 is normed with a norm denoted by $\|\cdot\|_2$. We intend to norm, if possible, the vector subspace $V = V_1 + V_2$ with the help of $\|\cdot\|_1$ and $\|\cdot\|_2$. Recall that

$$V_1 + V_2 = \{v_1 + v_2 : v_1 \in V_1, v_2 \in V_2\}.$$

A natural attempt to define a norm on V is to propose, for all $v \in V$,

$$N(v) = \inf\{\|v_1\|_1 + \|v_2\|_2 : v_1 \in V_1, v_2 \in V_2, v = v_1 + v_2\}. \quad (1)$$

1°) Check that N satisfies all the properties required for a norm, except this one:

$$(N(v) = 0) \Rightarrow (v = 0). \quad (2)$$

One therefore says that N is only a semi-norm on V.

2°) Illustration. Consider $E = \mathcal{C}([0,2], \mathbb{R})$; $V_1 = V_2 =$ the space of polynomial functions on $[0,2]$. Hence, we have $V = V_1 = V_2$ in this example.

We equip V_1 with the norm $\|P\|_1 = \max_{x\in[0,1]} |P(x)|$, and V_2 with the norm $\|P\|_2 = \max_{x\in[1,2]} |P(x)|$.

Let $f \in E$ be defined as follows:

$$f(x) = \begin{cases} 1 & \text{if } x \in [0,1], \\ 2-x & \text{if } x \in [1,2]. \end{cases}$$

We now consider a sequence of polynomial functions (P_n) converging towards f uniformly on $[0,2]$ (this is possible thanks to the Weierstrass approximation theorem).

(*a*) Show that (P_n) converges towards $Q = 1$ in $(V_1, \|\cdot\|_1)$ and towards $R = 2 - x$ in $(V_2, \|\cdot\|_2)$.

(*b*) Let $P = 1 - x$. By decomposing it as $P = (P_n - 1) + (2 - x - P_n)$, show that $N(P)$, as defined in (1), equals 0.

Hence, as a general rule, N is not a norm on V.

3°) What property should be added to the pair $(V_1, \|\cdot\|_1), (V_2, \|\cdot\|_2)$ to ensure that N, as defined in (1), is a norm on $V = V_1 + V_2$?

4°) Suppose that $(V_1, \|\cdot\|_1)$ and $(V_2, \|\cdot\|_2)$ are complete; suppose also that N is indeed a norm on $V = V_1 + V_2$.

Prove that V, equipped with the norm N, is complete.

Hints. 1°) Use properly the definition of $\mu = \inf A$ when $A \subset \mathbb{R}$ is a nonempty set bounded from below.

2°) (a). We have:

$$\begin{aligned} \|P_n - Q\|_1 &\leqslant \max_{x\in[0,2]} |P_n(x) - f(x)| \,; \\ \|P_n - R\|_2 &\leqslant \max_{x\in[0,2]} |P_n(x) - f(x)| \,. \end{aligned}$$

(b) We decompose $P = 1 - x \in V_1 + V_2$ as follows:

$$P = (P_n - 1) + (2 - x - P_n) = (P_n - Q)_{\in V_1} + (R - P_n)_{\in V_2} \,.$$

By definition of $N(P)$ as a lower bound,

$$N(P) \leqslant \|P_n - Q\|_1 + \|P_n - R\|_2 \,.$$

4°) For example, make use of the characterization of complete normed vector spaces with the help of series (see Tapa 29).

Answer. 3°) The difficulty (see, for example, 2°)) is that

$$\left(\begin{array}{c} P_n \to Q \text{ in } (V_1, \|\cdot\|_1) \\ \text{and } P_n \to R \text{ in } (V_2, \|\cdot\|_2) \end{array} \right) \not\Rightarrow (Q = R) \,.$$

One therefore should add a kind of "compatibility assumption" between $(V_1, \|\cdot\|_1)$ and $(V_2, \|\cdot\|_2)$, for example: if (z_k) is a sequence of elements in $V_1 \cap V_2$, then

$$\left(\begin{array}{c} z_k \to a \text{ in } (V_1, \|\cdot\|_1) \\ \text{and } z_k \to b \text{ in } (V_2, \|\cdot\|_2) \end{array} \right) \Rightarrow (a = b) \,. \qquad (\mathcal{C})$$

Comments. - One easily falls into the trap (as we did!) of believing one has proved that $(N(v) = 0) \Rightarrow (v = 0)$.

- The compatibility assumption $(\mathcal{C})$ is automatically satisfied in some practical situations, such as the one dealing with $L^p(\mathbb{R})$ spaces. For more on this, see:

J.-B. Hiriart-Urruty and P. Lassère, *When is $L^r(\mathbb{R})$ contained in $L^p(\mathbb{R}) + L^q(\mathbb{R})$?* American Math. Monthly, 55–61 (2013).

---

31. ★ *Non-continuity of the length of a curve*

The following example illustrating the non-continuity of $L : f \mapsto L(f)$ (= length of the graph of the function f) is gobsmacking... Consider the sequence $(f_n)_{n\geqslant 1}$ of functions whose graphs are drawn below:

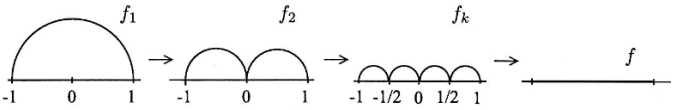

Figure 1

The graph of f_n is made of n half-circles of diameter $2/n$.

1°) Check that the sequence of functions (f_n) converges towards the function $f \equiv 0$ (identically equal to 0) uniformly on $[-1, 1]$.

2°) What are the lengths of the graphs of f_n? of the graph of f? Conclusion?

Answer. 2°) The length of the graph of f_n equals π for all n, while the length of the graph of f is 2.

Comments. - Even if the convergence of (f_n) towards f is uniform on $[a, b]$, one cannot secure that the length $L(f_n)$ converges to the length $L(f)$. The only "continuity" result on the length function that one could expect is its lower-semicontinuity, *i.e.,*

$$\liminf_{n\to+\infty} L(f_n) \geqslant L(f).$$

- For another counterexample of the same kind, with $\mathcal{C}^\infty$ functions f_n instead, consider

$$x \in [0, \pi] \mapsto f_n(x) = \frac{\sin(nx)}{n}.$$

---

32. ★ *A situation leading to a fixed point of a function f*

Let K be a compact subset of a metric space (E, d), let $f : K \to K$ be a function satisfying:

$$d\,[f(u), f(v)] < d(u, v) \text{ whenever } u \neq v \text{ in } K. \tag{1}$$

By minimizing the function

$$u \mapsto \varphi(u) = d\,[f(u), u]$$

on K, show that f has one, and only one, fixed point in K.

Comment. The function f is continuous on K but, in spite of the property (1), it is not a contraction on K.

Hints. A minimizer $\overline{u}$ of the continuous function φ on the compact set K is indeed a fixed point of f. To prove this, proceed by contradiction: if $f(\overline{u})$ is different from $\overline{u}$, apply (1) to $u = \overline{u}$ and $v = f(\overline{u})$; this leads to $\varphi\,[f(\overline{u})] < \varphi(\overline{u})$.

---

33. ★ *An equivalent definition of orthogonality*

Let H be an inner product space; $\langle \cdot, \cdot \rangle$ denotes the inner product and $\|\cdot\|$ stands for the norm associated with this inner product. Prove that two elements u and v in H are orthogonal if and only if

$$\|u\| \leqslant \|u + tv\| \text{ for all } t \in \mathbb{R}. \tag{1}$$

Hint. Simply express (1) as follows:

$$\|u\|^2 \leqslant \|u + tv\|^2 = \|u\|^2 + t^2 \|v\|^2 + 2t \langle u, v \rangle \text{ for all } t \in \mathbb{R}.$$

Comment. The relation (1) could be adopted for a possible definition of orthogonality in a normed vector space $(E, \|\cdot\|)$... However, it is not symmetrical in u and v.

---

34. ★ *A simple homeomorphism between the whole space and the open unit ball in a normed vector space*

Let $(E, \|\cdot\|)$ be a normed vector space. We intend to prove simply that the whole space E and the open unit ball $B(0,1)$ are homeomorphic.

1°) Warm-up. Check that

$$f : x \in \mathbb{R} \mapsto f(x) = \frac{x}{1+|x|} \tag{1}$$

is a homeomorphism from $\mathbb{R}$ onto $]-1,1[$ (the open interval with endpoints -1 and 1). For that, one inevitably will be led to determine the explicit form of $f^{-1}(y)$, $y \in]-1,1[$.

2°) General context. Prove that

$$f : x \in E \mapsto f(x) = \frac{x}{1+\|x\|} \tag{2}$$

defines a homeomorphism from E onto $B(0,1)$.

Answers. 1°) The proposed f is indeed a bijection from $\mathbb{R}$ onto $]-1,1[$; the inverse bijection is

$$y \in]-1,1[\mapsto f^{-1}(y) = \frac{y}{1-|y|}.$$

2°) As foreseen,

$$y \in B(0,1) \mapsto f^{-1}(y) = \frac{y}{1-\|y\|}$$

is the inverse of the function f proposed in (2).

---

35. ★ *Generalized parallelogram rule*

Let H be an inner product space; $\langle \cdot, \cdot \rangle$ denotes the inner product and $\|\cdot\|$ stands for the norm associated with this inner product. Consider two elements u and v in H, and two real numbers α and β adding up to 1.

1°) Show that

$$\alpha \|u\|^2 + \beta \|v\|^2 = \|\alpha u + \beta v\|^2 + \alpha\beta \|u - v\|^2 . \tag{1}$$

2°) What does (1) say when $\alpha = \beta = \frac{1}{2}$?

3°) Let x and y be two non-null vectors in H. Prove that

$$(\|x+y\| = \|x\| + \|y\|) \Rightarrow (\|y\|\, x = \|x\|\, y)\,. \tag{2}$$

Hints. 1°) Use the basic calculus rule $\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\,\langle x, y\rangle$.
3°) Consider $u = \|y\|\, x, v = \|x\|\, y, \alpha = \frac{\|x\|}{\|x+y\|}, \beta = \frac{\|y\|}{\|x+y\|}$.

Answer. 2°) When $\alpha = \beta = \frac{1}{2}$, the relation (1) becomes

$$2\left(\|u\|^2 + \|v\|^2\right) = \|u+v\|^2 + \|u-v\|^2\,,$$

which is the so-called parallelogram rule.

Comment. 3°) By squaring both sides in $(\|x+y\| = \|x\| + \|y\|)$, one checks that the assumed equality is equivalent to $(\langle x, y\rangle = \|x\| \times \|y\|)$. Hence, one recovers with (2) an equality case in the CAUCHY–SCHWARZ inequality.

---

36. ★ *Calculus in inner product spaces*

Let H be an inner product space; $\langle\cdot,\cdot\rangle$ denotes the inner product and $\|\cdot\|$ the norm associated with this inner product. Consider two elements u and v in H, and two real numbers α and β. Show that

$$\begin{aligned} & \alpha(1-\alpha)\,\|\beta u + (1-\beta)v\|^2 + \beta(1-\beta)\,\|\alpha u + (1-\alpha)v\|^2 \\ = \;& A(\alpha,\beta)\,\|u\|^2 + B(\alpha,\beta)\,\|v\|^2\,, \end{aligned}$$

where $A(\alpha,\beta)$ and $B(\alpha,\beta)$ are polynomial functions of α and β.

Hint. $A(\alpha,\beta)$ and $B(\alpha,\beta)$ should be given in a form which is as "readable" as possible, that is to say, as polynomial expressions they should be as factorized as possible.

Answer. We have:

$$\begin{aligned} A(\alpha,\beta) &= \alpha\beta(\alpha+\beta-2\alpha\beta); \\ B(\alpha,\beta) &= (1-\alpha)(1-\beta)(\alpha+\beta-2\alpha\beta). \end{aligned}$$

Comment. As expected from the definitions of $A(\alpha,\beta)$ and $B(\alpha,\beta)$ themselves, one has the following symmetry relations:

$$\begin{aligned} A(\alpha,\beta) &= A(\beta,\alpha);\ B(\alpha,\beta) = B(\beta,\alpha); \\ A(1-\alpha,1-\beta) &= B(\alpha,\beta);\ B(1-\alpha,1-\beta) = A(\alpha,\beta). \end{aligned}$$

---

37. ★ *Two competing sequences in an inner product space*

Let H be an inner product space; $\langle\cdot,\cdot\rangle$ denotes the inner product and $\|\cdot\|$ the norm associated with it. Consider two sequences (u_n) and (v_n) in H satisfying:

$$\begin{cases} \|u_n\| \leqslant 1 \text{ and } \|v_n\| \leqslant 1 \text{ for all } n; \\ \langle u_n, v_n\rangle \to 1 \text{ when } n \to +\infty. \end{cases}$$

Are the sequences $(\|u_n\|), (u_n), (\|v_n\|), (v_n)$ convergent?

Hints. Firstly use the expansion

$$\|u_n - v_n\|^2 = \|u_n\|^2 + \|v_n\|^2 - 2\langle u_n, v_n\rangle$$

to show that $u_n - v_n \to 0$ when $n \to +\infty$.

Then, show that no limit value of $(\|u_n\|)$ and $(\|v_n\|)$ can be less than 1.

Answer. Both sequences $(\|u_n\|), (\|v_n\|)$ are convergent with the same limit; their common limit is 1.

However, the sequences (u_n) and (v_n) themselves may not be convergent.

---

38. ★ *Angles in inner product spaces*

Let H be an inner product space; $\langle\cdot,\cdot\rangle$ denotes the inner product and $\|\cdot\|$ stands for the norm associated with this inner product.

1°) Recall quickly why

$$-1 \leqslant \frac{\langle u, v\rangle}{\|u\| \times \|v\|} \leqslant 1 \qquad (1)$$

whenever u and v are two non-null elements in H.

For two non-null elements u and v in H, one defines the angle between them as:

$$\text{angle}(u,v) = \arccos\left(\frac{\langle u,v\rangle}{\|u\| \times \|v\|}\right). \tag{2}$$

2°) Let u and v be two elements in H satisfying $\|u\| = 1$, $\|v\| = 1$. Prove that

$$(\|u\| + \|v\| \geqslant 1) \Leftrightarrow \left(0 \leqslant \text{angle}(u,v) \leqslant \frac{2\pi}{3}\right). \tag{3}$$

Hints. 2°) Because u and v are unitary vectors, having $\|u\| + \|v\| \geqslant 1$ amounts to having $\langle u,v\rangle \geqslant -1/2$. Moreover, the arccos function establishes a continuous decreasing bijection from $[-1,1]$ onto $[0,\pi]$.

---

39. ★ *Successive projections on vector spaces*

Let V_1 and V_2 be two closed vector spaces in a HILBERT space H, such that $V_1 \subset V_2$.

1°) Prove that:

$$p_{V_1} \circ p_{V_2} = p_{V_1}. \tag{1}$$

Here, p_V denotes the projection operator onto V.

2°) Show with a simple counterexample that (1) does not hold true with two closed convex sets C_1 and C_2 such that $C_1 \subset C_2$.

Hint. 1°) Use the characterization of $\overline{x} = p_V(x)$ on a closed vector space V of H:

$$(\overline{x} = p_V(x)) \Leftrightarrow \left(\overline{x} \in V \text{ and } x - \overline{x} \in V^{\perp}\right).$$

Comment. The result (1) can be illustrated with a line V_1 contained in a plane V_2 of $\mathbb{R}^3$. Draw a picture to understand why the result in such a context is called (in high school) "the theorem of three prependiculars".

---

40. ★ *A simple example of a non-compact unit sphere in a* HILBERT *space*
Let $H = \ell^2(\mathbb{R})$ be structured as a HILBERT space with the inner product

$$\langle x, y\rangle = \sum_{n=1}^{+\infty} x_n y_n, \text{ for } x = (x_n)_{n\geqslant 1} \text{ and } y = (y_n)_{n\geqslant 1} \text{ in } H.$$

We denote by $\|\cdot\|$ the norm derived from $\langle\cdot,\cdot\rangle$.

Let $(e_n)_{n\geqslant 1}$ be the sequence in H defined as follows: for $n \geqslant 1$, $e_n = (0, ..., 0, 1, 0....)$ [1 at the n-th position, 0 everywhere else].

1°) Check that there is no subsequence extracted from $(e_n)_{n\geqslant 1}$ which is convergent.

2°) Deduce from the above that the unit sphere S of H is not compact.

Answer. 1°) If $(e_{n_k})_{k\geqslant 1}$ is an extracted subsequence from the sequence $(e_n)_{n\geqslant 1}$, we have

$$\|e_{n_l} - e_{n_k}\| = \sqrt{2} \text{ for } l > k.$$

So, there is no hope that the sequence $(e_{n_k})_{k\geqslant 1}$ is convergent.

---

41. ★ *A simple example of a non-compact unit sphere in a* BANACH *space*
Consider $E = \mathcal{C}([0,1], \mathbb{R})$ equipped with the norm $\|\cdot\|_\infty$. It is a BANACH space. For any positive integer n, let $f_n \in E$ be defined as follows:

$$f_n(t) \quad = \quad \begin{cases} 0 & \text{if } t \in [1/n, 1], \\ 1 - nt & \text{if } t \in [0, 1/n]. \end{cases}$$

1°) Show that no extracted subsequence $(f_{n_k})_{k\geqslant 1}$ from the sequence $(f_n)_{n\geqslant 1}$ is convergent.

2°) Deduce from the above that the unit sphere S of E is not compact.

Hints. - As often with counterexamples involving particular (simple) functions f_n, a first helpful task is to draw the graphical representation of such f_n.

- 1°) First possible proof. Show that $(f_{n_k})_{k\geqslant 1}$ cannot be a CAUCHY sequence. For that purpose, check that $\|f_{n_l} - f_{n_k}\| = 1 - n_k/n_l$; one therefore has $\|f_{n_l} - f_{n_k}\| \geqslant 1/2$ if n_l is chosen larger than $2n_k$.

A second possible proof. Should $(f_{n_k})_{k\geqslant 1}$ converge to some f in $(E, \|\cdot\|_\infty)$, f would also be the pointwise limit of $(f_{n_k})_{k\geqslant 1}$; but this pointwise limit turns out to be discontinuous at 0.

---

42. ★ *A general parallelogram rule in inner product spaces*

Let H be an inner product space; $\langle \cdot, \cdot \rangle$ denotes the inner product and $\|\cdot\|$ is the norm associated with this inner product. Let $x_1, x_2, ..., x_n$ be n elements in H. Show that

$$\sum_{(\varepsilon_1,...,\varepsilon_n)\in\{-1,1\}^n} \left\| \sum_{i=1}^{n} \varepsilon_i x_i \right\|^2 = 2^n \sum_{i=1}^{n} \|x_i\|^2 . \tag{1}$$

Hints. - A first possibility is to prove (1) by induction on n.

- A second proof consists in developing $\left\| \sum_{i=1}^{n} \varepsilon_i x_i \right\|^2$ as $\sum_{i,j=1}^{n} \varepsilon_i \varepsilon_j \langle x_i, x_j \rangle$. Then,

$$\sum_{(\varepsilon_1,...,\varepsilon_n)\in\{-1,1\}^n} \left\| \sum_{i=1}^{n} \varepsilon_i x_i \right\|^2 = \sum_{i,j=1}^{n} \langle x_i, x_j \rangle \sum_{(\varepsilon_1,...,\varepsilon_n)\in\{-1,1\}^n} \varepsilon_i \varepsilon_j .$$

Now, we have:

$$\sum_{(\varepsilon_1,...,\varepsilon_n)\in\{-1,1\}^n} \varepsilon_i \varepsilon_j \text{ equals } 2^n \text{ if } i = j, \text{ and equals 0 if } i \neq j.$$

Comment. When $n = 2$, the relation (1) is nothing else than

$$\begin{aligned} 4\left(\|u\|^2 + \|v\|^2\right) &= \|u+v\|^2 + \|-u-v\|^2 + \|u-v\|^2 + \|-u+v\|^2 , \\ \text{or } 2\left(\|u\|^2 + \|v\|^2\right) &= \|u+v\|^2 + \|u-v\|^2 , \end{aligned}$$

which is the so-called parallelogram rule.

---

43. ★ *A variational form of the* Riesz *representation theorem*

Let $(H, \langle \cdot, \cdot \rangle)$ be a (real) Hilbert space; $\|\cdot\|$ denotes the norm associated with the inner product $\langle \cdot, \cdot \rangle$. Let l be a continuous linear form on H, and let $\theta : H \to \mathbb{R}$ be defined as:

$$\theta(h) = \frac{\|h\|^2}{2} - l(h).$$

1°) Show that there is one and only one minimizer of the convex function θ on H. This unique minimizer is denoted by $\overline{u}$.

2°) Check that

$$l(h) = \langle \overline{u}, h \rangle \text{ for all } h \in H.$$

Hints. 1°) θ is a strictly convex (even quadratic) 1-coercive function on H [*i.e.,* satisfying the following property: $\theta(h) \to +\infty$ as $\|h\| \to +\infty$].

2°) Show that the gradient vector $\nabla\theta(\overline{u})$ of θ at $\overline{u}$ equals 0.

---

44. ★ *An explicit form for the element provided by the* RIESZ *representation theorem*

Let $(H, \langle \cdot, \cdot \rangle)$ be a (real) HILBERT space; let $(e_n)_{n \geqslant 1}$ be an orthonormal basis for H. Consider a continuous linear form l on H.

Prove that

$$\overline{u} = \sum_{n=1}^{+\infty} l(e_n) e_n$$

is the unique element in H satisfying:

$$l(h) = \langle \overline{u}, h \rangle \text{ for all } h \in H.$$

Hint. Use the decompositions of vectors in H on the orthonormal basis $(e_n)_{n \geqslant 1}$.

---

45. ★ *Various norms on* $\mathcal{C}([0,1], \mathbb{R})$

In the vector space $E = \mathcal{C}([0,1], \mathbb{R})$, consider, defined for any $f \in E$:

$$\begin{aligned} N_1(f) &= \int_0^1 |f(t)| \, \mathrm{d}t; \; N_2(f) = \int_0^1 [f(t)]^2 \, \mathrm{d}t; \\ N_3(f) &= \int_0^1 |f(t)| \, e^t \, \mathrm{d}t; \; N_4(f) = |f(0)| + \int_0^1 |f(t)| \, \mathrm{d}t. \end{aligned}$$

1°) We claim (and proofs are not asked for) that among these four functions, only three are norms. Indicate, with a simple justification, which one is not a norm.

2°) We claim that among the three remaining norms, two are equivalent, but not equivalent to the third one.

(a) Determine the two equivalent norms by proposing the best inequality constants.

(b) Justify the non-equivalence with the remaining norm.

Hint. 2°) (b). Use the following functions f_n :

$$f_n(t) \;=\; \begin{cases} 0 & \text{if } t \in [1/n, 1]\,, \\ -n^2 t + n & \text{if } t \in [0, 1/n]\,. \end{cases}$$

Answers. 1°) N_2 is not a norm, because $N_2(\lambda f)$ may differ from $|\lambda|\, N_2(f\ )$.

2°) (a). N_1 and N_3 are equivalent norms:

$$N_1(f) \leqslant N_3(f) \leqslant e N_1(f) \text{ for all } f \in E.$$

The inequalities above, comparing N_1 and N_3, are the sharpest ones.

(b) N_4 and N_1 (or N_3) are not equivalent norms. Indeed, for the proposed $f_n \in E$ (in Hint):

$$N_4(f_n) = n + \frac{1}{2};\ N_1(f_n) = \frac{1}{2}.$$

Comment. A "classical" result on normed vector spaces is as follows: If E is a finite-dimensional vector space, then all the norms on E are equivalent. It is interesting to note that the converse is also true; put another way: If E is an infinite-dimensional vector space, then there are norms on E which are not equivalent.

---

46. ★ *An unusual norm on* $\mathcal{C}^2([0,1]\,,\mathbb{R})$

Let $E = \left\{ f \in \mathcal{C}^2([0,1]\,,\mathbb{R}) : f(0) = f'(0) = 0 \right\}$ be equipped with the norm $\|\cdot\|_\infty$. We make another proposal with:

$$f \in E \mapsto N(f) = \|f + 2f' + f''\|_\infty\,.$$

1°) (a) Check that $N(\cdot)$ defines a norm on E, and that this norm is finer than $\|\cdot\|_\infty$, *i.e.*, there is a constant $C > 0$ such that

$$\|f\|_\infty \leqslant C \times N(f) \text{ for all } f \in E. \tag{1}$$

(b) Determine the sharpest constant C in (1).

2°) Are the norms $\|\cdot\|_\infty$ and N equivalent?

Hints. 1°) (a). A key-role will be played by the scalar second-order differential equation $y'' + 2y' + y = 0$.

(b) If g is a given continuous function, the unique $f \in E$ satisfying

$$f'' + 2f' + f = g$$

can be expressed as

$$f(x) = e^{-x} \int_0^x (x-t)e^t g(t)\ \mathrm{d}t.$$

2°) Consider the functions $f_n(x) = x^n/n$ for $n \geqslant 2$.

Answers. 1°) (b). A possible constant C in (1) is $1 - 2/e$. This is the sharpest one since, for $x \in [0,1] \mapsto f_0(x) = 1 - e^{-x}(1+x)$ (f_0 does indeed belong to E),

$$\|f_0\|_\infty = 1,\ N(f_0) = 1 - 2/e.$$

2°) The norms $\|\cdot\|_\infty$ and N are not equivalent; for example, for the functions $f_n(x) = x^n/n$ with $n \geqslant 2$, we have:

$$\|f_n\|_\infty = \frac{1}{n},\ N(f_n) = \frac{1}{n} + n + 1.$$

---

47. ★ *A closed bounded set of functions which is not compact*

Let $E = \mathcal{C}([0,1], \mathbb{R})$ be structured as a normed vector space with the help of the norm $\|\cdot\|_\infty$. We consider

$$S = \{f \in E : f(0) = 0,\ f(1) = 1, \text{ and } 0 \leqslant f(x) \leqslant 1 \text{ for all } x \in [0,1]\}.$$

1°) Check that S is a closed bounded set in E.

2°) We intend to prove that S is not compact by exhibiting a continuous function $I : E \to \mathbb{R}$, bounded from below on S, and whose infimum on S is not attained.

Let I be defined on E as:

$$I(f) = \int_0^1 [f(x)]^2\ \mathrm{d}x.$$

(a) Check that I is continuous on E.

(b) Show that

$$\inf_{f\in S} I(f) = 0.$$

(c) Deduce from the above that S is not compact.

Hint. 2°) (b). Consider the functions $f_n(x) = x^n$, with $n \geqslant 1$.

Answers. 1°) For the closedness property of S, just use the characterization of closedness of sets via sequences.

2°) (a). We have, for example,

$$\begin{aligned} |I(f) - I(g)| &\leqslant \|f+g\|_\infty \times \|f-g\|_\infty \\ &\leqslant 2\|f\|_\infty \times \|f-g\|_\infty + \|f-g\|_\infty^2 . \end{aligned}$$

(b) We have: $I(f) \geqslant 0$ for all f, and $I(f_n) = \frac{1}{2n+1}$. Thus,

$$\inf_{f\in S} I(f) = 0.$$

(c) It is impossible to have $I(f) = 0$ for some f in S.

---

48. ★ *A continuous linear form on* $(\mathcal{C}([a,b],\mathbb{R}), \|\cdot\|_\infty)$

Let $E = \mathcal{C}([a,b],\mathbb{R})$ be normed with the norm $\|\cdot\|_\infty$.

Let s and t be such that $a \leqslant s < t \leqslant b$. We consider

$$A : f \in E \mapsto A(f) = f(s) - f(t). \tag{1}$$

Thus, A is a non-null linear form on E.

1°) Show that A is continuous, and provide a bound from above for the norm $N(A)$ of A.

We recall that N, sometimes called the "operator norm", is defined as follows: $N(A)$ is the greatest lower bound of those $L \geqslant 0$ satisfying:

$$|A(f)| \leqslant L\|f\|_\infty \text{ for all } f \in E. \tag{2}$$

2°) By choosing a particular $f_0 \in E$ in the inequality (2) above, provide a lower bound for $N(A)$.

3°) Deduce from the above the exact value of $N(A)$.

4°) What can be said about $V = \{f \in E : f(s) = f(t)\}$ from the vectorial viewpoint? from the topological viewpoint?

Hint. 2°) Consider $f_0 \in E$ defined as follows:

$$f_0(x) \quad = \quad \begin{cases} 1 & \text{if } a \leqslant x \leqslant s, \\ -1 & \text{if } t \leqslant x \leqslant b, \\ \text{linear} & \text{when } x \in [s,t]\,. \end{cases}$$

Then, $\|f_0\|_\infty = 1$ and $A(f_0) = 2$ (this has been done for that!).

Answers. 1°) Clearly,

$$|A(f)| \leqslant 2\,\|f\|_\infty \text{ for all } f \in E.$$

Thus, $N(A) \leqslant 2$.

2°) Since $|A(f_0)| \leqslant N(A)\,\|f_0\|_\infty$, we have that $N(A) \geqslant 2$.

3°) We deduce from the results of questions 1°) and 2°) that $N(A) = 2$.

4°) V is the kernel of the non-null continuous linear form A; it is therefore a closed hyperplane in E.

Comment (in the same vein as the comment at the end of Tapa 45). A "classical" result on normed vector spaces is as follows: If E is a finite-dimensional vector space, then all the linear forms on E are continuous. It is interesting to note that the converse is also true; put another way: If E is an infinite-dimensional vector space, then there are linear forms on E which are not continuous.

---

49. ★ *Calculation of norms of operators acting on subspaces of* $\mathcal{C}([0,+\infty),\mathbb{R})$

For $\alpha \geqslant 0$, let E_α denote the family of continuous functions $f : [0,+\infty) \to \mathbb{R}$ satisfying

$$\sup_{t\in[0,+\infty)} e^{\alpha t}\,|f(t)| < +\infty. \tag{1}$$

E_α is a (real) vector space, which can be normed with the following norm:

$$f \in E_\alpha \mapsto \|f\|_\alpha = \sup_{t\in[0,+\infty)} e^{\alpha t}\,|f(t)|\,.$$

Let $\beta \geqslant 0$ and let $b \in E_\beta$. One then defines an operator $B : E_\alpha \to \mathcal{C}([0,+\infty),\mathbb{R})$ as follows:

$$\text{For } f \in E_\alpha,\ B(f) : t \geqslant 0 \mapsto [B(f)](t) = b(t)f(t). \tag{2}$$

1°) Check that, for all $f \in E_\alpha$, $B(f) \in E_{\alpha+\beta}$.

2°) Show that B is a continuous linear operator from E_α into $E_{\alpha+\beta}$.

3°) Determine the "operator norm" N of B.

Hint. 2°) Show that

$$\|B(f)\|_{\alpha+\beta} \leqslant \|b\|_\beta \times \|f\|_\alpha \text{ for all } f \in E_\alpha.$$

3°) Recall that $N(B)$ is the greatest lower bound of those $L \geqslant 0$ satisfying:

$$\|B(f)\|_{\alpha+\beta} \leqslant L\,\|f\|_\alpha \text{ for all } f \in E_\alpha. \tag{3}$$

To get a lower bound on $N(B)$, it may be helpful to consider the specific functions $f_\alpha : t \geqslant 0 \mapsto f_\alpha(t) = e^{-\alpha t}$. Then, $f_\alpha \in E_\alpha$ and $\|f_\alpha\|_\alpha = 1$.

Answer. 3°) We have: $N(B) = \|b\|_\beta$.

---

50. ★ *A non-continuous limit of continuous linear forms on* $(\mathcal{C}([0,1],\mathbb{R}),\|\cdot\|_1)$

Let $E = \mathcal{C}([0,1],\mathbb{R})$ be normed with the norm $\|\cdot\|_1$.

For a positive integer n, define

$$\begin{aligned}\varphi_n \quad &: \quad E \to \mathbb{R} \\ &\qquad f \mapsto \varphi_n(f) = n\int_0^{\frac{1}{n}} f(t)\ \mathrm{d}t.\end{aligned}$$

Also define

$$\begin{aligned}\varphi \quad &: \quad E \to \mathbb{R} \\ &\qquad f \mapsto \varphi(f) = f(0).\end{aligned}$$

All the φ_n as well as φ are non-null linear forms on E.

1°) Check with the help of a counterexample that φ is not continuous.

2°) Prove that φ_n is continuous on E and determine its norm $N(\varphi_n)$. We recall that N, sometimes called the "operator norm", is defined as follows: $N(\varphi_n)$ is the greatest lower bound of those $L \geqslant 0$ satisfying:

$$|\varphi_n(f)| \leqslant L\,\|f\|_1 \text{ for all } f \in E.$$

3°) (*a*) Show that, for all $f \in E$,

$$|\varphi_n(f) - \varphi(f)| \leqslant \sup_{t \in \left[0, \frac{1}{n}\right]} |f(t) - f(0)| \,. \tag{1}$$

(*b*) Deduce from the above that, for all $f \in E$,

$$\varphi_n(f) \to \varphi(f) \text{ as } n \to +\infty. \tag{2}$$

Hints. 1°) Consider $f_n \in E$ defined as:

$$f_n(t) \quad = \quad \left\{ \begin{array}{ll} 0 & \text{if } t \in [1/n, 1] \,, \\ -n^2 t + n & \text{if } t \in [0, 1/n] \,. \end{array} \right.$$

2°) To determine a lower bound for $N(\varphi_n)$, use a particular f defined similarly as above:

$$f(t) \quad = \quad \left\{ \begin{array}{ll} 0 & \text{if } t \in [1/n, 1] \,, \\ -nt + n & \text{if } t \in [0, 1/n] \,. \end{array} \right.$$

3°) (*a*). To obtain the inequality (1), note that $\varphi(f)$ can also be written as an integral: $\varphi(f) = n \int_0^{\frac{1}{n}} f(0) \, \mathrm{d}t$.

Answer. 2°) We have: $N(\varphi_n) = n$.

Comments. - Although the sequence of continuous linear forms $(\varphi_n)_{n \geqslant 1}$ on E converges pointwise towards the linear form φ on E (see (2)), φ is not continuous.

- As an application of a more advanced result in Functional analysis, called the Banach–Steinhaus theorem (itself resulting from the so-called "uniform boundedness principle", see Tapa 222), we have the following result: Let $(E, \|\cdot\|_E)$ be a Banach space and let $(F, \|\cdot\|_F)$ be a normed vector space; if (f_n) is a sequence of continuous linear mappings from E to F converging pointwise to a mapping $f : E \to F$, then f is linear (easy) but also continuous. So, in the context of our Tapa here, one assumption is missing to apply this result: $(E = \mathcal{C}([0, 1], \mathbb{R}), \|\cdot\|_1)$ is not a Banach space (it is not complete).

51. ★ *Norms of positive linear operators (1)*

Consider the vector space $E = \mathcal{C}([0,1], \mathbb{R})$ normed with the norm $\|\cdot\|_\infty$. We say that a linear mapping (or operator) $A : E \to E$ is *positive* when:

$$(f \geqslant 0) \Rightarrow (A(f) \geqslant 0).$$

1°) Prove that a positive linear mapping is necessarily continuous, and that its "operator norm" N is:

$$N(A) = \|A(f_1)\|_\infty,$$

where f_1 denotes the function which is constantly 1 on $[0,1]$.

2°) Let $\varphi : (x,t) \in [0,1] \times [0,1] \mapsto \varphi(x,t) \in \mathbb{R}^+$ be continuous, possessing a nonnegative continuous partial derivative $\frac{\partial \varphi}{\partial x} : [0,1] \times [0,1] \mapsto \frac{\partial \varphi}{\partial x}(x,t) \in \mathbb{R}^+$. Then define $A_\varphi : E \to E$ by

$$[A_\varphi(f)](x) = \int_0^1 \varphi(x,t) f(t)\, \mathrm{d}t.$$

(*a*) Check that A_φ is positive and that $N(A_\varphi) = \int_0^1 \varphi(x,1)\, \mathrm{d}t$.

(*b*) Example with $\varphi(x,t) = \exp(xt)$. What is $N(A_\varphi)$?

Hints. 1°) - Recall that $N(A)$ is the greatest lower bound of those $L \geqslant 0$ satisfying:

$$\|A(f)\|_\infty \leqslant L \|f\|_\infty \text{ for all } f \in E.$$

- Choose $f \in E$ satisfying $\|f\|_\infty \leqslant 1$, that is to say, $-f_1 \leqslant f \leqslant f_1$. The positivity of A means that $-A(f_1) \leqslant A(f) \leqslant A(f_1)$, whence $\|A(f)\|_\infty \leqslant \|A(f_1)\|_\infty$.

Answers. 2°) (*a*). The function $x \mapsto h(x) = \int_0^1 \varphi(x,t)\, \mathrm{d}t$ is nonnegative (because $\varphi(x,t) \geqslant 0$) and increasing; the latter property comes from the evaluation of the derivative of h: $h'(x) = \int_0^1 \frac{\partial \varphi}{\partial x}(x,t)\, \mathrm{d}t \geqslant 0$. Whence we derive that

$$\|A_\varphi(f_1)\|_\infty = \max_{x \in [0,1]} \int_0^1 \varphi(x,t) \mathrm{d}t = h(1).$$

(*b*) In this example we obtain $\|A_\varphi(f_1)\|_\infty = e - 1$.

52. ★ *Norms of positive linear operators (2)*

Let $a > 0$ and let $f : \mathbb{R}^+ \to \mathbb{R}$ be a continuous function. Define a new continuous function $A(f) : \mathbb{R}^+ \to \mathbb{R}$ as follows:

$$\text{For all } x > 0,\ [A(f)](x) = e^{-ax} \int_0^x f(t)e^{at}\ \mathrm{d}t. \tag{1}$$

1°) (*a*) Check that A is a positive linear operator from $\mathcal{C}(\mathbb{R}^+, \mathbb{R})$ into itself. Positive means that:

$$(f \geqslant 0) \Rightarrow (A(f) \geqslant 0).$$

(*b*) Verify that $|A(f)| \leqslant A(|f|)$.

We now add some assumptions on f and compare various norms on f and $A(f)$.

2°) Assume that f is bounded on $\mathbb{R}^+$; we set $\|f\|_\infty = \sup_{x\in\mathbb{R}^+} |f(x)|$. Show that $A(f)$ is bounded, that $\|A(f)\|_\infty \leqslant \frac{1}{a}\|f\|_\infty$, and prove that this inequality is sharp.

3°) Assume that $|f|$ is integrable on $\mathbb{R}^+$; we set $\|f\|_1 = \int_0^{+\infty} |f(x)|\ \mathrm{d}x$. Show that $A(f)$ is integrable, that $\|A(f)\|_1 \leqslant \frac{1}{a}\|f\|_1$, and prove that this inequality is sharp.

4°) Assume that the square of f is integrable on $\mathbb{R}^+$; we set $\|f\|_2 = \sqrt{\int_0^{+\infty} [f(x)]^2\ \mathrm{d}x}$. Show that the square of $A(f)$ is integrable, that $\|A(f)\|_2 \leqslant \frac{1}{a}\|f\|_2$, and prove that this inequality is sharp.

Hint. 4°) It may help to note that $A(f)$ is, in fact, the unique solution of the following linear Cauchy problem:

$$\begin{cases} y' + ay = f, \\ y(0) = 0. \end{cases}$$

---

53. ★ *Closures and interiors of vector subspaces in a normed vector space*

Let $(E, \|\cdot\|)$ be a normed vector space and let V be a vector subspace of E.

1°) (*a*) Show that the closure $\overline{V}$ of V is still a vector subspace of E.

(*b*) What can be said about the interior $\mathrm{int}V$ of V?

2°) Example. Let $E = \mathcal{C}([0,1], \mathbb{R})$ be normed either by $\|\cdot\|_1$ or by $\|\cdot\|_\infty$. Consider

$$V = \{f \in E : f(0) = 0\}.$$

(a) Show that V is a closed vector subspace of $(E, \|\cdot\|_\infty)$.

(b) Prove that $\overline{V} = E$ when E is equipped with the norm $\|\cdot\|_1$.

Hints. 1°) (a). Use the fact that any $x \in \overline{V}$ can be obtained as a limit of a sequence of elements x_n in V.

2°) (b). For $f \in E$, define $f_n \in V$ as follows:

$$f_n(t) \quad = \quad \begin{cases} f(t) & \text{if } t \in [1/n, 1]\,, \\ nf(1/n)t & \text{if } t \in [0, 1/n]\,. \end{cases}$$

Then, $\|f - f_n\|_1 \leqslant \frac{2}{n} \|f\|_\infty \to 0$ as $n \to +\infty$.

Answer. 1°) (b). The interior of V is empty except when V is the whole space E (in that case, $\text{int}E = E$).

Comment. Question 2° illustrates the only two possibilities for a hyperplane V in E (kernel of a non-null linear form on E): either it is closed or it is everywhere dense in E.

---

54. ★ *Topological properties of sums of sets in a normed vector space*

Let $(E, \|\cdot\|)$ be a normed vector space, let A and B be two subsets of E. We recall the definition of the sum $A + B$ of A and B:

$$A + B = \{a + b : a \in A, b \in B\}\,.$$

1°) Draw some examples in the plane (with disks, boxes, line-segments, etc. as the sets A and B) in order to familiarize yourself with this addition operation.

2°) Show that:

(i) If A is open, then $A + B$ is open
[without any property required on B] ;

(ii) If A is closed and B is compact, then $A + B$ is closed;

(iii) If A and B are closed, then $A + B$ is not necessarily closed;

(iv) If A and B are compact, then $A + B$ is compact;

(v) If A and B are connected, then so is $A + B$;

(vi) If A and B are bounded, then so is $A + B$.

Hints. 2°) (i). Either use the basic definition of open sets (an open set is a neighborhood of each of its points), or write $A+B$ as $\cup_{b\in B}\{A+b\}$ (a union of open sets).

2°) $(iv), (v)$. The set $A \times B$ is compact (resp. connected), and $A + B$ is the image of $A \times B$ under the continuous mapping $(u, v) \mapsto u + v$.

$(ii), (iv)$. Use characterizations, with the help of sequences, of closedness and compactness of sets.

Answer. 2°) (iii). Consider in the plane the two closed (even convex) sets

$$A = \left\{(x, y) \in \mathbb{R}^2 : x > 0 \text{ and } y \geqslant 1/x\right\} \text{ and } B = \{0\} \times \mathbb{R}.$$

Then $A + B = \{(x, y) : x > 0 \text{ and } y \in \mathbb{R}\}$ is not closed.

---

55. ★ *Topological properties of the cone generated by a set in a normed vector space*

Let $(E, \|\cdot\|)$ be a normed vector space and let S be a nonempty subset of E. We denote by $\text{cone}(S)$ the cone with apex 0 generated by S, that is

$$\text{cone}(S) = \{tx : t \geqslant 0 \text{ and } x \in S\}.$$

1°) Draw some examples in the plane (with disks, finite sets of points, curves, etc. as sets S) in order to familiarize yourself with this construction.

2°) (a) Assume that S is compact and does not contain 0. Show that $\text{cone}(S)$ is closed.

(b) Show with a counterexample that $\text{cone}(S)$ may not be closed if $0 \in S$.

3°) Assume that S is closed and does not contain 0. Is $\text{cone}(S)$ closed?

Hints. 2°) (a). Use the characterization of closedness with the help of sequences.

(b) Consider in the plane a disk S tangent to the x-axis at 0.

Answers. 2°) (b). In the plane, if S is a disk tangent to the x-axis at 0, then the non-null elements on the x-axis do not belong to $\mathrm{cone}(S)$. We thus have that $\mathrm{cone}(S)$ is not necessarily closed.

3°) The answer is No. Consider, for example,

$$S = \left\{ (x, y) \in \mathbb{R}^2 : y \geqslant e^{-|x|} \right\}.$$

Then $\mathrm{cone}(S)$ does not contain the non-null elements on the x-axis. Therefore, the set $\mathrm{cone}(S)$ is not closed.

---

56. ★ *The vector space of bounded continuous functions converging to* 0 *at infinity*

Let $\mathcal{C}_b(\mathbb{R})$ denote the vector space of functions $f : \mathbb{R} \to \mathbb{R}$ which are continuous and bounded on $\mathbb{R}$. We equip $\mathcal{C}_b(\mathbb{R})$ with the norm $\|\cdot\|_\infty$.

1°) Check that $(\mathcal{C}_b(\mathbb{R}), \|\cdot\|_\infty)$ is a BANACH space.

2°) Let now

$$V = \left\{ f \in \mathcal{C}_b(\mathbb{R}) : \lim_{|x| \to +\infty} f(x) = 0 \right\}.$$

(a) Show that V is a closed vector space of $\mathcal{C}_b(\mathbb{R})$.

(b) Is $(V, \|\cdot\|_\infty)$ a BANACH space?

Hint. 2°) (a). Use the characterization of closedness via sequences; here, if $(f_n) \subset V$ converges to f in $(\mathcal{C}_b(\mathbb{R}), \|\cdot\|_\infty)$, prove that $f \in V$.

Answer. 2°) (b). The answer is Yes, because a closed vector subspace of a BANACH space is itself a BANACH space.

---

57. ★ *Continuity (or not) of the evaluation form* $f \mapsto f(x_0)$ *on* $\mathcal{C}^1([0, \pi], \mathbb{R})$

Questions 1°), 2°) and 3°) are independent. Starting from question 4°), one could freely use the inequalities (1) and (2) below.

1°) (a) Indicate why there exist constants $c_1 > 0$ and $c_2 > 0$ such that:

$$\text{For all } a, b \in \mathbb{R}, \quad c_1 \sqrt{a^2 + b^2} \leqslant |a| + |b| \leqslant c_2 \sqrt{a^2 + b^2}. \qquad (1)$$

(b) Give examples of such constants.

(c) Determine the best possible such constants (that is to say, the largest c_1 and the smallest c_2).

2°) With the help of the CAUCHY–SCHWARZ inequality, applied in the context of a prehilbertian space to be specified, show that:

$$\text{For all } f \in \mathcal{C}([0,\pi],\mathbb{R}), \qquad \int_0^\pi |f(x)|\,\mathrm{d}x \leqslant \sqrt{\pi} \times \sqrt{\int_0^\pi [f(x)]^2\,\mathrm{d}x}. \quad (2)$$

3°) On the vector space $E = \mathcal{C}^1([0,\pi],\mathbb{R})$, define the following symmetric bilinear form:

$$\langle f, g\rangle_1 = f(0)g(0) + \int_0^\pi f'(x)g'(x)\mathrm{d}x.$$

Check that $\langle\cdot,\cdot\rangle_1$ is a scalar product on E.

4°) Given x_0 in $[0,\pi]$, the so-called evaluation form at x_0 is defined as follows:

$$\delta_{x_0} : f \in E \mapsto \delta_{x_0}(f) := f(x_0) \in \mathbb{R}.$$

δ_{x_0} is indeed a non-null linear form on E. In this question, we intend to prove that δ_{x_0} is continuous on $(E, \|\cdot\|_1)$, where $\|\cdot\|_1 = \sqrt{\langle\cdot,\cdot\rangle_1}$ designates the norm derived from the scalar product $\langle\cdot,\cdot\rangle_1$.

(a) Prove the following inequality:

$$|f(x_0)| \leqslant |f(0)| + \int_0^\pi |f'(x)|\,\mathrm{d}x. \quad (3)$$

(b) With the help of inequalities (1), (2), (3), prove that there exists a constant $L > 0$ such that:

$$\text{For all } f \in E, \quad |f(x_0)| \leqslant L\,\|f\|_1. \quad (4)$$

(c) What can be deduced from this concerning the linear form δ_{x_0}?

5°) Let

$$V = \{f \in E\ ;\ f(x_0) = 0\}.$$

(a) What can be said about V from the vectorial viewpoint? from the topological viewpoint?

(b) We choose $x_0 = 0$ here.

- Propose a non-null element g in the orthogonal vector subspace $V^\perp$ of V.

- Determine $V^\perp$ completely.

Hints. 2°) Equip $\mathcal{C}([0,\pi],\mathbb{R})$ with the scalar product $\langle f, g\rangle = \int_0^\pi f(t)g(t)\,\mathrm{d}t$, and apply the CAUCHY–SCHWARZ inequality to $|f|$ and $g = 1$.

4°) (*a*). Use the evaluation $f(x) = f(0) + \int_0^{x_0} f'(x)\,\mathrm{d}x$.

4°) (*b*). Apply the inequality (1) with $a = |f(0)|$ and $b = \int_0^{\pi} |f'(x)|\,\mathrm{d}x$. Besides that, apply the inequality (2) to the function f' and square the resulting inequality.

Answers. 1°) (*a*). This follows from the equivalence of norms on $\mathbb{R}^2$.

(*c*) The optimal c_1 is 1, while the optimal c_2 is $\sqrt{2}$.

4°) (*b*). One obtains $L = \sqrt{2\pi}$.

(*c*) The linear form δ_{x_0} is continuous on $(E, \|\cdot\|_1)$.

Note that δ_{x_0} is not continuous if E is just endowed with the norm derived from the scalar product $\langle f, g\rangle = \int_0^{\pi} f(t)g(t)\,\mathrm{d}t$.

5°) (*a*). V is the kernel of the non-null continuous linear form δ_{x_0}; it is therefore a closed hyperplane in E (in the topology defined via $\|\cdot\|_1$).

(*b*) For example, the constant function $g = 1$ is orthogonal (in the $\langle\cdot,\cdot\rangle_1$ sense) to any element in V.

The orthogonal space $V^{\perp}$ is the vector line directed by the g function.

---

58. ★ *The vector space of* Lipschitz *continuous functions on* $[0, 1]$

We denote by E the family of functions $f : [0, 1] \to \mathbb{R}$ satisfying a Lipschitz property on $[0, 1]$. Clearly, E is a vector subspace of $\mathcal{C}([0, 1], \mathbb{R})$. For $f \in E$, we set

$$L(f) = \inf\{l \geqslant 0 : |f(x) - f(y)| \leqslant l\,|x - y| \text{ for all } x, y \text{ in } [0, 1]\};$$

(in words: for $f \in E$, $L(f)$ is the best (*i.e.,* smallest) Lipschitz constant).

1°) Does $L(\cdot)$ define a norm on E?

2°) We intend to decide whether $(E, \|\cdot\|_\infty)$ is complete or not. For that purpose, we consider a sequence $(f_n)_{n\geqslant 1}$ of functions defined as follows:

$$f_n : x \in [0, 1] \mapsto f_n(x) = \sqrt{x + \frac{1}{n}}.$$

(*a*) Show that $(f_n)_{n\geqslant 1}$ converges uniformly on $[0, 1]$ towards a function $g \in \mathcal{C}([0, 1], \mathbb{R})$ to be determined.

(*b*) Check that all the functions f_n belong to E. Does g belong to E?

(*c*) Show that $(f_n)_{n\geqslant 1}$ is a Cauchy sequence (in $(E, \|\cdot\|_\infty)$).

(*d*) Deduce from the above whether $(E, \|\cdot\|_\infty)$ is complete or not.

3°) Consider another proposal for a norm on E,

$$N(f) = \|f\|_\infty + L(f).$$

(a) Check that $N(\cdot)$ is indeed a norm on E.

(b) Prove that (E, N) is complete.

Hints. 2°) Use the inequality: $\sqrt{u+v} \leqslant \sqrt{u} + \sqrt{v}$ for all u, v in $[0,1]$.

3°) (b). It may help to use the property that $(\mathcal{C}([0,1], \mathbb{R}), \|\cdot\|_\infty)$ is complete.

Answers. 1°) The answer is No. For example, $L(f) = 0$ if $f(x) = 1$ for all $x \in [0,1]$.

2°) (a). The limit function is $g(x) = \sqrt{x}$.

(b) The function g does not satisfy a LIPSCHITZ property on $[0,1]$.

(c) We have for $m \geqslant n$: $\max_{x\in[0,1]} |f_m(x) - f_n(x)| \leqslant 1/(2m\sqrt{n})$. Hence, $(f_n)_{n\geqslant 1}$ is a CAUCHY sequence.

(d) $(E, \|\cdot\|_\infty)$ is not complete.

---

59. ★ *Minimizing an integral functional over a subset of* LIPSCHITZ *functions*

Let S denote the subset of $\mathcal{C}([0,1], \mathbb{R})$ consisting of functions $f : [0,1] \to \mathbb{R}$ which are 1-LIPSCHITZ and satisfy $f(0) = 0$. For $f \in S$, let

$$I(f) = \int_0^1 \left[f^2(x) - f(x)\right] \mathrm{d}x.$$

Solve the next optimization problem:

$$\text{Minimize } I(f) \text{ subject to } f \in S. \tag{1}$$

Hints. - Check that if $f \in S$, then $-x \leqslant f(x) \leqslant x$ for all $x \in [0,1]$.

- The function $u \in \mathbb{R} \mapsto u^2 - u$ is decreasing on $(-\infty, 1/2]$, negative on $(0,1)$, and minimized at $\overline{u} = 1/2$. So, intuitively, the minimal value in (1) should be attained by a function $\overline{f} \in S$ which approaches the value $1/2$ as fast as possible and stays there.

Answer. A solution to the optimization problem (1) is

$$\overline{f} : x \mapsto \left\{ \begin{array}{ll} x & \text{if } 0 \leqslant x \leqslant 1/2, \\ 1/2 & \text{if } 1/2 \leqslant x \leqslant 1. \end{array} \right.$$

The optimal value in problem (1) is $I(\overline{f}) = -5/24$.

---

60. ★ *An application of the fixed point theorem*

Let $E = \mathcal{C}([0,1], \mathbb{R})$ be equipped with the norm $\|\cdot\|_\infty$, so that $(E, \|\cdot\|_\infty)$ is a BANACH space.

1°) Consider the linear operator $A : E \to E$ defined as follows:

$$\text{For } f \in E,\ A(f) : x \in [0,1] \mapsto [A(f)](x) = \frac{1}{2} \int_0^{x^2} f(t)\ \mathrm{d}t. \qquad (1)$$

Show that A is continuous and determine its "operator norm" $N(A)$.

2°) Let $B : E \to E$ be defined as follows:

$$\text{For } f \in E,\ B(f) : x \in [0,1] \mapsto [B(f)](x) = e^x + \frac{1}{2} \int_0^{x^2} f(t)\ \mathrm{d}t. \qquad (2)$$

(a) Check that B is a contraction on E.

(b) Deduce from the above that there exists one and only one $f \in E$ such that $B(f) = f$ (*i.e.*, a fixed point of B).

3°) Consider the functional equation $(\mathcal{P})$ below:

$$(\mathcal{P}) \left\{ \begin{array}{c} \text{Find } f \in \mathcal{C}^1([0,1], \mathbb{R}) \text{ such that} \\ \left\{ \begin{array}{c} f'(x) = e^x + x f(x^2) \text{ for all } x \in [0,1], \\ f(0) = 1. \end{array} \right\} \end{array} \right.$$

(a) Calculate the derivative of the function $x \mapsto \theta(x) = \int_0^{x^2} f(t)\,\mathrm{d}t$, $f \in E$.

(b) Deduce from the above that $(\mathcal{P})$ is equivalent to the following problem $(\mathcal{P}')$:

$$(\mathcal{P}') \left\{ \begin{array}{c} \text{Find } f \in \mathcal{C}([0,1], \mathbb{R}) \text{ such that} \\ \left\{ f(x) = e^x + \frac{1}{2} \int_0^{x^2} f(t)\ \mathrm{d}t \text{ for all } x \in [0,1]. \right\} \end{array} \right.$$

(c) What can be deduced concerning the existence and the uniqueness of solutions to problem $(\mathcal{P})$?

Hints. 1°) Recall that $N(A)$ is the greatest lower bound of those $L \geqslant 0$ satisfying:

$$\|A(f)\|_\infty \leqslant L \|f\|_\infty \text{ for all } f \in E.$$

Answers. 1°) We have: $N(A) = \frac{1}{2}$.

2°) (a). We have:

$$\|B(f) - B(g)\|_\infty \leqslant \frac{1}{2} \|f - g\|_\infty \text{ for all } f \text{ and } g \text{ in } E.$$

B is therefore a contraction on E.

(b) This is an application of the classical fixed point theorem for contractions on complete metric spaces.

3°) (a). We have:

$$\theta'(x) = 2xf(x^2). \tag{3}$$

(b) In view of the relation (3) above:

- If f is a solution of $(\mathcal{P})$, $x \mapsto f(x)$, expressed as $f(x) = f(0) + \int_0^x f'(t)\mathrm{d}t$, is a solution of $(\mathcal{P}')$.

- If f is a solution of $(\mathcal{P}')$, $f \in \mathcal{C}^1([0,1], \mathbb{R})$ and f satisfies all the requirements in $(\mathcal{P})$.

(c) With what has been shown in (b) above, we conclude that the functional equation $(\mathcal{P})$ has one and only one solution.

---

61. ★ *Transformations of circles and lines by an inversion*

Let $(H, \langle \cdot, \cdot \rangle)$ be a (real) HILBERT space; $\|\cdot\|$ denotes the norm associated with the inner product $\langle \cdot, \cdot \rangle$. Let U be the open set of non-null elements x in H. Consider the so-called *inversion operation* F defined as follows:

$$\begin{aligned} F \quad : \quad & U \to U \\ & x \mapsto F(x) = \frac{x}{\|x\|^2}. \end{aligned}$$

1°) Let $a \neq 0$ and $V_a = \{x \in U : \langle a, x \rangle = 1\}$. Show that

$$F(V_a) = \left\{ y \in U : \left\| y - \frac{a}{2} \right\| = \frac{\|a\|}{2} \right\}.$$

2°) Let $a \neq 0$ and $C_a = \{x \in U : \|y - a\| = \|a\|\}$. Show that

$$F(C_a) = \left\{ y \in U : \langle a, y \rangle = \frac{1}{2} \right\}.$$

3°) Let $a \neq 0$ and $C_r = \{x \in U : \|y - a\| = r\}$, where $r > 0$ differs from $\|a\|$. Show that

$$F(C_r) = \left\{ y \in U : \left\| y - \frac{a}{\varepsilon} \right\| = \frac{r}{|\varepsilon|} \right\},$$

where $\varepsilon = \|a\|^2 - r^2$.

Comments. - The inversion operation F is an *involution* on U, *i.e.* satisfies $F \circ F = \mathrm{id}_U$, or $F = F^{-1}$. As a general rule, one can also check that:

$$\text{For all } x, y \text{ in } U, \quad \|F(x) - F(y)\| = \frac{\|x - y\|}{\|x\| \times \|y\|}. \tag{1}$$

The latter relation can be used to prove (for all x, y, z in H):

$$\|x\| \times \|y - z\| \leqslant \|z\| \times \|x - y\| + \|y\| \times \|z - x\|$$

and (for all t, u, v, w in H)

$$\|t - u\| \times \|w - v\| \leqslant \|t - v\| \times \|u - w\| + \|t - w\| \times \|v - u\|. \tag{2}$$

Inequality (2) is called Ptolemy's inequality because, in planar (Euclidean) geometry, for any convex quadrilateral $ABCD$, the Greek scientist Ptolemy (160–168) proved that

$$\left\|\overrightarrow{AC}\right\| \times \left\|\overrightarrow{BD}\right\| \leqslant \left\|\overrightarrow{AB}\right\| \times \left\|\overrightarrow{CD}\right\| + \left\|\overrightarrow{BC}\right\| \times \left\|\overrightarrow{AD}\right\|. \tag{3}$$

- Even in the context of the usual Euclidean plane, working with this specific transformation (the inversion) is very instructive (transforming lines into circles and vice versa, preserving angles between vectors, etc.); when I was young student, it was an important point in preparation for the Baccalauréat examination (in French high schools).

---

62. ★ *Inversion from the differential calculus viewpoint*

The context is the same as in the previous proposal.

1°) Prove that F is differentiable on U and give the expression of the differential $DF(x)$ of F at any $x \in U$.

2°) (a) Prove that F is a bijection from U onto U, actually a diffeomorphism of class $\mathcal{C}^\infty$ from U onto U.

(b) What is $[DF(x)]^{-1}$ ($\in \mathrm{Isom}(H)$)?

Answers. 1°) For $x \in U$, we have:

$$DF(x) : h \in H \mapsto DF(x)(h) = \frac{h}{\|x\|^2} - 2\frac{\langle x, h\rangle}{\|x\|^4} x.$$

2°) (b). We have:

$$[DF(x)]^{-1} : k \in H \mapsto [DF(x)]^{-1}(k) = -2\langle x, k\rangle x + \|x\|^2 k.$$

---

63. ★ *An application of the criterion of increasing slopes for convexity*

Let I be an open interval of $\mathbb{R}$ and let $f : I \to \mathbb{R}$ be a function.

Besides the definition itself, one of the simplest criteria for convexity (resp. strict convexity) of f is the following (called the criterion of increasing slopes): f is convex (resp. strictly convex) if and only if, for each $a \in I$, the slope-function

$$x \mapsto s_a(x) = \frac{f(x) - f(a)}{x - a} \tag{1}$$

is increasing (resp. strictly increasing) on $I \setminus \{a\}$. Moreover, if f is differentiable at a, one necessarily has:

$$s_a(x) \geqslant f'(a) \text{ (resp. } > f'(a)) \text{ for all } x \in I, x \neq a. \tag{2}$$

If f is twice differentiable on I, f is strictly convex on I whenever

$$f''(x) > 0 \text{ for all } x \in I. \tag{3}$$

1°) Let $f : I = (-1, +\infty) \to \mathbb{R}$ be defined by $f(x) = -\ln(1 + x)$.

Check that f is strictly convex on I.

2°) Using the criterion of increasing slopes for f with $a = 0$, show that

$$\left(1+\frac{1}{n}\right)^n < \left(1+\frac{1}{n+1}\right)^{n+1} < e \text{ for all integers } n \geqslant 1. \qquad (4)$$

Answers. 1°) We have: $f''(x) = 1/(1+x^2) > 0$ for all $x \in I$.

2°) With $a = 0$, the slope function for f is

$$s_0(x) = -\frac{\ln(1+x)}{x}.$$

We therefore have that

$$f'(0) = -1 < s_0\left(\frac{1}{n+1}\right) < s_0\left(\frac{1}{n}\right) \text{ for all integers } n \geqslant 1.$$

This leads directly to (4).

---

64. ★ *Convexity of the* $\Gamma \log \Gamma$ *function*

EULER's gamma function (denoted by Γ) is convex on $(0, +\infty)$; the same holds true for the function $\log \Gamma$.

1°) Another property in the same vein: prove that the product function $\Gamma \log \Gamma$ is convex on $(0, +\infty)$.

2°) Deduce from the above property the following inequality: for positive $a_1, a_2, ..., a_n$ and their mean value $\overline{a} = \frac{a_1+a_2...+a_n}{n}$,

$$\Gamma(a_1)^{\Gamma(a_1)}\Gamma(a_2)^{\Gamma(a_2)}... \Gamma(a_n)^{\Gamma(a_n)} \geqslant \Gamma(\overline{a})^{n\Gamma(\overline{a})}. \qquad (1)$$

---

65. ★ *An unusual convexity criterion for a function of a real variable*

Let $f : I \to \mathbb{R}$ be continuous on the open interval I.

Prove that f is convex on I if and only if for each line-segment $[a, b] \subset I$ and any real number r, the maximum of the function $f_r : x \in [a, b] \mapsto f_r(x) = f(x) - rx$ on $[a, b]$ is achieved either in a or in b.

Hint. (The announced condition is sufficient). Given $a < b$ in I, choose for r the mean value $[f(b) - f(a)]/(b - a)$, so that $f_r(a) = f_r(b)$. By assumption,

$$f_r[\lambda a + (1-\lambda)b] \leqslant f_r(a) = f_r(b) \text{ for all } \lambda \in [0, 1].$$

This leads to

$$f[\lambda a + (1-\lambda)b] \leqslant \lambda f(a) + (1-\lambda) f(b) \text{ for all } \lambda \in [0,1]\,.$$

Comments. - This criterion of convexity is helpful when some other basic criteria are inefficient, for example when proving the following propositions. We choose $I = \mathbb{R}$ for the sake of simplicity. Then:

• f is convex on $\mathbb{R}$ if and only if for all $x \in \mathbb{R}$ and $h > 0$,

$$f(x) \leqslant \frac{1}{2h} \int_{x-h}^{x+h} f(t)\ \mathrm{d}t. \tag{1}$$

• f is convex on $\mathbb{R}$ if and only if for all $x \in \mathbb{R}$ and $\varepsilon > 0$, there exists an $h \in (0, \varepsilon)$ such that

$$f(x) \leqslant \frac{1}{2}\left[f(x+h) + f(x-h)\right]. \tag{2}$$

- When f is convex, the two functions appearing in the right-hand sides of (1) and (2) are also convex.

---

66. ★ *Gradient of the* Rayleigh *quotient*

Let $A \in \mathcal{S}_n(\mathbb{R})$ and let f_A be the associated Rayleigh quotient function, that is:

$$f_A : x \neq 0 \in \mathbb{R}^n \mapsto f_A(x) = \frac{x^T A x}{\|x\|^2}. \tag{1}$$

Calculate the gradient $\nabla f_A(x)$ at any point $x \neq 0$, and compare it with the orthogonal projection of the vector Ax onto the hyperplane H_x orthogonal to x.

Hint. Check that

$$Ax = [Ax - f_A(x)x] + f_A(x)x$$

is the orthogonal decomposition of the vector Ax following the subspace H_x and the line $H_x^\perp = \mathbb{R}x$ directed by x.

Answer. We have:

$$\nabla f_A(x) = \frac{2}{\|x\|^2} \left[Ax - f_A(x)x \right], \tag{2}$$

which is, but for the coefficient $2/\|x\|^2$, the orthogonal projection of Ax onto the hyperplane H_x.

Comments. - The function f_A is constant along the lines directed by the vector $x \neq 0$; it is therefore normal that $\nabla f_A(x)$ be orthogonal to x.

- This Tapa is another opportunity to recall the variational formulations of the extreme eigenvalues of A in terms of the RAYLEIGH quotient function f_A :

$$\lambda_{\max}(A) = \sup_{x \neq 0} f_A(x);\ \lambda_{\min}(A) = \inf_{x \neq 0} f_A(x).$$

Moreover, in view of (2), the critical points of f_A, *i.e.,* those x for which $\nabla f_A(x) = 0$, are eigenvectors for A.

---

67. ★ *Differential calculus and applications to a minimization problem*

Let a and b be two different points in $\mathbb{R}^n$ (in the plane $\mathbb{R}^2$, for example) and let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f(x) = \|x - a\|^2 \times \|x - b\|^2,$$

where $\|\cdot\|$ denotes the usual Euclidean norm in $\mathbb{R}^n$.

1°) (a) Give a simple reason why f is of class $\mathcal{C}^\infty$ on $\mathbb{R}^n$.

(b) Determine the gradient vector $\nabla f(x)$ and the Hessian matrix $Hf(x)$ of f at any point $x \in \mathbb{R}^n$.

2°) (a) Determine the critical (or stationary) points of f.

(b) Determine the nature of each of these critical points (local minimizer, local maximizer, saddle-point).

Hints. 1°) (b). It is advisable to use first-order and second-order expansions of f around x.

2°) (a). There are three critical points, two are evident and the third one lies on the line joining a and b.

Answers. 1°) (b). We have:

$$\begin{aligned} \nabla f(x) &= 2\|x-a\|^2(x-b) + 2\|x-b\|^2(x-a); \\ 2Hf(x) &= 4(x-a)(x-b)^T + \|x-a\|^2 I_n + \|x-b\|^2 I_n. \end{aligned}$$

2°) There are three critical points: two minimizers a and b, one saddle-point $\frac{a+b}{2}$.

---

68. ★ *Minimization of a sum of products under a constraint on the sum*

Let $a_1 \leqslant a_2 \leqslant ... \leqslant a_n$ be nonnegative real numbers ($n \geqslant 2$) such that

$$a_1a_2 + ... + \ a_ia_{i+1} + ... + \ a_na_{n+1} = 1.$$

Determine the minimum value of $\sum_{i=1}^n a_i$.

Hints. - Make use of the necessary conditions provided by the LAGRANGE (or EULER–LAGRANGE) multipliers rule.

- Three cases have to be distinguished: $n = 2$ or 3, $n = 4$, and $n \geqslant 5$.

Answers. For $n = 2$ or 3, the minimal value of $\sum_{i=1}^n a_i$ is $\sqrt{n}$ and occurs when each a_i equals $1/\sqrt{n}$.

For $n = 4$, the minimal value of $\sum_{i=1}^n a_i$ is 2 and is attained when $a_1 = a_2, a_3 = a_4, a_1 + a_3 = 1$.

For $n \geqslant 5$, the minimal value of $\sum_{i=1}^n a_i$ is 2 and is achieved when $a_1 = a_2 = ... = a_{n-2} = 0, a_{n-1} = a_n = 1$.

---

69. ★ *Partial minimization of a quadratic form. Application to* BERGSTRÖM's *inequality*

Let $A \in \mathcal{S}_n(\mathbb{R})$ be a positive definite matrix and let $q : x \in \mathbb{R}^n \mapsto q(x) = x^TAx$ be the associated quadratic form. Consider the following minimization problem:

$$(\mathcal{P}) \ \left\{ \begin{array}{c} \text{Minimize } q(x) \\ \text{subject to } x_1 = 1, \end{array} \right.$$

where x_1 denotes the first component of $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$.

1°) Determine the optimal value in $(\mathcal{P})$.

2°) Use the form of the optimal value obtained above to prove the following inequality: if A and B are two positive definite matrices, then

$$\frac{\det(A+B)}{\det(A+B)_i} \geqslant \frac{\det A}{\det A_i} + \frac{\det B}{\det B_i}, \tag{1}$$

where $\det A_i$ stands for the cofactor of entry (i,i) in the matrix A.

Hints. 1°) - $(\mathcal{P})$ is a fairly simple convex minimization problem; therefore use characterizations of solutions provided by the LAGRANGE optimality conditions.

- The optimal value in $(\mathcal{P})$ is expressed with the help of entry $(1,1)$ of A^{-1}, which is actually $(\det A_1)\,/\,(\det A)$.

2°) Make use of the following inequality:

$$\min_{x_1=1} x^T(A+B)x \geqslant \min_{x_1=1} x^T A x + \min_{x_1=1} x^T B x. \tag{2}$$

Answers. 1°) The optimal value in $(\mathcal{P})$ is

$$\frac{1}{(A^{-1})_{1,1}} = \frac{\det(A)}{\det(A_1)}. \tag{3}$$

2) The form of the obtained optimal value in $(\mathcal{P})$, combined with the basic inequality (2), immediately gives rise to the proposed inequality (1) with $i=1$. The result is easily extended to any index $i \in \{1,2,...,n\}$.

Comment. Inequality (1) is known in the context of Matrix analysis as BERGSTRÖM's inequality (1949).

---

70. ★ *A convex function of eigenvalues of positive definite matrices*

Let $p \geqslant 0$ and $f_p : A \succ 0 \mapsto f_p(A) := \frac{[\lambda_{\max}(A)]^{p+1}}{[\lambda_{\min}(A)]^p}$, where $\lambda_{\max}(A)$ (resp. $\lambda_{\min}(A)$) stands for the largest (resp. the smallest) eigenvalue of the (symmetric) positive definite matrix A. The function f_p is clearly positively homogeneous of degree 1, *i.e.*, $f_p(\alpha A) = \alpha f_p(A)$ for all $\alpha > 0$ and $A \succ 0$.

Prove that f_p is a convex function.

Hints. - The function $A \succ 0 \mapsto \lambda_{\max}(A) > 0$ is convex, while the function $A \succ 0 \mapsto \lambda_{\min}(A) > 0$ is concave.

- A positively homogeneous function f of degree 1 is convex provided that it is sub-additive, *i.e.*, satisfies:

$$f(a+b) \leqslant f(a) + f(b) \text{ for all } a \text{ and } b.$$

Comments. - When $p \to \infty$, $(f_p)^{\frac{1}{p}}(A) \to c(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$ (called the *condition number* of $A \succ 0$). Hence, a fairly complicated function, namely the condition number c of matrices, can be approximated by $1/p$ powers of convex functions.

- See Tapa 200 for results in the same vein as this Tapa 70.

---

71. ★ *When pointwise convergence implies uniform convergence*

Let K be a compact subset of a metric space (E, d). We consider the set $\mathcal{F}$ of functions $f : K \to \mathbb{R}$ which are L-LIPSCHITZ on K, *i.e.,* satisfying:

$$|f(x) - f(y)| \leqslant Ld(x, y) \text{ for all } x, y \text{ in } K. \tag{1}$$

Consider a sequence (f_n) of functions in $\mathcal{F}$ converging pointwise to a function $f : K \to \mathbb{R}$.

1°) Check that the limit function f is in $\mathcal{F}$.

2°) Prove that the convergence of (f_n) towards f is uniform on K.

Hint. 2°) Proof by contradiction. Start with a sequence $(x_n) \subset K$ and $\varepsilon > 0$ such that

$$|f_n(x_n) - f(x_n)| \geqslant \varepsilon \text{ for all } n.$$

---

72. ★ OPIAL'S *inequality in a* HILBERT *space*

Let $(H, \langle \cdot, \cdot \rangle)$ be a (real) HILBERT space; $\|\cdot\|$ denotes the norm associated with the inner product $\langle \cdot, \cdot \rangle$. We suppose that the sequence (u_n) in H converges weakly towards $u \in H$.

Prove that for all $v \in H$, $v \neq u$, we have:

$$\liminf_{n \to +\infty} \|u_n - v\| > \liminf_{n \to +\infty} \|u_n - u\|. \tag{1}$$

Hint. Start with the classical decomposition

$$\begin{aligned}\|u_n - v\|^2 &= \|(u_n - u) + (u - v)\|^2 \\ &= \|u_n - u\|^2 + \|u - v\|^2 + 2\langle u_n - u, u - v\rangle .\end{aligned}$$

Comment. Inequality (1) bears the name of the Polish mathematician Z. Opial (1967).

---

73. ★ *Inverting reverses the order of positive definite operators*

Let $(H, \langle\cdot,\cdot\rangle)$ be a (real) Hilbert space. A linear operator $A : H \to H$ is said to be self-adjoint if $\langle Ax, y\rangle = \langle x, Ay\rangle$ for all x, y in H; it is said to be positive definite if $\langle Ax, x\rangle > 0$ for all $x \neq 0$ in H, and it is positive semidefinite if $\langle Ax, x\rangle \geqslant 0$ for all in H.

Consider two linear operators A and B from H into H, both assumed to be bijective, self-adjoint and positive definite.

1°) Prove that

$$(A - B \text{ is positive semidefinite}) \Rightarrow \left(B^{-1} - A^{-1} \text{ is positive semidefinite}\right). \tag{1}$$

2°) How do things simplify when $H = \mathbb{R}^n$ and A, B are symmetric positive definite matrices?

Hint. 1°) Start with the relation

$$\left\langle B(y - B^{-1}x), y - B^{-1}x\right\rangle \geqslant 0 \text{ for all } x, y \text{ in } H$$

and make successive use of the assumptions : $A - B$ is positive semidefinite, A is bijective, A is self-adjoint.

Answer. 2°) A positive definite symmetric matrix A is automatically invertible.

There are several ways of proving (1) with positive definite matrices A and B: using Matrix analysis, Optimization, etc. Here is one. If $x \mapsto q_S(x) = \frac{1}{2}x^T Sx$ is a quadratic form associated with the positive definite matrix S, it is easy to carry out the following calculation:

$$\max_{x\in\mathbb{R}^n}\left[y^T x - q_S(x)\right] = q_{S^{-1}}(y) = \frac{1}{2}y^T S^{-1} y. \tag{2}$$

The assumption in (1) is that $q_A \geqslant q_B$; the conclusion in (1) is that $q_{B^{-1}} \geqslant q_{A^{-1}}$, which immediately follows from (2).

Comments. - By symmetry, (1) is an equivalence, not just an implication.

- What (2) says in a hidden form is that $q_{S^{-1}}$ is the LEGENDRE transform of q_S; see Tapa 13 for a glimpse of this transformation for convex functions of a real variable.

---

74. ★ *The sum of two orthogonal closed vector spaces is closed*

Let $(H, \langle \cdot, \cdot \rangle)$ be a (real) HILBERT space; $\|\cdot\|$ denotes the norm associated with the inner product $\langle \cdot, \cdot \rangle$. Let V_1 and V_2 be two closed vector spaces of H. We assume that V_1 and V_2 are orthogonal. We set $V = V_1 + V_2$. We intend to prove that the vector space V is closed. For that purpose, we prove that V is complete.

Let $(x_n) \subset V_1$ and $(y_n) \subset V_2$ be two sequences such that $(x_n + y_n)$ is a CAUCHY sequence in V.

1°) (a) By just using the definition of a CAUCHY sequence and the basic calculus rule $\|u+v\|^2 = \|u\|^2 + \|v\|^2 + 2\langle u, v \rangle$, check that (x_n) is a CAUCHY sequence in V_1 and (y_n) a CAUCHY sequence in V_2.

(b) Conclude from the above.

2°) Check the following calculus rule on projections:

$$p_V = p_{V_1} + p_{V_2}. \tag{1}$$

Hint. 2°) Recall that if W is a closed vector space of H, we have the following characterization of $\overline{x} = p_W(x), x \in H$:

$$(\overline{x} = p_W(x)) \Leftrightarrow \left(\overline{x} \in W \text{ and } x - \overline{x} \in W^{\perp}\right).$$

Moreover, $(V_1 + V_2)^{\perp} = V_1^{\perp} \cap V_2^{\perp}$. Thus, (1) follows easily.

---

75. ★ *A convex function associated with an arbitrary set*

Let $(H, \langle \cdot, \cdot \rangle)$ be a (real) HILBERT space; $\|\cdot\|$ denotes the norm associated with the inner product $\langle \cdot, \cdot \rangle$. Let $S \subset H$ be an arbitrary nonempty set in H. Consider the function $\varphi_S : H \to \mathbb{R}$ defined as follows:

$$\text{For all } x \in H,\ \varphi_S(x) = \|x\|^2 - d_S^2(x), \tag{1}$$

where d_S denotes the distance function to S (associated, of course, with the Hilbertian norm $\|\cdot\|$).

Prove that φ_S is a continuous convex function on H.

Hint. Convexity of φ_S. Decomposing $\|x-c\|^2$ in the definition of $d_S^2(x) = \inf_{c\in S} \|x-c\|^2$, express φ_S as the pointwise supremum of a family of continuous affine functions:

$$x \mapsto \varphi_S(x) = \sup_{c\in S} \left[2\langle c, x\rangle - \|c\|^2 \right].$$

Comments. - It is surprising that φ_S is convex whatever the nature of the set S is... Among its various consequences, $d_S^2 = \|\cdot\|^2 - \varphi_S$ is always diff-convex, that is to say, a difference of two convex functions (one, $\|\cdot\|^2$, being of class $\mathcal{C}^\infty$).

- Let

$$P_S(x) = \{c \in S : \|x-c\| = d_S(x)\},$$

i.e., the (possibly empty) set of orthogonal projections of x onto S. Then, one easily checks that, whenever $c_x \in P_S(x)$,

$$\frac{1}{2}\varphi_S(y) \geqslant \frac{1}{2}\varphi_S(x) + \langle c_x, y-x\rangle \text{ for all } y \in H;$$

in words, c_x is the slope of a continuous affine minorant of the convex function $\varphi_S/2$, coinciding with it at x. When S is closed convex, $P_S(x)$ just contains one element for all x, and φ_S is differentiable on H. See Tapa 148 for more on this.

- This function φ_S was popularized by the Danish mathematician E. Asplund in 1969. It is one of my favorite convex functions!

---

76. ★ *Summing the squares of integers with alternating signs*

We wish to find a closed form formula, in terms of the positive integers n only, for the following sum:

$$S_n = 1 - 4 + 9 - 16 + ... + (-1)^{n+1} n^2.$$

Hint. Observe the first results with $n = 1, 2, 3, ...$ and guess the general formula. Then prove the result by induction.

Answer. We have

$$S_n = (-1)^{n+1}\frac{n(n+1)}{2}. \tag{1}$$

Comments. Recall that

$$1+2+3+...\ +n = \frac{n(n+1)}{2}, \tag{2}$$

while

$$1+4+9+16+...\ +n^2 = \frac{2n+1}{3}\times\frac{n(n+1)}{2}. \tag{3}$$

77. ★ *When the sum of the squares of three integers equals twice their product*

We look for nonnegative integers m, n, p satisfying

$$m^2+n^2+p^2 = 2\ m\times n\times p. \tag{1}$$

Hints. If one of the integers m, n, p is zero, the other two are zero as well. If m, n, p are positive integers satisfying (1), they are either even or odd. Work out these possibilities and obtain contradictions.

Answer. The only solution to (1) is $m = 0, n = 0, p = 0$.

78. ★ *Angles, double-angles and quadruple-angles in a triangle*

Let $\widehat{A}, \widehat{B}, \widehat{C}$ denote the angles in a triangle. Show that

$$\begin{aligned}
\sin\widehat{A}+\sin\widehat{B}+\sin\widehat{C} &= 4\cos(\widehat{A}/2)\times\cos(\widehat{B}/2)\times\cos(\widehat{C}/2);\\
\sin(2\widehat{A})+\sin(2\widehat{B})+\sin(2\widehat{C}) &= 4\sin\widehat{A}\times\sin\widehat{B}\times\sin\widehat{C};\\
\sin(4\widehat{A})+\sin(4\widehat{B})+\sin(4\widehat{C}) &= -4\sin(2\widehat{A})\times\sin(2\widehat{B})\times\sin(2\widehat{C}).
\end{aligned}$$

Hints. - The key-ingredient is that $\widehat{A}+\widehat{B}+\widehat{C}=\pi$, as also $\left(\pi-2\widehat{A}\right)+\left(\pi-2\widehat{B}\right)+\left(\pi-2\widehat{C}\right)=\pi$.

- Use the usual addition theorems for cosine and sine.

Comment. Beware... If $\widehat{A},\widehat{B},\widehat{C}$ are the angles in a triangle, $2\widehat{A},2\widehat{B},2\widehat{C}$ are not...

---

79. ★ *An unusual minimization problem in a triangle*

Let R be a point in the interior of a triangle ABC. Draw three lines passing through R and parallel to the sides of the triangle. This determines three sub-triangles of ABC (denoted RA_1A_2, RB_1B_2, RC_1C_2) and three parallelograms (see Figure 1).

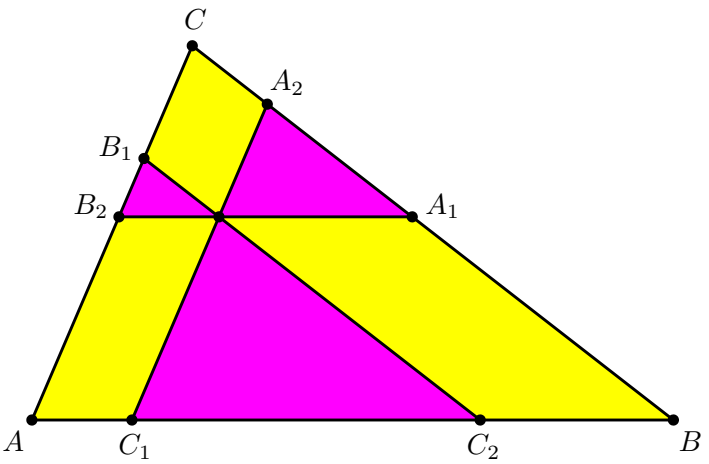

Figure 1

1°) Where should R be placed so that the sum of the areas of these three sub-triangles is minimal?

2°) What is the locus of points R such that the sum of areas of the three sub-triangles is equal to the sum of areas of the three parallelograms?

Hints. 1°) By construction, the three sub-triangles are homothetic to

ABC. The underlying minimization problem reduces to:

$$\text{Minimize } u^2+v^2+w^2$$
$$\text{subject to } u+v+w=C \text{ (a positive constant).}$$

Answers. 1°) R should be placed at the centroid (or isobarycenter) of the triangle. In the optimal configuration, the three sides of the triangle ABC are divided into three equal parts (see Figure 2).

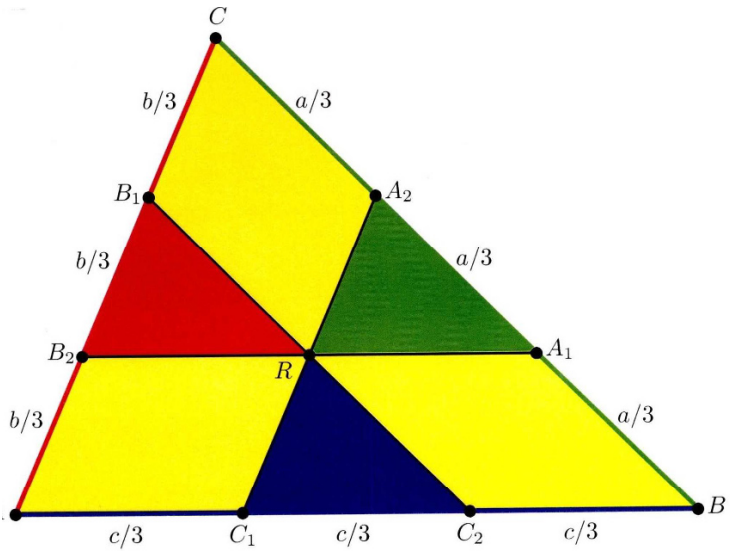

Figure 2

2°) The locus is the so-called STEINER ellipse $\mathcal{E}$ of the triangle ABC: $\mathcal{E}$ is centered at the centroid and tangent to the three sides at their mid-points; it is also the ellipse of maximal area contained in ABC.

Comments. - The STEINER ellipse associated with a triangle is a truly fascinating mathematical object; for example, its two focuses can be localized via calculus with complex numbers, see the comments at the end of Tapa 215 in

J.-B. HIRIART-URRUTY, *Mathematical tapas*, Volume 1 (for Undergraduates). Springer (2016).

Determining the barycentric coordinates of the focuses of the STEINER ellipse is more involved; it is the aim of the following note:

Section entitled "Questions and answers" in *Revue de la Filière Mathématique* (ex-Revue de Mathématiques Spéciales) n°1 (2015–2016), 110–114.

- From time to time, some newspapers propose mathematical challenges. The one in question 1°) was posed in the French newspaper *Le Monde* (Problem 936, fall 2015).

---

80. ★ *Triangles of largest area inscribed in a square*

Let DEF be a triangle contained in a square $ABCD$ (whose sides equal 1).

1°) No constraint on the shape of the triangle.

What is the largest possible area of DEF?

2°) The triangle is constrained to be equilateral.

Find an equilateral triangle of largest area DEF contained in the square $ABCD$? What is the maximal area?
Hint. 2°) Try with triangles placed as follows: one vertex on one of the vertices of the square, and the two other vertices on the opposed sides of the square.

Answers. 1°) The largest area is 1/2.

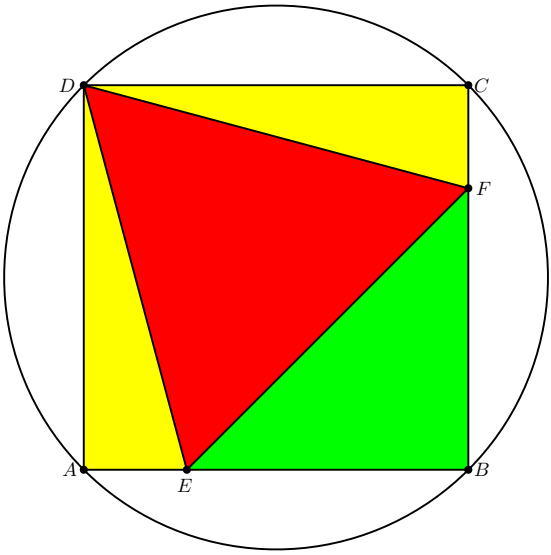

Figure 1

2°) The area of the triangle is maximized when one of its vertices is one of the vertices of the square (D for example in Figure 1), the triangle being symmetrical with respect to the diagonal of the square drawn from this vertex (DB on the figure). One then has: $\widehat{ADE} = \widehat{FDC} = \frac{\pi}{12} = \frac{1}{2}\widehat{EDF}$. The maximal area is $2\sqrt{3} - 3 \backsimeq 0.4641$.

Note that the areas in yellow and green (in Figure 1) are equal.

Comments. - The optimal DEF is sometimes called the Abul-Wafa triangle.

- Question 2°) was posed in the French newspaper *Le Monde* (Problem 961, Spring 2016).

---

81. ★ Viviani's *theorem in 3D*

A theorem (in 2D) due to V. Viviani (1660) asserts the following:

Let P be an arbitrary point inside an equilateral triangle, let $d_1(P)$, $d_2(P)$, $d_3(P)$ denote the distances from P to the three sides of the triangle. Then $d_1(P) + d_2(P) + d_3(P)$ is constant (*i.e,* does not depend on P), and equals the (common) altitude h of the triangle.

We extend this result to regular tetrahedrons (in 3D). Let P be an arbitrary point inside a regular tetrahedron, let $d_1(P), d_2(P), d_3(P), d_4(P)$ denote the distances from P to the four faces of the tetrahedron. Show that $d_1(P) + d_2(P) + d_3(P) + d_4(P)$ does not depend on P, and equals the (common) altitude h of the tetrahedron.

Hint. The tetrahedron can be divided by P into four tetrahedrons, each of them formed by a face of the original tetrahedron and the vertex P. For each of them, the volume V_i is $1/3\ \mathcal{A} \times d_i(P)$, where $\mathcal{A}$ denotes the (common) area of a face of the original tetrahedron. On the other hand, the volume of the whole original tetrahedron is

$$1/3\ \mathcal{A} \times [d_1(P) + d_2(P) + d_3(P) + d_4(P)] = 1/3\ \mathcal{A} \times h,$$

where h is the (common) altitude of the tetrahedron.

---

82. ★ *Minimization of a sum of angles in 3D*

In the usual affine Euclidean space $\mathbb{R}^3$, marked by the orthonormal basis $\left(O; \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k}\right)$, consider a point P in the first (or positive) orthant in $\mathbb{R}^3$, that is to say, a point $P = (x, y, z)$ with nonnegative Cartesian coordinates x, y, z. One then denotes by α, β, γ the three angles between the ray $\overrightarrow{OP}$ and the positive half-axes originating at O and directed by $\overrightarrow{i}$, $\overrightarrow{j}$ and $\overrightarrow{k}$, respectively.

Solve the following two optimization problems:

$$(\mathcal{P}) \quad \left\{ \begin{array}{c} \text{Maximize (resp. Minimize) } \alpha + \beta + \gamma, \\ \text{with } P \text{ in the positive orthant.} \end{array} \right.$$

Hint. If the Cartesian coordinates of P are x, y, z, the objective function $\alpha + \beta + \gamma$ in $(\mathcal{P})$ can be written as $f(x, y, z) = \arccos(x) + \arccos(y) + \arccos(z)$. The constraint set S in $(\mathcal{P})$ can be reduced, without loss of generality, to:

$$\left\{ \begin{array}{l} x \geqslant 0, y \geqslant 0, z \geqslant 0, \\ x^2 + y^2 + z^2 = 1. \end{array} \right. \qquad (\mathcal{S})$$

Answers. For the minimization problem, we have:

$$\inf_S f = 3 \arccos\left(\frac{\sqrt{3}}{3}\right),$$

a value achieved for $x = y = z = \sqrt{3}/3$.

For the maximization problem, we have:

$$\sup_S f = \pi,$$

a value attained on the following three pieces of curves in $(\mathcal{S})$:

$$\left\{\left(0, y, \sqrt{1 - y^2}\right), 0 \leqslant y \leqslant 1\right\};$$
$$\left\{\left(x, 0, \sqrt{1 - x^2}\right), 0 \leqslant x \leqslant 1\right\};$$
$$\left\{\left(x, \sqrt{1 - x^2}, 0\right), 0 \leqslant x \leqslant 1\right\}.$$

83. ★ *Minimization of an energy with a volume constraint*

A particular form of the energy of a particle in a rectangular parallelepiped of sides with lengths $a > 0, b > 0, c > 0$ is expressed as below:

$$E(a,b,c) = C\left(\frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{c^2}\right),$$

where $C > 0$ is defined in terms of PLANCK's constant and the mass of the particle.

Determine a parallelepiped of fixed volume

$$V(a,b,c) = abc = v \quad (v > 0 \text{ is fixed})$$

minimizing the energy $E(a,b,c)$.

Show that this optimization problem has one and only one solution, and determine it.

Answer. The unique solution is $(\overline{a}, \overline{b}, \overline{c}) = (v^{1/3}, v^{1/3}, v^{1/3})$, which means that the optimal parallelepiped is a cube with sides of equal lengths $v^{1/3}$. The minimal value of the energy is then $\frac{3C}{v^{2/3}}$.

---

84. ★ *Integrals of quadratic forms on the unit ball*

Let $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$ be the standard Euclidean space; let $\|\cdot\|$ denote the usual Euclidean norm. For $A \in \mathcal{S}_n(\mathbb{R})$, we consider $x \mapsto \langle Ax, x\rangle$, the associated quadratic form. We want to evaluate simply the integral $I(A) = \int\limits_{\|x\|\leqslant 1} \langle Ax, x\rangle \, \mathrm{d}x$ in terms of invariants associated with A.

1°) Prove that

$$I = \sum_{i=1}^{n} \int\limits_{\|x\|\leqslant 1} \lambda_i x_i^2 \, \mathrm{d}x,$$

where the λ_i' s are the eigenvalues of A.

2°) Deduce that

$$I = c_n \, \mathrm{tr}(A),$$

where c_n is a constant, independent of A, to be determined.

Hint. 1°) Perform a change of variables $x \mapsto y = Qx$, via an orthogonal matrix Q, so that the domain of integration remains unchanged and the quadratic form $x \mapsto \langle Ax, x\rangle$ is "reduced" to a linear combination of y_i^2.

Answer. 2°) We have

$$c_n = \int\limits_{\|x\| \leqslant 1} x_i^2 \, \mathrm{d}x,$$

which depends neither on A nor i.

---

85. ★ *Inverting a perturbed invertible matrix*

Let $A \in \mathcal{M}_n(\mathbb{R})$ be invertible. We perturb it to $A - UV$, where $U \in \mathcal{M}_{n,m}(\mathbb{R})$ and $V \in \mathcal{M}_{m,n}(\mathbb{R})$, and we wish to know how to calculate $(A - UV)^{-1}$ for various U and V, hence for various m. We suppose that $I_m - VA^{-1}U$ is invertible.

1°) Show that $A - UV$ is invertible, with

$$(A - UV)^{-1} = A^{-1} + A^{-1}U(I_m - VA^{-1}U)^{-1}VA^{-1}. \qquad (1)$$

2°) First application. Let u, v be non-null vectors in $\mathbb{R}^n$. The matrix A is perturbed by the rank one matrix uv^T.

What condition would ensure that $A + uv^T$ is invertible? In that case, what is the inverse of $A + uv^T$?

3°) Second application. We suppose, moreover, that A is symmetric. Let u, v be non-null vectors in $\mathbb{R}^n$ and let α and β be non-null real numbers. The matrix A is perturbed to $A + \alpha uu^T + \beta vv^T$.

We set

$$d = \left(1 + \alpha u^T A^{-1} u\right) \times \left(1 + \beta v^T A^{-1} v\right) - \alpha\beta \left(u^T A^{-1} v\right)^2.$$

Show that $\Sigma = A + \alpha uu^T + \beta vv^T$ is invertible whenever $d \neq 0$, with:

$$\begin{aligned} \Sigma^{-1} &= \left(A + \alpha uu^T + \beta vv^T\right)^{-1} \\ &= A^{-1} - \frac{\alpha}{d}\,(1 + \beta v^T A^{-1} v) A^{-1} uu^T A^{-1} \\ &\quad + \frac{\beta}{d}\,(1 + \alpha u^T A^{-1} u) A^{-1} vv^T A^{-1} \\ &\quad - \frac{\alpha\beta}{d}\,(u^T A^{-1} v)(A^{-1} vu^T A^{-1} + A^{-1} uv^T A^{-1}). \end{aligned} \qquad (2)$$

Hint. 3°) Make use of the general result (1) with $U = \begin{bmatrix} \alpha u & \beta v \end{bmatrix} \in \mathcal{M}_{n,2}(\mathbb{R})$ and $V = \begin{bmatrix} -u^T \\ -v^T \end{bmatrix} \in \mathcal{M}_{2,n}(\mathbb{R})$. Then, d is the determinant of

the 2×2 matrix $I_2 - VA^{-1}U$. The calculations leading to the formula (2) are somewhat long but not difficult.

Answers. 2°) $A + uv^T$ is invertible when $1 + v^T A^{-1} u \neq 0$. In that case,

$$\left(A + uv^T\right)^{-1} = A^{-1} - \frac{1}{1 + v^T A^{-1} u}\, A^{-1} u v^T A^{-1}. \qquad (3)$$

Comment. The general formula (1) has been illustrated here with $m = 1$ (first application) and $m = 2$ (second application); it is versatile and could be used in various contexts.

---

86. ★ *Behavior with n of some elements associated with balls in $\mathbb{R}^n$*

Consider the unit balls associated with the basic norms $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_\infty$ in $\mathbb{R}^n$:

$$\begin{aligned} B_1(n) &= \left\{ x = (x_1, ..., x_n) \in \mathbb{R}^n : \sum_{i=1}^{n} |x_i| \leqslant 1 \right\}, \\ B_2(n) &= \left\{ x = (x_1, ..., x_n) \in \mathbb{R}^n : \sum_{i=1}^{n} x_i^2 \leqslant 1 \right\}, \\ B_\infty(n) &= \left\{ x = (x_1, ..., x_n) \in \mathbb{R}^n : \max_{i=1,...,n} |x_i| \leqslant 1 \right\}. \end{aligned}$$

We know that

$$B_1(n) \subset B_2(n) \subset B_\infty(n),$$

but the behaviors with n of some elements associated with these balls may be surprising.

1°) The ball $B_1(n)$.

(a) How many extreme points (or vertices) does $B_1(n)$ have?

(b) What is the diameter of $B_1(n)$? [the diameter of a compact set S is the maximal (usual Euclidean) distance between two points in S].

(c) What is the LEBESGUE measure (= volume) of $B_1(n)$?

2°) The ball $B_\infty(n)$.

Consider the same questions $(a), (b), (c)$ as in 1°).

(d) What is the maximal distance from a point of $B_\infty(n)$ to $B_1(n)$?

3°) The ball $B_2(n)$.

Consider the same questions $(a), (b), (c)$ as in 1°).

(d) What is the maximal distance from a point of $B_2(n)$ to $B_1(n)$?

(e) What is the maximal distance from a point of $B_\infty(n)$ to $B_2(n)$?

Hint. 2°) (d), 3°) (d)–(e). For symmetry reasons, it is helpful to consider points of equal coordinates in the various balls.

Answers. 1°) (a). $B_1(n)$ has exactly $2n$ vertices, those marked as $(0, .., \pm 1, ., 0)$.

(b) The diameter of $B_1(n)$ is constantly equal to 2 [that is, the distance between two opposite vertices].

(c) The volume of $B_1(n)$ is $\frac{2^n}{n!}$.

2°) (a). $B_\infty(n)$ has exactly 2^n vertices, those marked as $(\pm 1, .., \pm 1, ., \pm 1)$.

(b) The diameter of $B_\infty(n)$ is equal to $2\sqrt{n}$ [that is, the distance between two opposite vertices].

(c) The volume of $B_\infty(n)$ is 2^n.

(d) The maximal distance from a point of $B_\infty(n)$ to $B_1(n)$ is $\frac{n-1}{\sqrt{n}}$.

3°) (a). The ball $B_2(n)$ has a "smooth" boundary, the unit sphere for the $\|\cdot\|_2$ norm. All the points in the unit sphere are extreme points of $B_2(n)$.

(b) The diameter of $B_2(n)$ is constantly equal to 2 [that is the distance between two diametrically opposed points].

(c) The volume of $B_2(n)$ is $\pi^{\frac{n}{2}}/\Gamma(\frac{n}{2}+1) = 2\pi^{\frac{n}{2}}/n\Gamma(\frac{n}{2})$.

(d) The maximal distance from a point of $B_2(n)$ to $B_1(n)$ is $\frac{\sqrt{n}-1}{\sqrt{n}}$.

(e) The maximal distance from a point of $B_\infty(n)$ to $B_2(n)$ is $\sqrt{n}-1$.

Comment. For more on the real series whose general term is the volume of $B_2(n)$, see Tapa 327 in

J.-B. Hiriart-Urruty, *Mathematical tapas*, Volume 1 (for Undergraduates). Springer (2016).

---

87. ★ *When the n-dimensional jack gets out of its box*

1°) In 2 dimensions. In the plane, consider a unit square centered at the origin, with equation $\max(|x|, |y|) \leqslant \frac{1}{2}$, hence of area 1. In this square, we place 4 (hockey) pucks centered at the points $\left(\pm\frac{1}{4}, \pm\frac{1}{4}\right)$ and of radius $\frac{1}{4}$. There remains some space for another small puck C_2 centered at the origin and tangent to the already placed 4 pucks.

What is the radius r_2 and the area λ_2 of C_2?

2°) In 3 dimensions. In the "usual" space, consider a unit cube centered at the origin, with equation $\max(|x|, |y|, |z|) \leqslant \frac{1}{2}$, hence of volume 1. In this cube, we place 8 pétanque balls centered at the points $\left(\pm\frac{1}{4}, \pm\frac{1}{4}, \pm\frac{1}{4}\right)$ and of radius $\frac{1}{4}$. There remains some space for another small ball (called a "cochonnet") C_3 centered at the origin and tangent to the 8 pétanque balls.

What are the radius r_3 and the volume λ_3 of C_3?

3°) In n dimensions. In the space $\mathbb{R}^n$, consider the unit cube centered at the origin, with equation $\max(|x_1|, |x_2|, ..., |x_n|) \leqslant 1/2$, hence of volume (*i.e.,* LEBESGUE measure) 1. In this cube, we place 2^n marbles (small balls) centered at the points $\left(\pm\frac{1}{4}, ..., \pm\frac{1}{4}\right)$ and of radius $\frac{1}{4}$. At the center of the box, there is still some space for a small marble, call it a jack (or an n-dimensional cochonnet) C_n centered at the origin and tangent to the 2^n small balls.

(*a*) What is the radius r_n of C_n? Check that for n large enough, a part of the jack gets out of its box!

(*b*) What is the volume λ_n (*i.e.,* LEBESGUE measure) of C_n? Using STIRLING's formula, provide an equivalent of λ_n and show that $\lambda_n \to +\infty$ when $n \to +\infty$.

(*c*) What is the volume μ_n of the space occupied by the 2^n balls? What is the behavior of μ_n as $n \to +\infty$?

Answers. 1°) $r_2 = \frac{\sqrt{2}-1}{4}$, about 10% of the side of the square; $\lambda_2 = \pi\left(\frac{\sqrt{2}-1}{4}\right)^2$, about 3.37% of the area of the square.

2°) $r_3 = \frac{\sqrt{3}-1}{4}$, about 18.3% of the side of the square; $\lambda_3 = \frac{4}{5}\pi\left(\frac{\sqrt{3}-1}{4}\right)^3$, about 2.57% of the volume of the cube.

3°) (*a*). We have $r_n = \frac{\sqrt{n}-1}{4}$, hence $r_n \to +\infty$ when $n \to +\infty$. Thus, for $n = 10$, a part of C_n is already out of the unit box!

(*b*) We have:

$$\lambda_n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}\left(\frac{\sqrt{n}-1}{4}\right)^n.$$

Using STIRLING's formula, we get

$$\lambda_n \sim \frac{\pi^{n/2}}{\sqrt{\pi n}\left(\frac{n}{2e}\right)^{n/2}}\left(\frac{\sqrt{n}-1}{4}\right)^n.$$

Hence, by considering $\ln(\lambda_n)$ for example, one sees that $\lambda_n \to +\infty$ when $n \to +\infty$.

(c) We have:

$$\begin{aligned}\mu_n &= \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}\left(\frac{1}{2}\right)^n,\\ \mu_n &\sim \frac{\pi^{n/2}}{\sqrt{\pi n}\left(\frac{n}{2e}\right)^{n/2}}\left(\frac{1}{2}\right)^n.\end{aligned}$$

By considering $\ln(\mu_n)$, for example, one observes that $\mu_n \to 0$ when $n \to +\infty$.

Thus, the small balls, whose number 2^n increases exponentially, occupy less and less space in the unit box!

Comment. As $n \to +\infty$, strange things, at least counterintuitive things, happen... Here, the distance of the origin to the facets of the unit cube remains constant (it equals $1/2$), while the distance of the origin to the vertices of the box is $\frac{\sqrt{n}}{2}$, which goes to $+\infty$ with n... Besides all these things, the unit ball remains a convex polytope, whose volume constantly equals 1.

Reference for a study in more detail:

J.-L. DUNAU and J.-B. HIRIART-URRUTY, *Quand le cochonnet n-dimensionnel déborde de sa boîte.* Bulletin de l'APMEP, n°444 (2003), 84–86.

---

88. ★ *A maximization problem on the unit-simplex in* $\mathbb{R}^k$

Let Λ_k denote the unit-simplex in $\mathbb{R}^k$, that is

$$\Lambda_k = \left\{(\lambda_1, .., \lambda_k) \in \mathbb{R}^k : \lambda_i \geqslant 0 \text{ for all } i, \text{ and } \lambda_1 + ... + \lambda_k = 1\right\}.$$

Given k positive integers $n_1, ..., n_k$ whose sum is denoted by N, we consider the function $(p_1, ..., p_k) \in \Lambda_k \mapsto f(p_1, ..., p_k) = p_1^{n_1} \times ... \times p_k^{n_k}$. We then want to solve the following optimization problem:

$$(\mathcal{P}) \quad \left\{\begin{array}{c}\text{Maximize } f(p_1, ..., p_k)\\ \text{over all } (p_1, ..., p_k) \in \Lambda_k.\end{array}\right.$$

1°) Check that an optimal solution $(\overline{p_1}, ..., \overline{p_k})$ in $(\mathcal{P})$ satisfies $\overline{p_i} > 0$ for all $i = 1, .., k$.

2°) Determine the unique solution in $(\mathcal{P})$.

Hints. It helps to "break" $f(p_1, ..., p_k)$ by taking its logarithm,

$$g(p_1, ..., p_k) = \ln f(p_1, ..., p_k) = n_1 \ln p_1 + ... + n_k \ln p_k.$$

At an optimal $(\overline{p_1}, ..., \overline{p_k})$, the gradient vector $\nabla g(\overline{p_1}, ..., \overline{p_k})$ should be orthogonal to the hyperplane with equation $p_1 + ... + p_k = 1$.

Answer. The unique solution in $(\mathcal{P})$ is

$$\overline{p_i} = \frac{n_i}{N} \text{ for all } i = 1, .., k. \tag{1}$$

Comment. The source of this Tapa is Statistics: Λ_k represents all the possible probability distributions, $f(p_1, ..., p_k)$ is the so-called *likelihood function*, and the maximization problem $(\mathcal{P})$ leads to a solution, expressed in (1), which is an estimator of the hidden probabilities $p_1, ..., p_k$.

---

89. ★ *Minimizing a differentiable convex function on the unit-simplex of* $\mathbb{R}^k$

Let Λ_k denote the unit-simplex in $\mathbb{R}^k$, that is,

$$\Lambda_k = \left\{(\lambda_1, .., \lambda_k) \in \mathbb{R}^k : \lambda_i \geqslant 0 \text{ for all } i, \text{ and } \lambda_1 + ... + \lambda_k = 1\right\}.$$

Let $f : (x_1, ..., x_k) \in \mathbb{R}^k \mapsto f(x_1, ..., x_k) \in \mathbb{R}$ be a convex differentiable function. We want to characterize the solutions of the following minimization problem:

$$(\mathcal{P}) \quad \left\{ \begin{array}{c} \text{Minimize } f(x_1, ..., x_k) \\ \text{over all } (x_1, ..., x_k) \in \Lambda_k. \end{array} \right.$$

For $x = (x_1, ..., x_k) \in \Lambda_k$, we denote by $I(x)$ the (possibly empty) set of indices i for which $x_i = 0$.

Show that a necessary and sufficient condition for $\overline{x} = (\overline{x}_1, ..., \overline{x}_1) \in \Lambda_k$ to be a solution of $(\mathcal{P})$ above is:

$$\begin{cases} \frac{\partial f}{\partial x_i}(\overline{x}) = \text{a constant } \overline{c} \text{ for all } i \notin I(\overline{x}), \\ \frac{\partial f}{\partial x_i}(\overline{x}) \geqslant \overline{c} \text{ for all } i \in I(\overline{x}). \end{cases}$$

Hint. Use the basic inequality $f(y) \geqslant f(x) + [\nabla f(x)]^T (y - x)$, valid for all x, y in $\mathbb{R}^k$, and the ways for a vector v to be "normal" to Λ_k at $x = (x_1, ..., x_k) \in \Lambda_k$.

Comment. A constraint set like Λ_k often appears in problems arising from Physical Chemistry ($\lambda_1, .., \lambda_k$ in Λ_k represents proportions of products).

---

90. ★ *Minimization of the quotient of two quadratic forms*

Let $\mathbb{R}^n$, $n \geqslant 2$, be equipped with the usual inner product and the associated Euclidean norm. Let a and b be two non-null elements in $\mathbb{R}^n$. Consider the function

$$0 \neq x \in \mathbb{R}^n \mapsto f(x) = \frac{\left(a^T x\right) \times \left(b^T x\right)}{\|x\|^2}. \tag{1}$$

1°) (a) Show that the infimum and the supremum of f on $\mathbb{R}^n \setminus \{0\}$ are attained.

(b) Determine $\mu = \inf_{x \neq 0} f(x)$ and $\nu = \sup_{x \neq 0} f(x)$.

2°) Let $M = ab^T + ba^T \in \mathcal{S}_n(\mathbb{R})$. Use the previous results to determine the smallest eigenvalue $\lambda_{\min}(M)$ and the largest eigenvalue $\lambda_{\max}(M)$ of M.

Hints. 1°) (a). We have: $f(tx) = f(x)$ for all $t \neq 0$. Hence, to minimize or maximize f, it suffices to confine oneself to the unit sphere of $\mathbb{R}^n$.

(b) Consider $\alpha = a/\|a\|, \beta = b/\|b\|$, and $u = \alpha + \beta$, $v = \alpha - \beta$. The Cauchy–Schwarz inequality, applied with u and v, plays an instrumental role here.

2°) Make use of the variational formulations of eigenvalues of M:

$$\lambda_{\min}(M) = \inf_{x \neq 0} \frac{x^T M x}{\|x\|^2} \text{ and } \lambda_{\max}(M) = \sup_{x \neq 0} \frac{x^T M x}{\|x\|^2}.$$

Answers. 1°) (b). We have:

$$\mu = \frac{1}{2}\left(a^T b - \|a\| \times \|b\|\right), \ \ \nu = \frac{1}{2}\left(a^T b + \|a\| \times \|b\|\right).$$

2°) We obtain

$$\lambda_{\min}(M) = a^T b - \|a\| \times \|b\| \, ; \, \lambda_{\max}(M) = a^T b + \|a\| \times \|b\| \, .$$

---

91. ★ *Minimization of a bi-quadratic function*

Let $\mathbb{R}^n$, $n \geqslant 2$, be equipped with the usual inner product and the associated Euclidean norm and let $A \in \mathcal{S}_n(\mathbb{R})$ be a positive semidefinite matrix. Consider the function

$$\begin{array}{rcl} f : \mathbb{R}^n \times \mathbb{R}^n & \to & \mathbb{R}, \\ (x, y) & \mapsto & f(x, y) = \left(x^T A x\right) \times \left(y^T A y\right) - \left(x^T A y\right)^2 . \end{array}$$

We intend to solve the following optimization problem:

$$(\mathcal{P}) \quad \left\{ \begin{array}{c} \text{Maximize } f(x, y) \\ (x, y) \in C, \end{array} \right.$$

where $C = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : \|x\| \leqslant 1 \text{ and } \|y\| \leqslant 1\}$.

1°) Determine the optimal value $\nu = \max_{(x,y)\in C} f(x, y)$ as well as a solution in $(\mathcal{P})$.

2°) What is the image of C under f?

Hints. - The function f is nonnegative, symmetric in x and y, and bi-quadratic (*i.e.,* $f(tx, ty) = t^4 f(x, y)$ for all $t \in \mathbb{R}$ and $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$). Whence we deduce that $\nu = \max_{\|x\|=1, \|y\|=1} f(x, y)$.

- The value ν involves the two largest eigenvalues of A.

Answers. 1°) Let $\lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_n$ be the eigenvalues of A; let $\{e_1, e_2, ..., e_n\}$ be an orthonormal basis of $\mathbb{R}^n$ made up of eigenvectors of A associated with the λ_i' s. Then $\nu = \lambda_1 \lambda_2$, and $(\overline{x}, \overline{y}) = (e_1, e_2)$ is a solution of $(\mathcal{P})$.

2°) The image of the convex compact set C under the continuous function f is the line-segment $[0, \lambda_1 \lambda_2]$.

---

92. ★ *Minimizing a quadratic form over the unit-simplex of* $\mathbb{R}^n$

Let Λ_n denote the unit-simplex of $\mathbb{R}^n$, that is,

$$\Lambda_n = \{(x_1, .., x_n) \in \mathbb{R}^n : x_i \geqslant 0 \text{ for all } i, \text{ and } x_1 + ... + x_n = 1\},$$

and let $f : \mathbb{R}^n \to \mathbb{R}$ be the quadratic form on $\mathbb{R}^n$ defined as

$$f(x_1, .., x_n) = \sum_{i \neq j} x_i x_j.$$

1°) Is f convex on $\mathbb{R}^n$? concave on $\mathbb{R}^n$? Consider the same questions for the restriction of f to Λ_n.

2°) Solve the following optimization problem:

$$(\mathcal{P}) \left\{ \begin{array}{c} \text{Maximize } f(x) \\ \text{over } \Lambda_n. \end{array} \right.$$

Hints. 2°) A first approach: by just playing with the CAUCHY–SCHWARZ inequality.

A second approach: by proving that a solution $x = (x_1, .., x_n)$ to $(\mathcal{P})$ must have $x_i > 0$ for all i, and then writing necessary conditions for optimality via the LAGRANGE theorem.

Answers. 1°) f is neither convex nor concave on $\mathbb{R}^n$. We have $f(x) = x^T A x$, where the matrix $A \in \mathcal{S}_n(\mathbb{R})$ has $\lambda_1 = n - 1$ as a simple eigenvalue and $\lambda_2 = -1$ as an eigenvalue of multiplicity $n - 1$.

By just developing $(x_1 + ... + x_n)^2$, we see that the restriction of f to Λ_n becomes $f(x_1, .., x_n) = 1 - \sum_{i=1}^{n} x_i^2$, which is strictly concave on Λ_n.

2°) The unique solution to $(\mathcal{P})$ is $(\frac{1}{n}, \frac{1}{n}, .., \frac{1}{n})$; the resulting optimal value is $1 - \frac{1}{n}$.

---

93. ★ *Minimizing an entropy-like function on the unit-simplex of* $\mathbb{R}^n$

Let Λ_n denote the unit-simplex of $\mathbb{R}^n$, that is,

$$\Lambda_n = \{(x_1, .., x_n) \in \mathbb{R}^n : x_i \geqslant 0 \text{ for all } i, \text{ and } x_1 + ... + x_n = 1\},$$

and let $f : \mathbb{R}^n \to \mathbb{R}$ be the convex function on $\mathbb{R}^n_+$ defined as

$$(x_1 \geqslant 0, x_2 \geqslant 0, ..., x_n \geqslant 0) \mapsto f(x_1, .., x_n) = \sum_{i=1}^{n} x_i \log(x_i).$$

Here, $0 \times \log(0)$ is deemed to equal 0. In applications where it arises, such f is called an *entropy function*.

Solve the following optimization problem:

$$(\mathcal{P}) \left\{ \begin{array}{c} \text{Minimize } f(x) \\ \text{over } \Lambda_n. \end{array} \right.$$

Hints. The objective function f is a continuous strictly convex one. Check that a solution $x = (x_1, .., x_n)$ to $(\mathcal{P})$ must have $x_i > 0$ for all i, and then write necessary and sufficient conditions for optimality provided by the LAGRANGE theorem.

Answer. The unique solution to $(\mathcal{P})$ is $(\frac{1}{n}, \frac{1}{n}, .., \frac{1}{n})$; the resulting minimal value of the entropy function is $-\log(n)$.

---

94. ★ *The* D'ALEMBERT–GAUSS *theorem via Differential calculus and Optimization*

The fundamental theorem of Algebra (better, on complex numbers), due to D'ALEMBERT and GAUSS (and some others), asserts that every polynomial with complex coefficients and of degree $d \geqslant 1$ has at least one root. We intend to prove this result by using basic tools from Differential calculus and Optimization.

Let $P(z) = a_0 + a_1 z + ... + a_d z^d$, where d is a positive integer and $a_0, a_1, ... , a_d$ are complex numbers (we suppose that $a_d \neq 0$). Let $f(z) = |P(z)|^2$.

1°) (a) Check that f is continuous on $\mathbb{C}$ and

$$f(z) \to +\infty \text{ whenever } |z| \to +\infty.$$

(b) Deduce from the above result that there exists a $z^* \in \mathbb{C}$ minimizing f over $\mathbb{C}$.

We wish to prove that $f(z^*) = 0$.

Let $Q(z) = b_0 + b_1 z + ... + b_d z^d$ be the polynomial of degree d defined as $Q(z) = P(z^* + z)$ [just a shift in the z variables]. The objective now is to prove that 0 is a root of Q, that is to say, $b_0 = 0$.

2°) For $\theta \in [0, 2\pi)$, consider the function

$$\varphi_\theta : t \in \mathbb{R} \mapsto \varphi_\theta(t) = \left|Q(te^{i\theta})\right|^2;$$

this is just the function f evaluated along a line passing through z^* and directed by the unit vector $e^{i\theta}$.

(a) Let s be the first integer $\geqslant 1$ for which $b_1 = ... \; b_{s-1} = 0$ and $b_s \neq 0$. Show that

$$\varphi_\theta^{(k)}(0) = 0 \text{ if } k = 1, ..., s-1; \quad (1)$$

$$\varphi_\theta^{(s)}(0) = 2s! \, \mathrm{Re}(\overline{b}_0 b_s e^{is\theta}). \quad (2)$$

(b) By exploiting the necessary conditions for optimality of higher order applied to φ_θ at 0, show that b_0 necessarily equals 0.

Hints. - There is no loss of generality in assuming that all the coefficients a_i in the polynomial P are real. This simplifies things a little, but not so much.

- 2°) A key-point is that 0 is a local minimizer of φ_θ for any $\theta \in [0, 2\pi)$.

- Note that the $s \geqslant 2$ defined in 2°) (a) does not depend on θ.

- The higher-order necessary condition for optimality (not so often used in Calculus) asserts that if 0 is a local minimizer of φ_θ, then the first k for which $\varphi_\theta^{(k)}(0) \neq 0$ is even and $\varphi_\theta^{(k)}(0) \geqslant 0$.

Answers. 2°) (a). By developing $Q(z) = b_0 + b_s z^s + ... + b_d z^d$ for $z = te^{i\theta}$, we obtain:

$$\varphi_\theta(t) = Q(te^{i\theta}) \times \overline{Q(te^{i\theta})} = |b_0|^2 + 2t^s \, \mathrm{Re}(\overline{b}_0 b_s e^{is\theta}) + \text{terms in } t^k \text{ with } k > s.$$

2°)(b). Proof by contradiction. Assuming that $b_0 \neq 0$, and knowing that $b_s \neq 0$ (by definition), the quantity $\mathrm{Re}(\overline{b}_0 b_s e^{is\theta})$ can be further developed to give rise to an expression like $\sqrt{u^2 + v^2} \cos(s\theta + \rho)$. This quantity, hence $\varphi_\theta^{(s)}(0)$ according to (2), takes values $\geqslant 0$ and < 0 as θ ranges over the interval $[0, 2\pi)$. This leads to a contradiction with the higher-order necessary condition for optimality applied to φ_θ at 0. Hence b_0 should be equal to 0.

Comment. Some references for proofs (of the fundamental theorem of Algebra) of a different type can be found in the note below; there is even a website devoted to collecting all the known proofs.

J.-B. Hiriart-Urruty, *Le théorème fondamental de l'algèbre : une démonstration par le calcul différentiel et l'optimisation.* Bulletin de L'APMEP, n° 466, 695–698 (2006).

---

95. ★ *Orthogonal projection onto a hyperplane, onto a half-space in a* Hilbert *space*

Let $(H, \langle \cdot, \cdot \rangle)$ be a (real) Hilbert space.

1°) For $c \neq 0$ in H and $b \in \mathbb{R}$, we define

- the closed affine hyperplane

$$V = \{x \in H : \langle c, x \rangle = b\}, \tag{1}$$

- the closed half-space

$$V^- = \{x \in H : \langle c, x \rangle \leqslant b\}. \tag{2}$$

1°) (a) Determine the orthogonal projection of $u \in H$ onto V as well as the distance from u to V.

(b) Determine the orthogonal projection of $u \in H$ onto V^- as well as the distance from u to V^-.

2°) First application. Given a positive integer n, let H denote the space of real polynomials of degree at most n, structured as a Euclidean space with the help of the inner product

$$\left(P = \sum_{i=0}^{n} a_i x^i, Q = \sum_{i=0}^{n} b_i x^i \right) \mapsto \langle P, Q \rangle = \sum_{i=0}^{n} a_i b_i.$$

Let

$$V = \{P \in H : P(1) = 0\}.$$

Determine the orthogonal projection of $1 \in H$ onto V as well as the distance from 1 to V.

3°) ★★ Second application. Let H be the Lebesgue space $L^2([0,1], \mathbb{R})$ equipped with the scalar product $\langle f, g \rangle = \int_{[0,1]} f(t) g(t) \, \mathrm{d}\lambda(t)$, and let

$$V = \left\{ f \in H : \int_{[0,1]} f(t) \, \mathrm{d}\lambda(t) = 0 \right\}.$$

Determine the orthogonal projection of $f_k : t \in [0,1] \mapsto f_k(t) = t^k, k \geqslant 0$, onto V as well as the distance from f_k to V.

Hints. 1°) (a). The affine V is parallel to the closed (vectorial) hyperplane

$$V_0 = \{x \in H : \langle c, x\rangle = 0\},$$

and $(V_0)^\perp = \mathbb{R}c$, $H = V_0 \oplus \mathbb{R}c$.

2°) Determine a polynomial Q such that

$$\{P \in H : P(1) = 0\} = \{P \in H : \langle P, Q\rangle = 0\}.$$

3°) If g_1 denotes the function which constantly equals 1 on $[0,1]$, we have the representation (1) of V with $c = g_1$.

Answers. 1°) (a). We have:

$$p_V(u) = u - \frac{\langle c, u\rangle - b}{\|c\|^2}c \text{ and } d_V(u) = \frac{|\langle c, u\rangle - b|}{\|c\|}.$$

(b) We have:

$$p_{V^-}(u) = u - \frac{(\langle c, u\rangle - b)^+}{\|c\|^2}c \text{ and } d_{V^-}(u) = \frac{(\langle c, u\rangle - b)^+}{\|c\|}.$$

2°) We have :

$$p_V(1) = 1 - \frac{1}{n+1}\sum_{i=0}^{n} x^i \text{ and } d_V(u) = \frac{1}{\sqrt{n+1}}.$$

3°) We have:

$$p_V(f_k) : t \in [0,1] \mapsto t^k - \frac{1}{k+1} \text{ and } d_V(f_k) = \frac{1}{k+1}.$$

---

96. ★ *Non-existence of an orthogonal projection onto a closed vector space in a* Banach *space*

Let $E = \{f \in \mathcal{C}([0,1], \mathbb{R}) : f(0) = 0\}$ be structured as a Banach space with the help of the norm $\|\cdot\|_\infty$. We consider

$$V = \left\{f \in E : \int_0^1 f(t)\ \mathrm{d}t = 0\right\}$$

and

$$a : t \mapsto a(t) = t.$$

We intend to show that the infimum of $\|a-f\|_\infty$ for $f \in V$ is not attained, that is to say: there is no orthogonal projection of $a \in E$ onto V.

1°) (a) Check that V is a closed vector space (even a hyperplane) of E.

(b) Let $g \in E$, not identically equal to 0. Show that

$$\left| \int_0^1 g(t)\ \mathrm{d}t \right| < \|g\|_\infty . \tag{1}$$

(c) Let $f \in V$. Check that

$$\|a-f\|_\infty > \frac{1}{2}. \tag{2}$$

2°) Let $\varepsilon \in \left(0, \frac{1}{2}\right)$ and $\alpha = 2\varepsilon^2$. Define $f_\varepsilon : [0,1] \to \mathbb{R}$ as follows:

$$f_\varepsilon(t) \quad = \quad \begin{cases} \left(1 - \frac{1}{2\alpha}\right) t & \text{if } 0 \leqslant t \leqslant \alpha, \\ t - \frac{1}{2} & \text{if } \alpha \leqslant t \leqslant 1-\varepsilon, \\ \frac{1}{2} - \varepsilon & \text{if } 1-\varepsilon \leqslant t \leqslant 1. \end{cases}$$

(a) Check that $f_\varepsilon \in V$.

(b) Determine $g_\varepsilon = a - f_\varepsilon$ and evaluate $\|g_\varepsilon\|_\infty$.

(c) Conclusion?

(d) What is, for example, a missing assumption on $(E, \|\cdot\|)$ which would ensure the existence of an orthogonal projection of (an arbitrary) a onto V, that is: $\overline{f} \in V$ such that

$$\left\|a - \overline{f}\right\| = \inf_{f \in V} \|a-f\| ?$$

Hints. 1°) (a). By definition, V is the kernel of the continuous linear form $u : f \in E \mapsto u(f) = \int_0^1 f(t)\,\mathrm{d}t$.

(b) Proceed by contradiction. Contradicting (1) leads to $\int_0^1 [\|g\|_\infty - g(t)]\ \mathrm{d}t = 0$; whence g is constant on $[0,1]$, *i.e.*, $g(t) = g(0) = 0$ for all $t \in [0,1]$.

(c) Proof by contradiction. We have $\int_0^1 (a-f)(t)\,\mathrm{d}t = \frac{1}{2}$. Using the result of the previous question with $g = a - f$, to have $\|a-f\|_\infty \leqslant \frac{1}{2}$ would lead to $\frac{1}{2} < \frac{1}{2}$.

2°) The functions f_ε and g_ε are piecewise affine; sketching their graphs helps to grasp them.

Answers. 2°) (*b*). We have that $\|g_\varepsilon\|_\infty = \frac{1}{2} + \varepsilon$.

(*c*) We proved that $\inf_{f \in V} \|a - f\|_\infty = \frac{1}{2}$ but this infimum is not attained.

(*d*) The norm $\|\cdot\|_\infty$ on E does not derive from a scalar product; we are not in a HILBERT space situation.

---

97. ★ *An orthogonal decomposition problem in a prehilbertian space*

Let $H = \mathcal{C}^2([0,1], \mathbb{R})$ be structured as a (real) prehilbertian space thanks to the inner product

$$\langle f, g \rangle = \int_0^1 \left[f(t)g(t) + f'(t)g'(t) \right] \, \mathrm{d}t.$$

We denote by $\|\cdot\|$ the norm (on H) derived from $\langle \cdot, \cdot \rangle$.

Let $V = \{f \in H : f(0) = 0 \text{ and } f(1) = 0\}$ and let $W = \{f \in H : f = f''\}$.

1°) (*a*) Give the explicit form of functions in W.

(*b*) Deduce from the above the dimension of W.

2°) (*a*) Prove the following:

$$\begin{gathered} V \text{ and } W \text{ are orthogonal;} \\ V \cap W = \{0\}\,; \\ H = V + W. \end{gathered}$$

(*b*) Application. Let $h \in H$. Determine the orthogonal projection of h onto W.

(*c*) A second application. Let $V_1 = \{f \in H : f(0) = 0 \text{ and } f(1) = 1\}$. Solve the following optimization problem:

$$(\mathcal{P}) \quad \left\{ \begin{array}{c} \text{Minimize } I(f) = \int_0^1 \left[f(t)^2 + f'(t)^2 \right] \, \mathrm{d}t \\ \text{over } f \in V_1. \end{array} \right.$$

Hints. 1°) (*a*). V and W are vector subspaces of H. The elements of W are the solutions of the simple differential equation $f'' - f = 0$.

2°) (*a*). With $f \in V$ and $g \in W$, use integration by parts on

$$\langle f, g \rangle = \int_0^1 \left[f(t)g''(t) + f'(t)g'(t) \right] \, \mathrm{d}t.$$

3°) V_1 is an affine space, a translated version of V.

Answers. 1°) (a). The functions f in W are of the form

$t \in [0,1] \mapsto f(t) = \lambda e^t + \mu e^{-t}$, with λ and μ real numbers,
or, equivalently,
$t \in [0,1] \mapsto f(t) = \alpha \cosh(t) + \beta \sinh(t)$, with α and β real numbers.

(b) W is of dimension 2, hence it is a closed vector space.

2°) (a) - With $f \in V$ and $g \in W$,

$$\begin{aligned} \langle f, g \rangle &= \int_0^1 [f(t)g''(t) + f'(t)g'(t)] \, \mathrm{d}t = [fg']_0^1 \\ &= 0 \text{ (because } f(0) = f(1) = 0). \end{aligned}$$

- If $f(0) = f(1) = 0$ is imposed on $f \in W$, we obtain the null function.

- For $h \in H$, we set

$$t \in [0,1] \mapsto f(t) = h(0)\cosh(t) + \frac{h(1) - h(0)\cosh(1)}{\sinh(1)} \sinh(t); \quad (1)$$

$$t \in [0,1] \mapsto g(t) = h(t) - f(t). \quad (2)$$

Then, $f \in W$ and $g \in V$.

We retain from the results here that $W = V^\perp$.

(b) The orthogonal projection of $h \in H$ onto W is the f function exhibited in (1) above.

(c) V_1 is an affine space whose direction is the vector space V. In the optimization problem $(\mathcal{P})$, the objective function is just $I(f) = \|f\|^2$; so the posed question is to find the projection of the origin on V_1. Due to the results in 2°) (a), one easily sees that the solution is the unique element f_0 in $V_1 \cap W$, namely

$$t \in [0,1] \mapsto f_0(t) = \frac{\sinh(t)}{\sinh(1)}.$$

The optimal value is $I(f_0) = \frac{\cosh(1)}{\sinh(1)}$.

---

98. ★ *Characterizing all the possible solutions of a (scalar) linear differential equation of order n*

Let I be an open interval in $\mathbb{R}$ and let n be a positive integer. In this Tapa, we intend to characterize all the functions $y : I \to \mathbb{R}$ which are n times differentiable on I and such that there exist continuous functions $a_1, ..., a_n : I \to \mathbb{R}$ for which:

$$y^{(n)} + a_1 y^{(n-1)} + ... + \ a_{n-1} y' + a_n y = 0. \qquad (LDE)$$

(In short, y is solution of the linear differential equation (LDE).)

1°) Let $y : I \to \mathbb{R}$ be n times differentiable on I and satisfy (LDE) for a collection $a_1, ..., a_n$ of continuous functions on I. Prove that, if y is not identically equal to 0, then

$$[y(t)]^2 + [y'(t)]^2 + ... + \left[y^{(n-1)}(t)\right]^2 > 0 \text{ for all } t \in I. \qquad (\mathcal{C})$$

2°) Let $y : I \to \mathbb{R}$ be n times differentiable on I and satisfy the condition $(\mathcal{C})$. We set, for all $i = 1, .., n$,

$$t \in I \mapsto a_i(t) = -\frac{y^{(n-i)}(t) \times y^{(n)}(t)}{[y(t)]^2 + [y'(t)]^2 + ... + \left[y^{(n-1)}(t)\right]^2}.$$

Check that y is a solution of (LDE) written with these a_i's.

3°) Conclusion. Characterize the set of all the possible solutions of a linear differential equation like (LDE).

Answers. 1°) If $[y(t_0)]^2 + [y'(t_0)]^2 + ... + \left[y^{(n-1)}(t_0)\right]^2 = 0$ for some $t_0 \in I$, then the solution y of (LDE) is identically equal to 0; this comes directly from the existence and uniqueness result for a CAUCHY problem.

2°) It suffices to check that

$$\sum_{i=1}^{n-1} a_i(t) y^{(n-i)}(t) = -y^{(n)}(t).$$

3°) Except for the null function, a function $y \in \mathcal{C}^n(I, \mathbb{R})$ is a solution of a differential equation like (LDE) if and only if the condition $(\mathcal{C})$ holds true.

---

99. ★ WIRTINGER's *inequality. Application to lower bounds for the periods of an autonomous differential equation.*

Let $\mathbb{R}^d$ be equipped with the usual Euclidean norm $\|\cdot\|$.

1°) Let $f : \mathbb{R} \to \mathbb{R}^d$ be continuously differentiable, T-periodic (with $T > 0$), and satisfy $\int_0^T f(t)\ \mathrm{d}t = 0$.

- Prove the inequality:

$$\left(\frac{1}{T}\int_0^T \|f(t)\|^2\ \mathrm{d}t\right)^{\frac{1}{2}} \leqslant \left(\frac{1}{T}\int_0^T \|f'(t)\|^2\ \mathrm{d}t\right)^{\frac{1}{2}} \times \frac{T}{2\pi}. \qquad (1)$$

- Let $T = 2\pi$. Check that equality holds in (1) if and only if f is a linear combination of the sine and cosine functions.

2°) Application. Let $d \geqslant 2$ and let $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ be continuously differentiable and LIPSCHITZ on $\mathbb{R}^d$ with a constant $L > 0$.

(a) If there exists a solution $x : \mathbb{R} \to \mathbb{R}^d$ to the vectorial autonomous differential equation $x' = \varphi(x)$ which is T-periodic and non-constant, show that necessarily

$$T \geqslant \frac{2\pi}{L}. \tag{2}$$

(b) Check with an example that the inequality (2) above is sharp.

(c) Why has d been chosen to be $\geqslant 2$?

Hints. 1°) As often in such a context: develop f and f' in FOURIER series, use relationships between the FOURIER coefficients of f and f', and apply PARSEVAL's formula.

2°) (a). Apply inequality (1) to x' (this is possible since x' is continuously differentiable and $\int_0^T x'(t)\ \mathrm{d}t = x(T) - x(0) = 0$).

Answers. 2°) (b). Consider the following vectorial (with $d = 2$) linear differential equation:

$$\begin{cases} x' = Ly, \\ y' = -Lx, \end{cases} \quad \text{with } L > 0.$$

Here, $\varphi(x, y) = (Ly, -Lx)$ is LIPSCHITZ with constant L. A non-constant periodic solution of the differential equation above is:

$$t \in \mathbb{R} \mapsto \begin{cases} x(t) = \cos(Lt), \\ y(t) = -\sin(Lt). \end{cases}$$

This solution is periodic, of period $T = \frac{2\pi}{L}$.

(c) The case where $d = 1$ is somewhat particular. Indeed, if $\varphi : \mathbb{R} \to \mathbb{R}$ is a continuous function, any solution $x : I \subset \mathbb{R} \to \mathbb{R}$ of the scalar autonomous differential equation $x' = \varphi(x)$ is necessarily either increasing or decreasing on I. Therefore, there is no non-constant periodic solution x.

Comments. - Inequality (1) is due to W. WIRTINGER. A result in the same vein, for twice continuously differentiable functions, is as follows. Let $f : \mathbb{R} \to \mathbb{R}$ be twice continuously differentiable, 2π-periodic; then:

$$2\int_0^{2\pi} f^2(t)\ \mathrm{d}t + \int_0^{2\pi} \left[f''(t)\right]^2 \mathrm{d}t \geqslant 3\int_0^{2\pi} \left[f'(t)\right]^2 \mathrm{d}t + \frac{1}{\pi}\left[\int_0^{2\pi} f(t)\ \mathrm{d}t\right]^2; \tag{3}$$

equality holds in (3) if and only if f is a linear combination of the constant function and of the sine and cosine functions.

Again, here the technique of proof is: develop f, f' and f'' in FOURIER series, use relationships between FOURIER coefficients of f, f' and f'', and apply PARSEVAL's formula.

- The result (2) is known as J. YORKE's inequality (1963).

---

100. ★ *Minimizing an integral functional*

Let $a < b$, A and B be real numbers, and

$$\begin{aligned} E &= \left\{f \in \mathcal{C}^1(\mathbb{R}) : f(a) = A \text{ and } f(b) = B\right\}, \\ f &\in E \mapsto I(f) = \int_a^b \left[f^2(t) + (f')^2(t)\right]\ \mathrm{d}t. \end{aligned}$$

1°) Check that I is bounded from below on E and that there exists one and only one $\overline{f} \in E$ such that

$$I(\overline{f}) = \inf_{f \in E} I(f).$$

2°) Is I bounded from above on E?

Hints. 1°) To answer the first question, consider the unique solution $\overline{f}$ of the boundary value problem below:

$$\begin{cases} f'' - f = 0, \\ f(a) = A,\ f(b) = B. \end{cases} \tag{1}$$

Then, by using integration by parts, show that $I(f) \geqslant I(\ \overline{f}\ )$ for all $f \in E$. However, beware that the considered f are only $\mathcal{C}^1$, not $\mathcal{C}^2$.

2°) Consider functions whose graphs are made of pieces of parabolas connected by line-segments.

Answers. 1°) Clearly, I is nonnegative on E. Using some calculations on integrals, like integration by parts, one shows that

$$I(f) - I(\overline{f}) = \int_a^b \left\{ [f(t) - \overline{f}(t)]^2 + [f'(t) - \overline{f}'(t)]^2 \right\} \, \mathrm{d}t.$$

Hence, $I(f) > I(\ \overline{f}\ )$ for all $f \neq \overline{f}$ in E.

2°) For r large enough and $a' < b'$ in the interior of $[a, b]$, consider $f_r \in E$ defined as follows:

$$f_r(t) = \begin{cases} r + \frac{A-r}{(a-a')^2}(t-a')^2 & \text{if } t \in [a, a'], \\ r & \text{if } t \in [a', b'], \\ r + \frac{B-r}{(b-b')^2}(t-b')^2 & \text{if } t \in [b', b]. \end{cases}$$

This construction has been made so that

$$I(f_r) \geqslant \int_{a'}^{b'} r^2 \, \mathrm{d}t = r^2(b' - a').$$

Consequently, I is not bounded from above on E.

---

101. ★ *Wild behavior of the higher-order derivatives of a function at a point*

1°) Warm-up 1. Let $f : x \in \mathbb{R} \mapsto f(x) = \arctan(x)$. What are the higher-order derivatives $\arctan^{(2p+1)}(0)$ for positive integers p?

2°) Warm-up 2. Let $f_0 : \mathbb{R} \to \mathbb{R}$ be the function defined as follows:

$$f_0(x) = \begin{cases} e^{-1/x^2} & \text{if } x > 0, \\ 0 & \text{if } x < 0. \end{cases} \tag{1}$$

Prove that f_0 is of class $\mathcal{C}^\infty$ on $\mathbb{R}$ with $f_0^{(n)}(0) = 0$ for all $n \geqslant 1$.

3°) Consider now a sequence $(a_n)_{n\in\mathbb{N}}$ of real numbers. We intend to prove that there exists a $\mathcal{C}^\infty$ function $f : \mathbb{R} \to \mathbb{R}$ such that

$$f^{(n)}(0) = a_n \text{ for all } n. \tag{2}$$

For that, consider a function $\theta : \mathbb{R} \to \mathbb{R}$ of class $\mathcal{C}^\infty$ such that:

$$\theta(x) = \begin{cases} 0 & \text{if } |x| \geqslant 1, \\ 1 & \text{if } |x| \leqslant 1/2. \end{cases}$$

Let

$$\varepsilon_n = \begin{cases} 1 & \text{if } 0 \leqslant |a_n| \leqslant 1, \\ 1/|a_n| & \text{if } |a_n| \geqslant 1; \end{cases} \tag{3}$$

$$u_n(x) = a_n \times \theta\left(\frac{x}{\varepsilon_n}\right)\frac{x^n}{n!}; \tag{4}$$

$$f(x) = \sum_{n=0}^{+\infty} u_n(x). \tag{5}$$

(a) Prove that the function f defined in (5) above answers the posed question: f is of class $\mathcal{C}^\infty$ and $f^{(n)}(0) = a_n$ for all n.

(b) Is the answer to the question unique?

Hints. 1°) Use the development of f around 0 as a sum of a power series.

2°) This is an exercise to be done by every student-reader at least once in his life. Proof by induction on n.

3°) Calculate $u_n^{(k)}(x)$ for all x, obtain an upper bound of $\sup_{x\in\mathbb{R}}\left|u_n^{(k)}(x)\right|$, and prove that the series of functions $(u_n^{(k)})$ is normally convergent on $\mathbb{R}$.

Answers. 1°) We have:

$$\arctan^{(2p+1)}(0) = \pm(2p)!$$

So, even for such a nice (and smooth) function around 0, the higher-order derivatives at 0 may blow up fast with the order of derivation.

3°) - By definitions: $0 < \varepsilon_n \leqslant 1$ for all n; u_n is of class $\mathcal{C}^\infty$, $u_n(x) = 0$ for all $|x| \geqslant \varepsilon_n$ (hence for all $|x| \geqslant 1$) and $u_n(x) = a_n\frac{x^n}{n!}$ for all $|x| \leqslant \varepsilon_n/2$.

- Calculation of $u_n^{(k)}(x)$. We have:

$$\begin{aligned} &\text{if } |x| \geqslant \varepsilon_n,\ u_n^{(k)}(x) = 0 \text{ for all } k \text{ and } n; && (6)\\ &\text{if } |x| \leqslant \varepsilon_n \text{ and } n \geqslant k+1, \\ &\qquad u_n^{(k)}(x) = \sum_{j=0}^{k}\left(\begin{array}{c} k \\ j \end{array}\right)\frac{a_n}{\varepsilon_n^{k-j}}\theta^{(k-j)}\left(\frac{x}{\varepsilon_n}\right)\frac{x^{n-j}}{(n-j)!}. && (7)\end{aligned}$$

Let $M_j = \sup_{x\in[-1,1]}\left|\theta^{(k)}(x)\right|$. By bounding from above $k+1-j$ factors $|x|$ by ε_n and the other ones by 1, and taking into account the

inequality $|a_n| \times \varepsilon_n \leqslant 1$ (*cf.* (3)), one infers from (6)–(7) that

$$\begin{aligned}\left|u_n^{(k)}(x)\right| &\leqslant \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) \frac{|a_n|}{\varepsilon_n^{k-j}} M_{k-j} \frac{\varepsilon_n^{k+1-j} \times 1^{n-k-1}}{(n-j)!} \\ &\leqslant \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) \frac{M_{k-j}}{(n-j)!}.\end{aligned}$$

Therefore, for every fixed k, the series of functions with general term $u_n^{(k)}$ is normally convergent on $\mathbb{R}$. As a consequence, the sum function f is of class $\mathcal{C}^\infty$ on $\mathbb{R}$ with

$$f^{(k)}(x) = \sum_{n=0}^{+\infty} u_n^{(k)}(x) \text{ for all } x \in \mathbb{R}.$$

It remains to check that $u_n^{(k)}(0) = 0$ if $k \neq n$ and $u_k^{(k)}(0) = a_n$.

The solution to the posed problem is not unique, it suffices to add the function f_0 exhibited in Question 2°) to one proposed solution.

Comments. - The mathematical result that any sequence (a_n) can be the higher-order derivatives at 0 of a $\mathcal{C}^\infty$ function is due to G. Peano (1884) and E. Borel (1895).

- Here are some consequences of the proved result:

• Even if the function f is of class $\mathcal{C}^\infty$, the series with general term $\frac{f^{(n)}(0)}{n!}x^n$ may be wildly divergent for $x \neq 0$;

• Even if the series with general term $\frac{f^{(n)}(0)}{n!}x^n$ is nicely convergent for $x \neq 0$, its sum may differ from $f(x)$.

• If f is of class $\mathcal{C}^\infty$ and even on $\mathbb{R}$, there exists a $\mathcal{C}^\infty$ function such that $f(x) = g(x^2)$ for all $x \in \mathbb{R}$.

---

102. ★ *A Hölderian inequality on integrals*

Let f, g, h be three nonnegative functions from $\mathbb{R}^n$ into $\mathbb{R}$. We assume that these three functions are integrable on $\mathbb{R}^n$ and that there exists an $\alpha \in (0,1)$ such that:

$$h(x) \leqslant [f(x)]^{1-\alpha} \times [g(x)]^{\alpha} \text{ for all } x \in \mathbb{R}^n. \tag{1}$$

Prove that:

$$\int_{\mathbb{R}^n} h(x)\, \mathrm{d}x \leqslant \left[\int_{\mathbb{R}^n} f(x)\, \mathrm{d}x\right]^{1-\alpha} \times \left[\int_{\mathbb{R}^n} g(x)\, \mathrm{d}x\right]^{\alpha}. \tag{2}$$

Hint. Let $p = \frac{1}{1-\alpha}$, $q = \frac{1}{\alpha}$. Then p and q are conjugate ($1/p + 1/q = 1$) and $f^{1-\alpha} \in L^p(\mathbb{R}^n)$, $g^\alpha \in L^q(\mathbb{R}^n)$. It remains to apply HÖLDER's inequality to $f^{1-\alpha} \times g^\alpha$.

Comment. There is also an inequality similar to (1) but in the reverse sense (that is why it is sometimes called the anti-Hölderian inequality); it is due to PRÉKOPA and LEINDLER (1972–1973). Here it is. With the same initial assumptions on the functions f, g, h, we change the assumption (1) into:

$$h\left[(1-\alpha)x + \alpha y\right] \geqslant \left[f(x)\right]^{1-\alpha} \times \left[g(y)\right]^{\alpha} \text{ for all } x, y \in \mathbb{R}^n. \tag{3}$$

Then,

$$\int_{\mathbb{R}^n} h(x)\ \mathrm{d}x \geqslant \left[\int_{\mathbb{R}^n} f(x)\ \mathrm{d}x\right]^{1-\alpha} \times \left[\int_{\mathbb{R}^n} g(x)\ \mathrm{d}x\right]^{\alpha}. \tag{4}$$

---

103. ★ *Equality of integrals of functions vs equality of functions*

1°) Let f and g be two continuous functions from $\mathbb{R}^2$ into $\mathbb{R}$ such that: for every box $[a,b] \times [c,d] \subset \mathbb{R}^2$ we have

$$\iint\limits_{[a,b]\times[c,d]} f(x,y)\ \mathrm{d}x\mathrm{d}y = \iint\limits_{[a,b]\times[c,d]} g(x,y)\ \mathrm{d}x\mathrm{d}y. \tag{1}$$

Prove that $f = g$.

2°) SCHWARZ's theorem for twice continuously differentiable functions.

Let $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be a twice continuously differentiable function.

(a) Show that: for every box $[a,b] \times [c,d] \subset \mathbb{R}^2$ we have

$$\iint\limits_{[a,b]\times[c,d]} \frac{\partial^2 \varphi}{\partial y \partial x}(x,y)\ \mathrm{d}x\mathrm{d}y = \iint\limits_{[a,b]\times[c,d]} \frac{\partial^2 \varphi}{\partial x \partial y}(x,y)\ \mathrm{d}x\mathrm{d}y, \tag{2}$$

and calculate this common value (in terms of values of φ at vertices of the box $[a,b] \times [c,d]$).

(b) Deduce from the above that

$$\frac{\partial^2 \varphi}{\partial y \partial x}(x,y) = \frac{\partial^2 \varphi}{\partial x \partial y}(x,y) \text{ for all } (x,y) \in \mathbb{R}^2. \tag{3}$$

Hints. 1°) - Since only integrals of continuous functions θ on boxes $[a,b]\times[c,d]$ are at stake, according to FUBINI's theorem, we have:

$$\begin{aligned}\iint\limits_{[a,b]\times[c,d]} \theta(x,y)\ \mathrm{d}x\mathrm{d}y &= \int_a^b \left[\int_c^d \theta(x,y)\ \mathrm{d}y\right]\ \mathrm{d}x \\ &= \int_c^d \left[\int_a^b \theta(x,y)\ \mathrm{d}x\right]\ \mathrm{d}y. \qquad (4)\end{aligned}$$

- Proof by contradiction. Let $h(x,y)=f(x,y)-g(x,y)$. Suppose that $h(\overline{x},\overline{y})\neq 0$ at some point $(\overline{x},\overline{y})\in\mathbb{R}^2$ and obtain a contradiction.

Answers. 1) Consider the continuous function $h=f-g$. We proceed with a proof by contradiction. Suppose that $h(\overline{x},\overline{y})>0$ at some point $(\overline{x},\overline{y})\in\mathbb{R}^2$. Due to the continuity of h, one can find a box $[a,b]\times[c,d]$ containing $(\overline{x},\overline{y})$ such that

$$h(x,y)>h(\overline{x},\overline{y}) \text{ for all } (x,y) \text{ in } [a,b]\times[c,d]\,.$$

This implies that $\iint\limits_{[a,b]\times[c,d]} h(x,y)\ \mathrm{d}x\mathrm{d}y>0$, which contradicts the initial assumption on the integrals of $h=f-g$ on boxes in $\mathbb{R}^2$ (see (1)).

2) (a). By calculating iterated integrals (using FUBINI's theorem), one finds in both cases

$$\begin{aligned}\iint\limits_{[a,b]\times[c,d]} \frac{\partial^2\varphi}{\partial y\partial x}(x,y)\ \mathrm{d}x\mathrm{d}y &= \iint\limits_{[a,b]\times[c,d]} \frac{\partial^2\varphi}{\partial y\partial x}(x,y)\ \mathrm{d}x\mathrm{d}y \\ &= \varphi(b,d)-\varphi(a,d)+\varphi(a,c)-\varphi(b,c).\end{aligned}$$

Comment. SCHWARZ's theorem (that is to say, equality (3)), proved here for twice continuously differentiable functions, actually holds true for any twice differentiable function (a result somewhat overlooked): whenever $\varphi:\mathbb{R}^2\to\mathbb{R}$ is twice differentiable at $(\overline{x},\overline{y})$ (in the usual (also called FRÉCHET) sense), we have $\frac{\partial^2\varphi}{\partial x\partial y}(\overline{x},\overline{y})=\frac{\partial^2\varphi}{\partial y\partial x}(\overline{x},\overline{y})$.

Part 2. Two-starred tapas

"A good proof is a proof that makes us wiser"

Yu. Manin (1937–)

104. ★★ *Fair division of a pizza into eight parts*

Let us consider a circular pizza that we want to divide into eight parts, four for yourself, four for your friend, both getting the same amount of pizza. One usually does this using a knife, making four straight cuts through the center of the pizza.

Suppose now that the cuts all radiate from a point P in the pizza, not necessarily the center: with four adjacent cuts rotated by an angle of $\pi/4 = 45^\circ$, we get eight different pieces, not necessarily sectors of a circle but similar; for simplicity we nevertheless call them "sectors". You take every second "sector" (for example the white ones in Figure 1) and your friend the remaining ones (light red in Figure 1). So, each of you has four pieces of pizza, of different shapes of course.

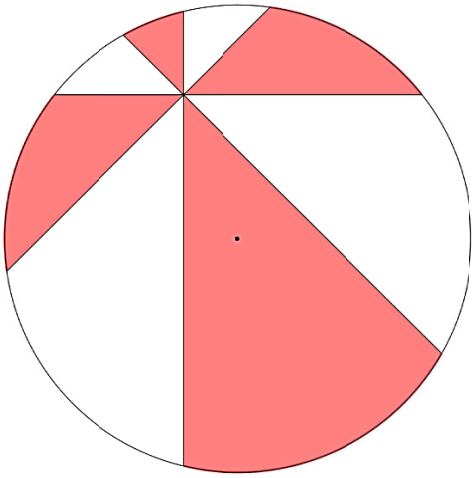

Figure 1

Prove that the pizza has been equally divided, that is to say: the total area of the four pieces of pizza is the same for you and your friend!

Hints. - When considering P as a pole, the outer border of the pizza can be described in polar coordinates as $r = r(\theta)$; for each "sector", θ sweeps across an interval of length $\pi/4$. We do need to know the

expression for $r(\theta)$, which can be fairly complicated. Recall nevertheless that the area of a "sector" is

$$\mathcal{A}_i = \frac{1}{2}\int_{\theta_i}^{\theta_i+\pi/4} r^2(\theta)\ \mathrm{d}\theta. \tag{1}$$

- A key point is the following result on every orthogonal pair of chords in a circle (see Figure 2):

$$(PA)^2 + (PB)^2 + (PC)^2 + (PD)^2 = a^2 + b^2 + c^2 + d^2 = 4r^2, \tag{2}$$

where r is the radius of the circle. This is fairly easy to prove by using properties of lengths in rectangular triangles (like PYTHAGORAS' theorem). Now, imagine orthogonal pairs of chords moving and rotating

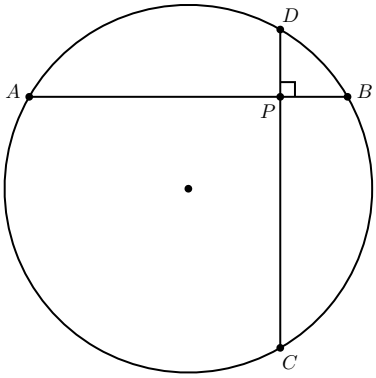

Figure 2

by an angle of $\pi/4$ (like in some wall chronometers that you can see in swimming pools), and use formula (1) for each of the four "sectors".

Answer. The total area of your four pieces of pizza is

$$\begin{aligned}\mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3 + \mathcal{A}_4 \;&=\; \frac{1}{2}\int_0^{\pi/4} r^2(\theta)\ \mathrm{d}\theta + \frac{1}{2}\int_{\pi/2}^{3\pi/4} r^2(\theta)\ \mathrm{d}\theta \\ &\quad + \frac{1}{2}\int_{\pi}^{5\pi/4} r^2(\theta)\ \mathrm{d}\theta + \frac{1}{2}\int_{3\pi/2}^{7\pi/4} r^2(\theta)\ \mathrm{d}\theta.\end{aligned}$$

For your friend, there is a similar formula with just a shift by $\pi/4$ in the endpoints of the intervals of integration.

With the help of simple linear changes of variable on θ, we transform the three integrals $\mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4$ into integrals with the same bounds as in $\mathcal{A}_1$:

$$\mathcal{A}_2 = \frac{1}{2}\int_0^{\pi/4} r^2\left(\theta+\frac{\pi}{2}\right)\,\mathrm{d}\theta, \ldots\,,\ \mathcal{A}_4 = \frac{1}{2}\int_0^{\pi/4} r^2\left(\theta+3\frac{\pi}{2}\right)\,\mathrm{d}\theta.$$

Let us define $\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3 + \mathcal{A}_4$. Then,

$$\mathcal{A} = \frac{1}{2}\int_0^{\pi/4} \left[r^2(\theta) + r^2\left(\theta+\frac{\pi}{2}\right) + r^2\left(\theta+2\frac{\pi}{2}\right) + r^2\left(\theta+3\frac{\pi}{2}\right)\right]\,\mathrm{d}\theta.$$

Now, using the key result on every orthogonal pair of chords in a circle (see (2) in Hints), we get (this is magic!):

$$\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3 + \mathcal{A}_4 = \frac{1}{2}\int_0^{\pi/4} \left[4r^2\right]\,\mathrm{d}\theta = \frac{\pi r^2}{2}. \qquad (3)$$

So, you get exactly half of the pizza... as does your friend.

Comments. - This (surprising) result is true when n, the number of pieces of pizza, is $8, 12, 16, ...$, but false for $n = 2, 4$ (easy to see!), $6, 10, 14, ...$

- For more details on this problem, see:

H. Humenberger, *Dividing a pizza into equal parts – an easy job?* The Mathematics Enthusiast, vol. 12, 389–403 (2015).

---

105. ★★ *Behavior of coefficients of a polynomial with only real roots*

Let $P(x) = \sum_{i=0}^{n} \binom{n}{i} a_i x^i$ be a real polynomial function of degree $n \geqslant 2$, having only real roots. We intend to prove that, for all $k = 1, 2, ..., n-1$,

$$a_k^2 \geqslant a_{k-1} \times a_{k+1}. \qquad (1)$$

1°) Warm-up.

(a) What does (1) say for $n = 2$, *i.e.* for the polynomial $a_0 + 2a_1x + a_2x^2$ having only real roots?

(b) If P is a polynomial function of degree n with only real roots, what about the polynomials $P', P'', ..., P^{(n-1)}$? or the new polynomial $Q(x) = x^n P(\frac{1}{x})$.

2°) - For a given $k = 1, 2, ..., n-1$, first calculate $D(x) = P^{(k-1)}(x)$, then $E(x) = x^{n-k+1} D(\frac{1}{x})$, and finally $E^{(n-k-1)}(x)$.

- Conclude from the obtained result.

Hints. Beware that the coefficient of x^i in P is $\left(\begin{array}{c} n \\ i \end{array}\right) a_i$, not a_i alone.

2°) Begin by illustrating the process in a simple case, with $n = 4$ and $k = 2$, for example. Starting with the polynomial function $P(x) = a_0 + 4a_1x + 6a_2x^2 + 4a_3x^3 + a_4x^4$, we successively obtain:

$$\begin{aligned} D(x) &= P'(x), E(x) = x^3 D\left(\frac{1}{x}\right); \\ E'(x) &= 12a_1x^2 + 24a_2x + 12a_3 = \frac{24}{2}(a_1x^2 + 2a_2x + a_3). \end{aligned}$$

Answers. 1°) (*a*). In this case, (1) says that $(2a_1)^2 - 4a_0a_2 \geqslant 0$, *i.e.* the discriminant of the quadratic polynomial P is nonnegative.

(*b*) By ROLLE's theorem, if the polynomial P of degree n has only real roots, the same holds true for $P', P'', ..., P^{(n-1)}$.

The results of the trick passing from P to Q are the following: if $r \neq 0$ is a root of P, $\frac{1}{r}$ is a root of Q; if 0 is a root of order p of Q, the degree of Q is decreased by p, but no new root is added. As a result, if P has only real roots, so does Q.

2°) $D = P^{(k-1)}$ is a polynomial of degree $n-(k-1) = n-k+1$ having only real roots. The construction of E, followed by that of $E^{(n-k-1)}$, ends up with a polynomial of degree (at most) $n-k+1-(n-k-1) = 2$, which has only real roots. This final polynomial is:

$$\frac{n!}{2}(a_{k-1}x^2 + 2a_kx + a_{k+1}).$$

Expressing that the discriminant of this quadratic polynomial is nonnegative leads directly to the desired inequality (1).

---

106. ★★ *Closedness of the conical hull of a finite set of vectors*

Given m vectors $a_1, ..., a_m$ in $\mathbb{R}^n$, we denote by $\text{cone}(a_1, ..., a_m)$ the (convex) conical hull of $\{a_1, ..., a_m\}$, that is:

$$\text{cone}(a_1, ..., a_m) = \left\{ \sum_{i=1}^{m} \lambda_i a_i : \lambda_i \geqslant 0 \text{ for all } i = 1, ..., m \right\}; \qquad (1)$$

it is the smallest convex cone containing the vectors $a_1, ..., a_m$. It can also be viewed as $\mathbb{R}^+\text{co}(a_1, ..., a_m)$, *i.e.*, the set of all nonnegative multiples of convex combinations of $a_1, ..., a_m$.

Our objective is to prove that $\text{cone}(a_1, ..., a_m)$ is always closed.

1°) Suppose that $0 \notin \text{co}(a_1, ..., a_m)$. Show that $\text{cone}(a_1, ..., a_m)$ is closed.

2°) Suppose that $0 \in \text{co}(a_1, ..., a_m)$. Check that

$$\text{cone}(a_1, ..., a_m) = \bigcup_{j=1}^{n} \text{cone}\,(a_i : i \neq j)\,. \tag{2}$$

3°) Prove, by induction on m, that $\text{cone}(a_1, ..., a_m)$ is closed.

Hints. 2°) The path we follow is similar to that for proving a theorem by CARATHÉODORY stating that any $x \in \text{co}S$, with $S \subset \mathbb{R}^n$, can be represented as a convex combination of $n+1$ elements of S.

3°) The union of a finite number of closed sets is closed.

---

107. ★★ GORDAN's *alternative theorem*

An alternative (or transposition) theorem is a set of two statements such that each one is false when the other is true. More precisely, let P and Q be two logical propositions. They are said to form an *alternative* if one and only one of them is true:

$$(P \Rightarrow \text{not } Q) \text{ and } (\text{not } P \Rightarrow Q)$$

or, just as simply:

$$(P \Leftrightarrow \text{not } Q) \text{ or } (\text{not } P \Leftrightarrow Q)\,.$$

In the whole spectrum of alternative theorems, we choose GORDAN's alternative theorem, maybe the oldest one with propositions involving linear inequalities (1873). Let $a_1, ..., a_m$ be m vectors in $\mathbb{R}^n$; consider the following two statements:

(P) : There exists an $x \in \mathbb{R}^n$ such that $a_i^T x < 0$ for all $i = 1, ..., m$;

(Q) : There are nonnegative λ_i's, not all zero, such that $\displaystyle\sum_{i=1}^{m} \lambda_i a_i = 0$.

GORDAN's alternative theorem expresses that (not $P \Leftrightarrow Q$). As is usually the case in such equivalences, the implication in one direction

is easy (and without much interest), while the converse is more difficult (and is the interesting part). Here, the implication ($Q \Rightarrow$ not P) is easy and does not offer much interest. The converse implication (not $P \Rightarrow Q$) is what requires more effort to prove. We shall do that in an original way by using necessary conditions for approximate optimality (*cf.* Tapa 220 in J.-B. HIRIART-URRUTY, *Mathematical tapas,* Volume 1 (for Undergraduates). Springer (2016)): Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and bounded from below on $\mathbb{R}^n$; then, for any $\varepsilon > 0$, there exists an $x_\varepsilon \in \mathbb{R}^n$ such that

$$\|\nabla f(x_\varepsilon)\| \leqslant \varepsilon.$$

The starting point here is (not P), that is to say:

$$(\text{not } P) : \max_{i=1,\ldots,m} a_i^T x > 0 \text{ for all } x \in \mathbb{R}^n.$$

Let now $f : \mathbb{R}^n \to \mathbb{R}$ be defined as follows:

$$f(x) = \sum_{i=1}^{m} \exp(a_i^T x).$$

By making use of the necessary conditions for approximate optimality recalled earlier, applied to f above, prove that (not $P \Leftrightarrow Q$).

Answer. Assuming (not P), the function f is bounded from below by 1. For any positive integer k, there exists an x_k such that

$$\left\| \nabla f(x_k) = \sum_{i=1}^{m} a_i \exp(a_i^T x_k) \right\| \leqslant \frac{1}{k}. \tag{1}$$

As a consequence, since $f(x_k) \geqslant 1$,

$$\left\| \sum_{i=1}^{m} a_i \frac{\exp(a_i^T x_k)}{f(x_k)} \right\| \leqslant \frac{1}{k}. \tag{2}$$

The coefficients $\lambda_i^{(k)} = \frac{\exp(a_i^T x_k)}{f(x_k)}$ are all positive and add up to 1. So, by extracting a convergent subsequence from $\left(\lambda_1^{(k)}, \lambda_2^{(k)}, ..., \lambda_m^{(k)}\right)_k$, we can pass to the limit in k in (2) and therefore obtain the desired result (Q).

---

108. ★★ *On conditions necessarily satisfied at a minimizer of* $\max(f_1, ..., f_m)$

Let $f_1, ..., f_m : \mathbb{R}^n \to \mathbb{R}$ be m differentiable functions on $\mathbb{R}^n$, and let $f = \max(f_1, ..., f_m)$.

1°) Provide a necessary and sufficient condition on the gradients $\nabla f_i(x_0)$ for f to be differentiable at x_0.

2°) Let $x_0 \in \mathbb{R}^n$ and let $I(x_0) = \{i : f(x_0) = \max_{i=1,...,m} f_i(x_0)\}$ (*i.e.*, the set of indices i where f_i "touches" f at x_0).

Show that the directional derivative $f'(x_0, d)$ of f at x_0 exists in all directions $d \in \mathbb{R}^n$, with:

$$f'(x_0, d) = \max_{i \in I(x_0)} [\nabla f_i(x_0)]^T d. \tag{1}$$

3°) Let x_0 be a local minimizer of f.

(a) Show that:

$$f'(x_0, d) = \max_{I(x_0)} [\nabla f_i(x_0)]^T d \geqslant 0 \text{ for all } d \in \mathbb{R}^n. \tag{2}$$

(b) Deduce that there exist nonnegative λ_i's, $i \in I(x_0)$, summing up to 1, such that

$$\sum_{i \in I(x_0)} \lambda_i \nabla f_i(x_0) = 0. \tag{3}$$

In other words, 0 belongs to the convex hull of the $\nabla f_i(x_0)$' s, $i \in I(x_0)$.

Hints. 1°) Taking the maximum of a finite number of functions preserves some properties like continuity, the (local) LIPSCHITZ property and convexity (for example, $f = \max(f_1, ..., f_m)$ is continuous if all the f_i's are continuous), but destroys differentiability, except in the situation where there is an index i_0 such that $f_{i_0}(x_0) > f_i(x_0)$ for all $i \neq i_0$. When there are several i_0 for which $f_{i_0}(x_0) = \max_{i=1,...,m} f_i(x_0)$, very likely there is a "kink" in the graph of f at x_0.

2°) The directional derivative $f'(x_0, d)$ of f at x_0 in the direction d is defined as

$$f'(x_0, d) = \lim_{t > 0 \to 0} \frac{f(x + td) - f(x)}{t}.$$

3°) (a). As in the differentiable case, use the first-order development of f at x_0 in the d direction:

$$\text{For } t > 0,\ f(x_0 + td) = f(x_0) + tf'(x_0, d) + t\varepsilon_d(t),$$

where $\varepsilon_d(t) \to 0$ when $t > 0 \to 0$.

(b) Make use of GORDAN's alternative theorem (see the previous Tapa 107).

Answers. 1°) Let $I(x_0) = \{i : f(x_0) = \max_{i=1,...,m} f_i(x_0)\}$. We have: f is differentiable at x_0 if and only if all the gradients $\nabla f_i(x_0), i \in I(x_0)$, are equal. In that case, $\nabla f(x_0)$ equals the common vector $\nabla f_i(x_0), i \in I(x_0)$. This is certainly the case when $I(x_0) = \{i_0\}$.

2°) – 3°). For details of proofs, see:

J.-B. HIRIART-URRUTY, *What conditions are satisfied at points minimizing the maximum of a finite number of differentiable functions?* In Nonsmooth optimization: methods and applications, pp. 166–172. Gordon and Breach (1992).

Comments. - Objective functions (to be minimized) of the type $f = \max(f_1, ..., f_m)$ play an important role in Approximation theory (think of $|f| = \max(f, -f)$), Mathematical economics, etc.

- The result developed in this Tapa (a necessary condition for optimality (3)) is a good argument for considering the compact convex polyhedron $\operatorname{co}\{\nabla f_i(x_0), i \in I(x_0)\}$ as a substitute for the (non-existent) gradient of f at x_0. This is indeed the case for the analysis and optimization (theory, algorithms) of such functions.

---

109. ★★ *When is a convex polyhedron described as $Ax \leqslant b$ nonempty?*

Given $A \in \mathcal{M}_{m,n}(\mathbb{R})$ and $b \in \mathbb{R}^m$, we consider the closed convex polyhedron $P(A, b)$ defined as $Ax \leqslant b$, that is, the conjunction of the m linear inequalities

$$a_i^T x \leqslant b_i,\ i = 1, .., m$$

[the vectors a_i are the m lines of the matrix A].

Provide a necessary and sufficient condition, involving A and b, for $P(A, b)$ to be nonempty.

Answer. $P(A, b)$ is nonempty if and only if

$$\left(\sum_{i=1}^{m} \lambda_i a_i = 0 \text{ with } \lambda_i \geqslant 0 \text{ for all } i = 1, ..., m\right) \Rightarrow \left(\sum_{i=1}^{m} \lambda_i b_i \geqslant 0\right). \quad (1)$$

Due to the "homogeneity" in $\lambda = (\lambda_1, ..., \lambda_m)$ of the relation above, a reformulation of the presented result is as follows: $P(A, b)$ is empty

if and only if there exists a vector $\lambda = (\lambda_1, ..., \lambda_m)$ with nonnegative components λ_i such that

$$A^T\lambda = 0 \text{ and } b^T\lambda = -1.$$

Comment. Condition (1) is satisfied, whatever the right-hand side vector b is, if

$$\left(\sum_{i=1}^{m} \lambda_i a_i = 0 \text{ with } \lambda_i \geqslant 0 \text{ for all } i = 1, ..., m\right) \Downarrow \quad (\lambda_i = 0 \text{ for all } i = 1, ..., m)\,. \tag{2}$$

---

110. ★★ *When is a convex polyhedron described as $Ax \leqslant b$ bounded?*

Given $A \in \mathcal{M}_{m,n}(\mathbb{R})$ and $b \in \mathbb{R}^m$, we consider the nonempty closed convex polyhedron $P(A, b)$ defined as $Ax \leqslant b$, that is, the conjunction of the m linear inequalities

$$a_i^T x \leqslant b_i,\ i = 1, .., m.$$

Prove that, for $P(A, b)$ to be bounded, a necessary and sufficient condition is as follows:

$$\left(a_i^T d \leqslant 0 \text{ for all } i = 1, .., m\right) \Leftrightarrow (d = 0)\,. \tag{1}$$

Hint. The directions $d \in \mathbb{R}^n$ satisfying $\left(a_i^T d \leqslant 0 \text{ for all } i = 1, .., m\right)$ are exactly those for which, given $x \in P(A, b)$,

$$x + td \in P(A, b) \text{ for all } t > 0;$$

actually, this condition does not depend on $x \in P(A, b)$.

Comments. - Observe that the necessary and sufficient condition (1) does not depend on the right-hand side vector b; this means that if $P(A, b_0)$ is nonempty and bounded for some $b_0 \in \mathbb{R}^m$, any other polyhedron $P(A, b)$, with $b \in \mathbb{R}^m$, is also bounded (but possibly empty).

- The condition (1) could be written in some equivalent "dual" form, as follows:

$$\operatorname{cone}(a_1, ..., a_m) = \mathbb{R}^n. \tag{2}$$

- It is interesting to visualize the results in Tapas 109 and 110 in the particular cases where $n = 2$ (in the plane) and $m = 3$ (with triangles).

---

111. ★★ *Non-existence of an orthogonal projection onto a closed vector space in a prehilbertian space*

Let $E = \mathcal{C}([-1, 1], \mathbb{R})$ be structured as a prehilbertian space with the help of the inner (or scalar) product $\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\,\mathrm{d}t$. The derived norm $\sqrt{\langle \cdot, \cdot \rangle}$ on E is denoted by $\|\cdot\|$.

Let

$$V = \left\{ u \in E : \int_0^1 u(t)\ \mathrm{d}t = 0 \right\}.$$

1°) Check that V is a closed vector space in E.

2°) Let $f_0 : t \in [-1, 1] \to \mathbb{R}$ be the constant function 1, *i.e.*, $f_0(t) = 1$ for all $t \in [-1, 1]$. Prove that f_0 has no orthogonal projection onto V, that is to say: there is no $\overline{u}$ in V such that

$$\|f_0 - \overline{u}\| = \inf_{u \in V} \|f_0 - u\|.$$

3°) What is, for example, a missing assumption on $(E, \langle \cdot, \cdot \rangle)$ which ensures that there always exists an orthogonal projection of $f \in E$ onto V?

Hints. 1°) Use the characterization of closed sets with the help of sequences, or observe that V is the kernel of a continuous linear form on E. Actually, V is a closed hyperplane in E.

2°) Three points to prove:

- Firstly,

$$d_V(f_0) = \inf_{u \in V} \|f_0 - u\| \geqslant 1.$$

- Secondly, $d_V(f_0) \leqslant 1$ with the help of functions $u_\varepsilon \in V$ defined, for $0 < \varepsilon < 1$, as:

$$u_\varepsilon(t) = \begin{cases} 1 & \text{if} \quad -1 \leqslant t \leqslant -\varepsilon, \\ -\frac{t}{\varepsilon} & \text{if} \quad -\varepsilon \leqslant t \leqslant 0, \\ 0 & \text{if} \quad 0 \leqslant t \leqslant 1. \end{cases}$$

- Finally,

$$\|f_0 - u\| > 1 \text{ for all } u \in V.$$

Answer. 3°) Although the used norm $\|\cdot\|$ derives from a scalar product, $(E, \langle\cdot,\cdot\rangle)$ is not a HILBERT space (that is: $(E, \|\cdot\|)$ is not complete).

---

112. ★★ *Determining the projection on a closed vector space (of codimension 2) in a prehilbertian space*

Let $E = \mathcal{C}([0,1], \mathbb{R})$ be structured as a prehilbertian space with the help of the inner (or scalar) product $\langle f, g\rangle = \int_0^1 f(t)g(t)\,\mathrm{d}t$. The derived norm $\sqrt{\langle\cdot,\cdot\rangle}$ on E is denoted by $\|\cdot\|$.

Let

$$V = \left\{ u \in E : \int_0^1 u(t)\ \mathrm{d}t = 0 \text{ and } \int_0^1 tu(t)\ \mathrm{d}t = 0 \right\}. \tag{1}$$

1°) - Express V in the form

$$V = \{u \in E : \langle f, u\rangle = 0 \text{ and } \langle g, u\rangle = 0\}, \tag{2}$$

with appropriate f and g in E.

- Check that V is a closed vector space of codimension 2.

2°) Let $f_0 : t \in [0,1] \to \mathbb{R}$ be the function defined as $f_0(t) = t^2$ for all $t \in [0,1]$. Determine the orthogonal projection $\overline{u}$ of f_0 on V, as well as

$$\|f_0 - \overline{u}\| = \inf_{u \in V} \|f_0 - u\| = d_V(f_0).$$

Hints. 2°) The convex minimization problem to solve is:

$$(\mathcal{P}) \quad \left\{ \begin{array}{c} \text{Minimize } \theta(u) = \frac{1}{2}\|f_0 - u\|^2 \\ \text{subject to } \langle f, u\rangle = 0 \text{ and } \langle g, u\rangle = 0. \end{array} \right.$$

Use the characterization of solutions provided by LAGRANGE's theorem.

Answers. 1°) We obtain the representation (2) of V with, for example, the following two functions: $f(t) = 1$ and $g(t) = t$ for all $t \in [0,1]$.

V is the kernel of the continuous linear mapping

$$A : u \in E \mapsto A(u) = \begin{pmatrix} \langle f, u\rangle \\ \langle g, u\rangle \end{pmatrix} \in \mathbb{R}^2.$$

It is of codimension 2 since f and g are two linearly independent elements in E.

2°) The system of optimality conditions resulting from LAGRANGE's theorem provides one, and only one, solution in $(\mathcal{P})$; it is

$$\overline{u} : t \in [0,1] \to \overline{u}(t) = t^2 - t + \frac{1}{6}.$$

As a consequence,

$$d_V(f_0) = \frac{\sqrt{7}}{6}.$$

---

113. ★★ *An optimal control problem treated as a projection problem onto a vector space in a prehilbertian space*

The position at time $t \geqslant 0$ of the shaft of an electrical engine fed by a direct current $u(t)$, defined by an angle $\theta(t)$, is governed by the following differential equation:

$$\theta''(t) + \theta'(t) = u(t). \qquad (DE)$$

The initial conditions for (DE), that is to say, the initial position $\theta(t_0)$ as well as the initial velocity $\theta'(t_0)$ at time $t_0 = 0$, are $\theta(t_0) = 0$ and $\theta'(t_0) = 0$.

By acting on the "control function" $u(t)$, we want to bring $\theta(t)$ to the final position $\theta_f = 1$ and final velocity $\theta_f^{'} = 0$ at final time $t_f = 1$, while minimizing the energy used (via $u(t)$). Since the energy used is, in our model, proportional to the integral $\int_0^1 u^2(t)\,\mathrm{d}t$, the posed problem is written in mathematical terms as follows:

$$\begin{aligned}
\text{Minimize } I(u) &= \int_0^1 u^2(t)\mathrm{d}t, \text{ subject to:} && (1)\\
\theta''(t) + \theta'(t) &= u(t) \text{ for all } t \in [0,1], && (2)\\
\theta(0) &= 0,\ \theta'(0) = 0, && (3)\\
\theta(1) &= 1,\ \theta'(1) = 0. && (4)
\end{aligned}$$

Let $E = \mathcal{C}([0,1], \mathbb{R})$ be structured as a prehilbertian space with the help of the inner (or scalar) product $\langle f, g\rangle = \int_0^1 f(t)g(t)\ \mathrm{d}t$. The derived norm $\sqrt{\langle \cdot, \cdot \rangle}$ on E is denoted by $\|\cdot\|$.

1°) For $u \in E$, determine the solution θ_u of the CAUCHY problem defined by (2)–(3).

2°) Express the imposed terminal conditions (4), which are $\theta_u(1) = 1$ and $\theta_u'(1) = 0$, in the form

$$\langle f, u\rangle = 0 \text{ and } \langle g, u\rangle = 1,$$

with appropriate f and g in E.

3°) Solve the posed problem after having formulated it as a projection problem in $(E, \langle\cdot,\cdot\rangle)$: that of 0 on the closed affine subspace in E defined by the equations

$$\langle f, u\rangle = 0 \text{ and } \langle g, u\rangle = 1. \tag{5}$$

Hints. 1°) A change of functions $v_u = \theta'_u$ allows us to transform (2) into a first-order differential equation with constant coefficients. Since u is not given, express $\theta_u(t)$ in an integral form. In doing so, the initial conditions on θ_u can be taken into account.

3°) Since an optimization problem with two equality constraints (see (5)) is involved, use optimality conditions provided by LAGRANGE's theorem.

Answers. 1°) The solution θ_u of the CAUCHY problem defined by (2)–(3) is:

$$\theta_u(t) = \int_0^t \left[e^{-\tau} \int_0^\tau e^s u(s)\, \mathrm{d}s \right] \mathrm{d}\tau.$$

2°) After some calculus on integrals, the imposed terminal conditions (4) can be expressed in the form $\langle f, u\rangle = 0$ and $\langle g, u\rangle = 1$, with:

$$f(t) = e^{t-1},\ g(t) = 1 - e^{t-1}.$$

3°) The posed problem is that of minimizing $I(u) = \|u\|^2$ subject to the equality constraints $\langle f, u\rangle = 0$ and $\langle g, u\rangle = 1$.

With the help of optimality conditions provided by LAGRANGE's theorem, one obtains the solution, the so-called "optimal control"

$$t \in [0,1] \mapsto \overline{u}(t) = \frac{1}{3-e}(1 + e - 2e^t).$$

Comment. In real applications, some technological constraints could be added like $u(t) \in [0, u_{\max}], 0 \leqslant t \leqslant 1$. In that case, some other mathematical tools from Optimal control theory, like the so-called PONTRYAGIN maximum principle (PMP), should be used to determine the optimal control $t \in [0,1] \mapsto \overline{u}(t)$.

114. ★★ *Variations on the projections onto two closed vector subspaces in a* HILBERT *space*

Let $(H, \langle \cdot, \cdot \rangle)$ be a HILBERT space; the topology used on H is the one inherited from the norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.

1°) Let P be a continuous linear mapping from H onto itself. We denote by P^* the adjoint of P. We want to delineate when P is an "orthogonal projection".

Prove the following equivalence:

$$(P \circ P = P \text{ and } P^* = P\ ) \iff (P = P_V),$$

i.e., P is the orthogonal projection mapping onto a closed vector subspace V in H.

2°) Consider two closed vector subspaces M and N in H. With the help of the characterization obtained in the question above, prove:

(1) $(P_M \circ P_N$ is an orthogonal projection$) \Leftrightarrow (P_M$ and P_N commute); in that case, $P_M \circ P_N = P_{M \cap N}$.

(2) $(P_M + P_N$ is an orthogonal projection$) \Leftrightarrow (P_M \circ P_N = 0)$.

(3) If P_M and P_N commute, then $P_M + P_N - P_M \circ P_N$ is an orthogonal projection.

(4) If P_M and P_N commute, then $P_M + P_N - 2P_M \circ P_N$ is an orthogonal projection.

Hints. 1°) - The closed vector space to be considered is clearly

$$V = \{u \in H : P(u) = u\}.$$

- Use the basic definition of P^*, which is:

$$\langle u, P(v) \rangle = \langle P^*(u), v \rangle \text{ for all } u, v \text{ in } H.$$

Comment. One may complete the result in (1) in the following way. The next six assertions are equivalent:

P_M and P_N commute;
P_M and $P_{N^\perp}$ commute;
$P_{M^\perp}$ and P_N commute;
$P_{M^\perp}$ and $P_{N^\perp}$ commute;
M equals $M \cap N + M \cap N^\perp$;
N equals $M \cap N + M^\perp \cap N$.

115. ★★ GATEAUX *vs* FRÉCHET *differentiability of a convex function on a* BANACH *space*

1°) Let $E = \mathcal{C}([0,1])$ denote the vector space of continuous real-valued functions on $[0,1]$, equipped with the norm

$$\|u\|_\infty = \text{maximum of } |u(x)| \text{ on } [0,1].$$

Let $J : E \to \mathbb{R}$ be the convex function on E defined by

$$u \in E \mapsto J(u) = \int_0^1 \sqrt{1 + [u(x)]^2}\, \mathrm{d}x. \tag{1}$$

(*a*) Prove that J is FRÉCHET-differentiable on $(E, \|\cdot\|_\infty)$ and determine $DJ(u)h$ for $u, h \in E$.

(*b*) Prove that J is actually twice continuously differentiable on E and determine $D^2J(u)(h,k)$ for u, h and k in E.

2°) Let now E be the LEBESGUE space $L^1([0,1])$ equipped with the norm

$$\|u\|_1 = \sqrt{\int_{[0,1]} |u(x)|\, \mathrm{d}x}.$$

Let J be defined on E exactly as in (1).

(*a*) Check that J is GATEAUX-differentiable on $(E, \|\cdot\|_1)$ and determine $D_G J(u)h$ for $u, h \in E$.

(*b*) ★★★ Prove that J is not FRÉCHET-differentiable at any $u \in E$.

Hints. 1°) $(E, \|\cdot\|_\infty)$ is a BANACH space and J is indeed a (continuous) convex function on E.

(*a*) Use properties of the following function θ (of the real variable)

$$t \mapsto \theta(t) = \sqrt{1 + t^2}.$$

Because $J(u) = \int_0^1 \theta\,[u(x)]\, \mathrm{d}x$, convexity as well as TAYLOR developments of θ up to the third order allow us to control and bound the expression

$$\begin{aligned} & J(u+h) - J(u) - \int_0^1 \theta'\,[u(x)]\, h(x)\, \mathrm{d}x \\ = \quad & J(u+h) - J(u) - \int_0^1 \frac{u(x)}{\sqrt{1 + [u(x)]^2}} h(x)\, \mathrm{d}x. \end{aligned} \tag{2}$$

(b) Since $\theta''(t) = \frac{1}{(1+t^2)\sqrt{1+t^2}}$, a natural candidate for the continuous bilinear form $D^2J(u)$ (the second FRÉCHET differential of J at u) is

$$\begin{aligned} D^2J(u) \quad &: \quad (h,k) \in E \times E \mapsto \int_0^1 \theta'' \left[u(x)\right] h(x)k(x)\, \mathrm{d}x \\ &= \int_0^1 \frac{1}{\left(1+\left[u(x)\right]^2\right)\sqrt{1+\left[u(x)\right]^2}} h(x)k(x)\, \mathrm{d}x. \end{aligned} \tag{3}$$

2°) (a). GATEAUX-differentiability is easier to access than FRÉCHET-differentiability. We have to prove that $\frac{J(u+th)-J(u)}{t}$ has a limit when $t \to 0$ and that the resulting expression is a continuous function of h (the result is then denoted $D_GJ(u)h$).

Again use the properties of the θ function, like the next one:

$$|\theta(t+\varepsilon) - \theta(t)| \leqslant |\varepsilon| \text{ for all } t, \varepsilon.$$

Since the underlying space is $L^1([0,1])$, results like LEBESGUE's theorem on a sequence of functions bounded from above by an integrable function have to be called.

(b) If J were FRÉCHET-differentiable at u, the FRÉCHET differential $DJ(u)$ would equal the GATEAUX differential $D_GJ(u)$ given above.

The idea of the proof is to exhibit a sequence (h_n) of functions converging to 0 in $L^1([0,1])$ and such that

$$\frac{|J(u+h_n) - J(u) - D_GJ(u)h_n|}{\|h_n\|_1} \text{ does not tend to 0.}$$

Answers. 1°) (a). Since $\theta'(t) = \frac{t}{\sqrt{1+t^2}}$, we have

$$\left|\theta(t+\varepsilon) - \theta(t) - \theta'(t)\varepsilon\right| \leqslant \frac{\varepsilon^2}{2} \text{ for all } t, \varepsilon.$$

Consequently,

$$\left| J(u+h) - J(u) - \int_0^1 \frac{u(x)}{\sqrt{1+\left[u(x)\right]^2}} h(x)\, \mathrm{d}x \right| \leqslant \frac{\|h\|^2}{2}.$$

This inequality proves that J is FRÉCHET-differentiable at $u \in E$ with:

$$h \in E \mapsto DJ(u)h = \int_0^1 \frac{u(x)}{\sqrt{1+\left[u(x)\right]^2}} h(x)\, \mathrm{d}x. \tag{4}$$

(b) We have that

$$\left|\theta'(t+\varepsilon)-\theta'(t)-\theta''(t)\varepsilon\right| \leqslant \frac{3}{2}\varepsilon^2 \text{ for all } t,\varepsilon.$$

Therefore,

$$\left|DJ(u+h)k - DJ(u)k - \int_0^1 \frac{1}{\left(1+[u(x)]^2\right)\sqrt{1+[u(x)]^2}} h(x)k(x)\ \mathrm{d}x\right| \leqslant \frac{3}{2}\|h\|^2\|k\| \text{ for all } u,h,k \text{ in } E.$$

As a result, DJ is FRÉCHET-differentiable at $u \in E$, and since $D^2J(u)(h,k)$ is the differential at u of the differentiable function $v \mapsto DJ(v)h$ applied at k (a technical trick to calculate $D^2J(u)(h,k)$, somewhat overlooked), we get that

$$D^2J(u)(h,k) = \int_0^1 \frac{1}{\left(1+[u(x)]^2\right)\sqrt{1+[u(x)]^2}} h(x)k(x)\ \mathrm{d}x.$$

From this expression of $D^2J(u)$ and the expression of θ''', one obtains that D^2J is LIPSCHITZ continuous on E with LIPSCHITZ constant 3 (hence continuous on E).

2°) (a). As expected, J is GATEAUX-differentiable on E with the following expression of the GATEAUX differential $D_GJ(u)$ at $u \in E$:

$$h \in E \mapsto D_GJ(u)h = \int_{[0,1]} \frac{u(x)}{\sqrt{1+[u(x)]^2}} h(x)\ \mathrm{d}x \tag{5}$$

(the same as in (4), even if the underlying space E is larger and the norm different).

Comments. - It is a bit disconcerting that even for (nice) continuous convex functions on a BANACH space E, there might be a gap between GATEAUX and FRÉCHET differentiability. This phenomenon does not occur when E is finite-dimensional.

- For long time, actually since my student years, I had written GÂTEAUX with a circumflex accent on the A... I only recently learnt that GATEAUX's birth certificate, as well as his own signature, did not contain any such accent; so...

116. ★★ *Discrepancy between two norms in second-order optimality conditions*

Let E be the LEBESGUE space $L^\infty([0,1])$ equipped with the two norms:
$\|u\|_2 = \sqrt{\int_{[0,1]} [u(x)]^2 \ \mathrm{d}x}$;
$\|u\|_\infty$ = essential supremum of $|u(x)|$ on $[0,1]$.
Let $J : E \to \mathbb{R}$ be defined by

$$u \in E \mapsto J(u) = \int_{[0,1]} \sin[u(x)] \ \mathrm{d}x. \tag{1}$$

1°) Show that J is (FRÉCHET-) differentiable on $(E, \|\cdot\|_2)$ as well as on $(E, \|\cdot\|_\infty)$, and that the differential $DJ(u)$ of J at $u \in E$ is given (in both cases) by:

$$h \in E \mapsto DJ(u)(h) = \int_{[0,1]} h(x) \cos[u(x)] \ \mathrm{d}x. \tag{2}$$

2°) Consider the following unconstrained minimization problem:

$$(\mathcal{P}) \ \left\{ \begin{array}{c} \text{Minimize } J(u) \\ u \in E. \end{array} \right.$$

(a) Check that the constant function $\overline{u}(x) = -\frac{\pi}{2}$ is a global minimizer of J on E, but that it is not a strict local minimizer of J on $(E, \|\cdot\|_2)$.

(b) Show that J is twice (FRÉCHET-) differentiable on $(E, \|\cdot\|_\infty)$, and that there exists an $\alpha > 0$ such that:

$$D^2J(\overline{u})(h,h) \geqslant \alpha \left(\|h\|_2\right)^2 \text{ for all } h \in E. \tag{3}$$

(c) Show, with the help of the result in (a) and the second-order sufficiency conditions for strict local optimality, that J cannot be twice (FRÉCHET-) differentiable at $\overline{u}$ when E is equipped with the norm $\|\cdot\|_2$.

Hints. 1°) - Using the properties of the sine and cosine functions, prove that

$$\left| J(u+h) - J(u) - \int_{[0,1]} h(x) \cos[u(x)] \ \mathrm{d}x \right| \leqslant \left(\|h\|_2\right)^2 / 2. \tag{4}$$

- We have $\|\cdot\|_2 \leqslant \|\cdot\|_\infty$.

2°) (c). The second-order sufficient conditions for strict local optimality say the following: Let E be equipped with the norm $\|\cdot\|_2$; if J is

twice (FRÉCHET-) differentiable at $\overline{u}$ and if the following conditions are satisfied:

$$(5)\ \begin{cases} DJ(\overline{u}) = 0; \\ \text{there exists an } \alpha > 0 \text{ with } D^2J(\overline{u})(h,h) \geqslant \alpha \left(\|h\|_2\right)^2 \text{ for all } h \in E, \end{cases}$$

then $\overline{u}$ is a strict local minimizer of J.

Answers. 2°) (a). We have:

$$J(u) \geqslant J(\overline{u}) = -1 \text{ for all } u \in E.$$

For $0 < \varepsilon < 1$, let $\overline{u_\varepsilon} \in E$ be defined as:

$$\overline{u_\varepsilon}(x) = \begin{cases} 3\pi/2 & \text{if } 0 \leqslant x \leqslant \varepsilon; \\ -\pi/2 & \text{if } \varepsilon < x \leqslant 1. \end{cases}$$

We have:

$$J(\overline{u_\varepsilon}) = J(\overline{u}) \text{ and } \|\overline{u_\varepsilon} - \overline{u}\|_2 \leqslant \sqrt{2\pi} \times \varepsilon^{1/4},$$

which shows that $\overline{u}$ is not a strict local minimizer on $(E, \|\cdot\|_2)$.

(b) J is twice (FRÉCHET-) differentiable on $(E, \|\cdot\|_\infty)$ with:

$$(h,k) \in E \times E \mapsto D^2J(u)(h,k) = -\int_{[0,1]} h(x)k(x)\sin\left[u(x)\right]\ \mathrm{d}x. \quad (6)$$

In particular, $D^2J(\overline{u})(h,h) = \left(\|h\|_2\right)^2$.

(c) Here, E is equipped with the norm $\|\cdot\|_2$. If J were twice (FRÉCHET) differentiable at $\overline{u}$, its second differential $D^2J(u)(h,k)$ would be the same as the one expressed in (6) (this is easy to check), and the second-order sufficient conditions for local optimality (recalled in (5)) would ensure that $\overline{u}$ is a strict local minimizer on $(E, \|\cdot\|_2)$, which is not the case.

Comment. Enlarging (or reducing) the underlying space L and changing the norm $\|\cdot\|$ equipping it may play tricks when differential calculus on $(L, \|\cdot\|)$ is concerned... One easily falls into a trap.

117. ★★ *The* HAUSDORFF *distance between bounded closed sets in a metric space*

Let (E,d) be a metric space and let $\mathcal{F}$ be the family of all nonempty bounded closed subsets of E. For A and B in $\mathcal{F}$, we set:

$$\begin{aligned} e_H(A,B) &= \sup_{x\in A} d_B(x); \\ d_H(A,B) &= \max\left(e_H(A,B), e_H(B,A)\right). \end{aligned}$$

1°) With the help of simple examples in the plane (equipped with the usual Euclidean distance), illustrate geometrically what $e_H(A,B) \leqslant \varepsilon$ and $d_H(A,B) \leqslant \varepsilon$ mean.

2°) Prove that d_H is a distance on $\mathcal{F}$ (called the HAUSDORFF distance).

3°) Question for thinking over. Is the HAUSDORFF distance an appropriate notion for expressing that a set is close to another one?

Hints. 2°) - The notations and calculations are a bit lighter if one considers a normed vector space $(E,\|\cdot\|)$ instead of a metric space (E,d).

- The main tools are the basic properties of the distance function d and the definition of $\nu = \sup S$ when $S \subset \mathbb{R}$ is bounded from above.

Answer. 3°) It depends on the collection of subsets that one is considering... If A and B are two bounded closed convex subsets of a normed vector space, to have $d_H(A,B) \leqslant \varepsilon$ indeed means (geometrically at least) that A and B are close to each other. However, if B is obtained from A simply by adding some "whiskers", A and B can be considered close to each other while $d_H(A,B)$ may be large.

---

118. ★★ *A shrinking family of closed subsets in a complete metric space*

Let $(F_n)_{n\geqslant 1}$ be a family of nonempty closed subsets in a complete metric space (E,d) satisfying:

$$\begin{gathered} (i)\ F_{n+1} \subset F_n \text{ for all } n; \\ (ii) \text{ the diameter of } F_n \text{ tends to 0 as } n \to +\infty. \end{gathered}$$

Prove that $\cap_{n\geqslant 1} F_n$ reduces to one and only one point.

Hints. Recall that the diameter $\delta(A)$ of A is $\sup_{x,y\in A} d(x,y)$.

- Let F denote $\cap_{n\geqslant 1} F_n$. The assumption (ii) in the statement of the Tapa implies that $\delta(F) = 0$. One then has to prove that F is nonempty (that is the key point). For this, choose $u_n \in F_n$ for all n; then (u_n) is a CAUCHY sequence, hence it converges towards $u \in E$. Check then that $u \in \overline{F_n} = F_n$ for all n; thus $u \in F$.

Comment. The result put into perspective in this Tapa turns out to be helpful as a mathematical tool in Analysis.

---

119. ★★ *The problem of farthest points in a compact subset of a* BANACH *space*

Let K be a nonempty compact subset of a Banach space $(E, \|\cdot\|)$. For all $x \in E$, we set:

$$Q_K(x) = \left\{ \overline{y} \in K : \|x - \overline{y}\| = \sup_{y \in K} \|x - y\| \right\}.$$

($Q_K(x)$ is the set of points in K which are farthest from x).

Prove that if $Q_K(x)$ is a singleton (*i.e.,* reduces to one element) for all $x \in E$, then K itself is a singleton.

Hints. - The posed challenge is a bit simpler to solve in a finite-dimensional context (*i.e.,* the mathematical tools to be used may be simpler).

- With the assumption made, $Q_K(x) = \{q_K(x)\}$ for all $x \in E$, so that q_K is actually a mapping from E onto K. The first step consists in proving that q_K is actually continuous.

- The second step consists in resorting to some fixed point theorem, like that of SCHAUDER and TIKHONOV: if f is continuous from a compact convex set C into itself, there then exists an $\overline{x} \in C$ such that $f(\overline{x}) = \overline{x}$.

Answer. Let C be the closed convex hull of K; thus C is a compact convex set in E. We reason by contradiction. Suppose that K is not a singleton. Then, $q_K(x) \neq x$ for all $x \in K$. This contradicts the fixed point theorem (of SCHAUDER and TIKHONOV, for example) applied to $q_K : C \to C$ (q_K does have a fixed point in C).

Comments. - The assumption "$Q_K(x)$ is a singleton (*i.e.,* reduces to one element) for all $x \in K$" is not strong enough to secure that K is a singleton; as a counterexample, consider a doubleton $K = \{a, b\}, a \neq b$ in E.

- The so-called "farthest point conjecture" can be stated as follows: Let S be a nonempty closed bounded subset in a normed vector space E; if $Q_S(x)$ is a singleton for all $x \in E$, then S itself is a singleton. This conjecture, formulated by V. KLEE in the 1960s, remains unsolved in its full generality.

---

120. ★★ *Approximating a continuous function on* $[0,1]$ *by "quadratic inf-convolution"*

Let $f : [0,1] \to \mathbb{R}$ be a continuous function. Associated with f, one defines a new function $T(f) : [0,1] \to \mathbb{R}$ as follows:

$$\text{For all } x \in [0,1],\ T(f)(x) = \inf\left[f(y) + (x-y)^2\right]. \tag{1}$$

1°) Show that $T(f)$ is LIPSCHITZ on $[0,1]$.

2°) (a) Solve the fixed point equation $T(f) = f$.

(b) Check that T is a continuous mapping from $(\mathcal{C}([0,1],\mathbb{R}), \|\cdot\|_\infty)$ into itself.

3°) Let (f_n) be the sequence of functions in $\mathcal{C}([0,1],\mathbb{R})$ defined as follows:

$$\begin{cases} f_0 = f, \\ f_{n+1} = T(f_n). \end{cases}$$

Prove that (f_n) converges uniformly on $[0,1]$ towards a specific function in $\mathcal{C}([0,1],\mathbb{R})$ to be determined.

Hints. 1°) To see how the transformation T works, it may be wise to calculate $T(f)$ for some particular functions f, for example f constant or $f(x) = \alpha x^2$.

3°) The sequence (f_n) is decreasing. DINI's theorem asserts that if we have a monotone sequence of continuous functions on a compact space K, and if the limit function is also continuous, then the convergence is uniform on K.

Answers. 1°) $T(f)$ is a LIPSCHITZ function on $[0,1]$, with LIPSCHITZ constant 2.

2°) We have: $T(f) = f$ if and only if f is a constant function.

3°) The sequence (f_n) converges uniformly on $[0,1]$ towards the constant function $\mu = \min_{x\in[0,1]} f(x)$.

Comment. The name "quadratic inf-convolution" comes from three terms: firstly, *inf* and *quadratic* because it is f perturbed by a quadratic term which is minimized in the definition (1) of $T(f)(x)$; secondly, *convolution* is suggested by the resemblance of the formula (1) with the classical convolution formula (with an integral instead of an infimum).

---

121. ★★ *Successively averaging a continuous function on* $[0,1]$

Let $f : [0,1] \to \mathbb{R}$ be a continuous function. We define a new function $T(f) : [0,1] \to \mathbb{R}$ as follows:

$$\begin{cases} \text{For all } x \in (0,1],\ T(f)(x) = \dfrac{1}{x}\displaystyle\int_0^x f(t)\ \mathrm{d}t, \\ T(f)(0) = f(0). \end{cases} \tag{1}$$

1°) Check that $T(f)$ is indeed continuous on $[0,1]$.

2°) (a) Give examples of functions f for which $T(f) = f$.

(b) Check that T is a linear continuous mapping from $(\mathcal{C}([0,1],\mathbb{R}), \|\cdot\|_\infty)$ into itself.

3°) Let (f_n) be the sequence of functions in $\mathcal{C}([0,1],\mathbb{R})$ defined as follows:

$$\begin{cases} f_0 = f, \\ f_{n+1} = T(f_n). \end{cases}$$

Prove that (f_n) converges uniformly on $[0,1]$ towards a specific function in $\mathcal{C}([0,1],\mathbb{R})$, to be determined.

Hints. 1°) To see how the transformation T works, it may be wise to calculate $T(f)$ for some particular f, for example for polynomial functions f.

3°) If $f(0) = 0$, one also has $T(f)(0) = 0$. If u denotes the function which constantly equals 1, one easily checks by induction that f_n can be decomposed as:

$$f_n = f(0)u + g_n,$$

where $g_n \in \mathcal{C}([0,1],\mathbb{R})$ satisfies $g_n(0) = 0$.

Suggestion: Show that the sequence (g_n) converges uniformly on $[0,1]$ towards the null function.

Answers. 1°) The limit of $T(f)(x)$ as $x > 0$ tends to 0 is the derivative at 0 of the function $x \mapsto \int_0^x f(t)\ \mathrm{d}t$, that is, $f(0)$.

2°) (a). Constant functions f satisfy $T(f) = f$.

(b) The mapping T is clearly linear and:

$$\text{For all } f \in \mathcal{C}\left([0,1], \mathbb{R}\right), \|T(f)\|_\infty \leqslant \|f\|_\infty .$$

It is therefore continuous.

3°) The sequence (f_n) converges uniformly on $[0,1]$ towards the constant function $x \mapsto f(0)$.

Comment. The proofs are a bit easier to carry out when the functions f are polynomials. Then, one can use the density of the polynomial functions in $\mathcal{C}\left([0,1], \mathbb{R}\right)$ (WEIERSTRASS' theorem), combined with the continuity of T.

---

122. ★★ *Comparing the lengths of boundaries of convex sets in the plane*

Let C_1 and C_2 be two compact convex sets in the plane whose boundaries are denoted by Γ_1 and Γ_2, respectively. We suppose that $C_1 \subset C_2$.

Prove that:

$$\text{length of } \Gamma_1 \leqslant \text{length of } \Gamma_2. \tag{1}$$

Hint. Given a compact convex set C in the plane, the so-called parameterized support function of C is

$$\theta \in [0, 2\pi] \mapsto h_C(\theta) = \max_{(x,y)\in C} \left[x\cos\theta + y\sin\theta\right]; \tag{2}$$

it is the (so-called) *support function* of C in the unitary direction $\overrightarrow{u_\theta} = (\cos\theta, \sin\theta)$. For example: if C is the ball centered at 0 and of radius $1/2$, the function h_C constantly equals $1/2$; if C is the square with vertices $(\pm 1, \pm 1)$, $h_C(\theta) = |\cos\theta| + |\sin\theta|$. A beautiful result from Integral convex geometry tells us that the length of the boundary Γ of C is given by the following simple formula:

$$\ell(\Gamma) = \int_0^{2\pi} h_C(\theta)\ \mathrm{d}\theta. \tag{3}$$

Answer. Since $C_1 \subset C_2$, we have that $h_{C_1} \leqslant h_{C_2}$ (see the definition (2)). It then follows from (3) that

$$\ell(\Gamma_1) \leqslant \ell(\Gamma_2).$$

Comments. - As a further application of the result above, consider the following situation: a convex function $f : I \to \mathbb{R}$, $a < b$ in the interval I, the two tangent lines to the graph of f at points $A = (a, f(a))$ and $B = (b, f(b))$ which meet at the point M (two supporting lines can be defined even if f is not differentiable at a or b); then the length of the graph of f between A and B is less than the sum of two lengths $AM + MB$. This is a sort of "curvilinear triangle inequality".

- For more on (2) and (3), see:

T. Bayen and J.-B. Hiriart-Urruty, *Objets convexes de largeur constante (en 2D) ou d'épaisseur constante (en 3D) : du neuf avec du vieux.* Annales des Sciences Mathématiques du Québec, Vol. 36, n° 1, 17–42 (2012).

---

123. ★★ *Convex sets of constant width in the plane*

Let C be a compact convex set in the plane, with nonempty interior; we denote by Γ its boundary. The support function σ_C of C is defined as follows:

$$d \in \mathbb{R}^2 \mapsto \sigma_C(d) = \max_{c \in C} c^T d. \tag{1}$$

The width of C in the unitary direction $\overrightarrow{u} \in \mathbb{R}^2$ is the quantity

$$e_C(\overrightarrow{u}) = \max_{c \in C} c^T \overrightarrow{u} - \min_{c \in C} c^T \overrightarrow{u} = \sigma_C(\overrightarrow{u}) + \sigma_C(-\overrightarrow{u}). \tag{2}$$

In geometrical terms, $e_C(\overrightarrow{u})$ represents the distance between two parallel lines orthogonal to $\overrightarrow{u}$ "squeezing" C (or tangent to C if the contact points are smooth).

We say that C is a convex set of constant width $\Delta > 0$ when $e_C(\overrightarrow{u}) = \Delta$ for all unitary vectors $\overrightarrow{u}$ in $\mathbb{R}^2$. Such compact convex sets of constant width in $\mathbb{R}^2$ are sometimes called *orbiforms* in the literature.

1°) Give at least two examples of compact convex sets in the plane with constant width.

2°) Prove that if C is of constant width Δ, then its perimeter $l(\Gamma)$ always equals $\pi\Delta$.

Hint. 2°) The length of the boundary Γ of C is given by the following formula:

$$l(\Gamma) = \int_0^{2\pi} \sigma_C(\cos\theta, \sin\theta)\ \mathrm{d}\theta, \tag{3}$$

i.e., the integral over $[0, 2\pi]$ of the support function σ_C of C in the unitary directions $\overrightarrow{u} = (\cos\theta, \sin\theta) \in \mathbb{R}^2$. This has already been used in Tapa 122.

Answers. 1°) Firstly the disks of diameter Δ. Secondly, the so-called curvilinear triangles of F. Reuleaux (1829–1905). The latter are built from equilateral triangles with sides of length Δ as follows: draw three circular arcs of radius Δ from the three vertices of the triangle; the resulting "inflated" triangle is a convex set of constant width Δ.

2°) From the relation (2), it follows by (3) that

$$l(\Gamma) = \int_0^{2\pi} \sigma_C(\cos\theta, \sin\theta)\ \mathrm{d}\theta + \int_0^{2\pi} \sigma_C(-\cos\theta, -\sin\theta)\ \mathrm{d}\theta = 2\pi\Delta. \tag{4}$$

But $\sigma_C(-\cos\theta, -\sin\theta) = \sigma_{-C}(\cos\theta, \sin\theta)$, so that the second integral in (4) is the perimeter of $-C$, which is the same as that of C. In brief, we infer from (4) that $2l(\Gamma) = 2\pi\Delta$.

Comments. - The surprising result of question 2°) is due to J. Barbier (1839–1889).

- Among the compact convex sets in the plane of constant width Δ, the disks are those of maximal area ($= \pi\frac{\Delta^2}{4}$) and the Reuleaux curvilinear triangles are those of minimal area ($= \frac{1}{2}(\pi - \sqrt{3})\Delta^2$).

---

124. ★★ *Convex sets of maximal area*

Let C be a compact convex set in the plane, whose boundary Γ is given by an equation in polar coordinates: $\rho = \rho(\theta), \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$; see Figure 1 below. We recall that the area $\mathcal{A}(C)$ of C is given by the formula

$$\mathcal{A}(C) = \frac{1}{2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \rho^2(\theta)\ \mathrm{d}\theta. \tag{1}$$

1°) Verify that

$$\mathcal{A}(C) = \frac{1}{2}\int_0^{\frac{\pi}{2}} \left[\rho^2(\theta) + \rho^2(\theta - \frac{\pi}{2})\right]\ \mathrm{d}\theta. \tag{2}$$

2°) Deduce that if the diameter of C is bounded from above by Δ, then $\mathcal{A}(C)$ is bounded from above by $\pi\frac{\Delta^2}{4}$.

Answers. 1°) We have:

$$\mathcal{A}(C) = \frac{1}{2}\int_{-\frac{\pi}{2}}^{0} \rho^2(\theta)\ \mathrm{d}\theta + \frac{1}{2}\int_{0}^{\frac{\pi}{2}} \rho^2(\theta)\ \mathrm{d}\theta.$$

A simple change of variable in the first integral, $\theta' = \theta + \pi/2$, leads to the announced formula (2).

2°) In the rectangular triangle OPQ (see Figure 1 below),

$$OP^2 + OQ^2 = PQ^2 \leqslant \Delta^2.$$

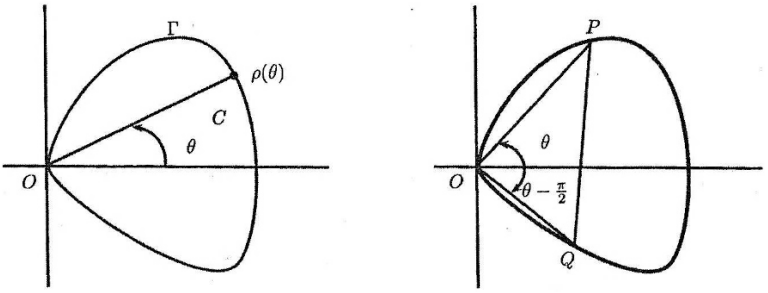

Figure 1

Hence, by using the expression (2) of $\mathcal{A}(C)$,

$$\mathcal{A}(C) \leqslant \frac{1}{2} \times \frac{\pi}{2}\Delta^2 = \pi\frac{\Delta^2}{4}.$$

As a result: among the compact convex sets in the plane whose diameter is bounded from above by Δ, the disks of diameter Δ are those which maximize the area.

Comment. The clever proof displayed in this Tapa 124 is due to J.E. Littlewood (1953).

125. ★★ *Measuring the volume of the polar of a convex set with the help of its support function*

In $\mathbb{R}^d$, we denote by λ the LEBESGUE measure ($\lambda(A)$ is therefore the "volume" of the set $A \subset \mathbb{R}^d$). Let $C \subset \mathbb{R}^d$ be a compact convex set with the origin 0 in its interior. We denote by C° the *polar set* of C, that is:

$$C^\circ = \left\{ y \in \mathbb{R}^d : y^T x \leqslant 1 \text{ for all } x \in C \right\}. \tag{1}$$

The set C° is again compact convex with 0 in its interior. We intend to evaluate the volume $\lambda(C^\circ)$ of the polar set C° with the help of the support function σ_C of C. Recall that the support function σ_C of C is defined as follows:

$$d \in \mathbb{R}^d \mapsto \sigma_C(d) = \max_{c \in C} c^T d. \tag{2}$$

1°) An example to begin with. Let $A \in \mathcal{S}_d(\mathbb{R})$ be positive definite, let $\mathcal{E}_A$ be the associated convex elliptic ball

$$\mathcal{E}_A = \left\{ x \in \mathbb{R}^d : x^T A x \leqslant 1 \right\}.$$

What is the support function of $\mathcal{E}_A$? the polar set $(\mathcal{E}_A)^\circ$ of $\mathcal{E}_A$? the volume of $\mathcal{E}_A$?

2°) Prove that

$$\lambda(C^\circ) = \frac{1}{d!} \int_{\mathbb{R}^d} e^{-\sigma_C(x)} \, \mathrm{d}x. \tag{3}$$

Hint. 2°) Use changes of variables in integrals; for example, given a measurable function $f : \mathbb{R}^+ \to \mathbb{R}^+$, if $\varphi : \mathbb{R}^d \to \mathbb{R}^+$ is convex positively homogeneous, then

$$\int_{\mathbb{R}^d} f\left[\varphi(x)\right] \, \mathrm{d}x = d \times \lambda(B) \times \int_{\mathbb{R}^+} r^{d-1} f(r) \, \mathrm{d}r, \tag{4}$$

where $B = \left\{ x \in \mathbb{R}^d : \varphi(x) \leqslant 1 \right\}$.

Answers. 1°) We have:

$$\begin{array}{rcl} u & \in & \mathbb{R}^d \longmapsto \sigma_{\mathcal{E}_A}(u) = \sqrt{u^T A^{-1} u}; \ (\mathcal{E}_A)^\circ = \mathcal{E}_{A^{-1}}; \\ \lambda(\mathcal{E}_A) & = & \dfrac{1}{\sqrt{\det A}} \times \dfrac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}. \end{array}$$

2°) Use formula (4) with $\varphi = \sigma_C$ (so that $\left\{x \in \mathbb{R}^d : \varphi(x) \leqslant 1\right\} = C^\circ$) and $f(r) = e^{-r}$. One then obtains

$$\int_{\mathbb{R}^d} e^{-\sigma_C(x)}\, \mathrm{d}x = d \times \lambda(C^\circ) \times (d-1)!$$

---

126. ★★ *Calculating two improper integrals*

Some integrals of continuous real-valued functions are called *improper* (or *generalized*) when: either the interval on which the integration is performed is unbounded, or, even if the interval of integration is bounded, the function to be integrated is unbounded at the endpoints of the interval. Beyond the existence of such integrals (called convergent in such cases), the main point is to be able to calculate them. In this Tapa, we taste two important pet techniques: integration by parts and change of variables.

1°) Let $f : \left(0, \frac{\pi}{2}\right] \to (-\infty, 0]$ be defined as $f(x) = \ln(\sin x)$. We intend to calculate the improper integral $J = \int_0^{\frac{\pi}{2}} f(x)\, \mathrm{d}x$.

(*a*) Use the trick $\sin(2u) = 2 \sin u \cdot \cos u$, so that replicas of J show up in the calculation of J.

(*b*) Determine J.

2°) Let $g : [0, +\infty) \to (0, +\infty)$ be defined as:

$$g(x) = \begin{cases} \left(\frac{\arctan x}{x}\right)^2 & \text{if } x \neq 0; \\ 1 & \text{if } x = 0. \end{cases}$$

We intend to calculate the improper integral $I = \int_0^{+\infty} g(x)\, \mathrm{d}x$.

(*a*) By performing successively an integration by parts, a change of variables and again an integration by parts, prove that I and J are very simply related.

(*b*) Determine I.

Hints. 1°) After using the decomposition $\ln(\sin x) = \ln 2 + \ln(\sin \frac{x}{2}) + \ln(\cos \frac{x}{2})$, perform simple changes of variables like $y = x/2$ and $z = \pi/2 - y$.

2°) Since $\frac{\arctan x}{x} \to 1$ when $x \to 0$, g is indeed a continuous function.

(*a*) Begin with an integration by parts on $g(x) = (\arctan x)^2 \times \frac{1}{x^2} = u(x) \times v'(x)$. Then perform the change of variable $y = \arctan x$. Finally, integrate by parts on $\frac{y}{\tan y} = y \times [\ln(\sin y)]' = u(y) \times v'(y)$.

Answers. 1°) We obtain

$$J = \frac{\pi \ln 2}{2} + 2\int_0^{\frac{\pi}{2}} \ln(\sin t)\ \mathrm{d}t = \frac{\pi \ln 2}{2} + 2J.$$

Hence, $J = -\frac{\pi \ln 2}{2}$.

2°) (a). We obtain

$$\begin{aligned} I &= 2\int_0^{+\infty} \frac{\arctan x}{x(1+x^2)}\ \mathrm{d}x \\ &\quad \text{(via the integration by parts proposed in Hints);} \\ &= 2\int_0^{\frac{\pi}{2}} \frac{y}{\tan y}\ \mathrm{d}y \\ &\quad \text{(via the change of variables } y = \arctan x\text{);} \\ &= -2\int_0^{\frac{\pi}{2}} \ln(\sin y)\ \mathrm{d}y \\ &\quad \text{(via the integration by parts proposed in Hints).} \end{aligned}$$

(b) We have that $I = -2J$, whence $I = \pi \ln 2$.

---

127. ★★ *Surjectivity of the normal unit to a compact surface in* $\mathbb{R}^n$

Consider $\mathbb{R}^n, n \geqslant 2$, equipped with the usual scalar product and the associated Euclidean norm $\|\cdot\|$. We denote by S the unit-sphere of $\mathbb{R}^n$. Let $h : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function, and let

$$\Sigma = \{x \in \mathbb{R}^n : h(x) = 0\}.$$

We suppose that Σ is nonempty, bounded, and that $\nabla h(x) \neq 0$ for all $x \in \Sigma$. The normal unit vector to Σ at $x \in \Sigma$ is

$$\nu(x) = \frac{\nabla h(x)}{\|\nabla h(x)\|}.$$

Prove that the mapping $\nu : \Sigma \to S$ is surjective.

Hints. - A drawing in the plane helps to understand what is going on.

- Let B be a closed ball containing Σ. Since $n \geqslant 2$, the complementary set $B^c = \{x \in \mathbb{R}^n : x \notin B\}$ is arcwise-connected; as a consequence, h is of a constant sign on the set B^c.

- Let $u \in S$ and let $f(x) = u^T x$. Consider the problem of maximizing $f(x)$ over $x \in \Sigma$: solutions do exist; at a maximizer $\overline{x}$, one can write a LAGRANGE optimality condition, *i.e.,* there exists a real λ such that $u = \lambda \nabla h(\overline{x})$. Then, study the function $t \in \mathbb{R} \mapsto h(\overline{x} + tu)$ to conclude that $\lambda = 1/\left\|\nabla h(\overline{x})\right\|$.

Comment. The assumption on the dimension n, namely $n \geqslant 2$, has been essential; it has been used to ensure that the complementary set of a ball is arcwise-connected. Here is a Tapa of the same vein, based on the same topological property.

Let $n \geqslant 2$, let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function.

(a) Suppose that f is surjective, *i.e.*, $f(\mathbb{R}^n) = \mathbb{R}$. Then the set $\Sigma = \{x \in \mathbb{R}^n : f(x) = 0\}$ is (closed but) unbounded.

(b) Suppose that f is convex and that $\Sigma = \{x \in \mathbb{R}^n : f(x) = 0\}$ is nonempty and bounded. Then $f(x) \to +\infty$ as $\|x\| \to +\infty$.

---

128. ★★ *A mean value theorem in equality form for differentiable vector-valued functions*

Let $X : [a, b] \subset \mathbb{R} \to \mathbb{R}^n$ be continuous on $[a, b]$ and differentiable on (a, b). We denote by $X'(t) \in \mathbb{R}^n$ the derivative of X at $t \in (a, b)$. We intend to express (exactly) the mean value $\frac{X(b)-X(a)}{b-a}$ of X on $[a, b]$ in terms of some derivatives $X'(t_i), i = 1, ..., k$.

1°) Let $d \in \mathbb{R}^n$ and $X_d : t \in [a, b] \mapsto X_d(t) = [X(t)]^T d$ (the "scalarized" version of X). By using the usual mean value theorem (in equality form) for the real-valued function X_d, show that:

$$\left[\frac{X(b) - X(a)}{b - a}\right]^T d \leqslant \sup_{t \in (a,b)} \left[X'(t)\right]^T d. \tag{1}$$

2°) (a) By using the geometrical form of the separation theorem (HAHN–BANACH) for convex sets in $\mathbb{R}^n$ or, equivalently in the present context, the properties of the orthogonal projection onto a closed convex set in $\mathbb{R}^n$, deduce from (1) that

$$\frac{X(b) - X(a)}{b - a} \in \overline{\text{co}}\,(X'(t) : t \in (a, b))\,. \tag{2}$$

(Here, $\overline{\text{co}}\,(S)$ denotes the closed convex hull of S.)

(b) Given any nonempty convex set $C \subset \mathbb{R}^n$, there exists a unique smallest affine set containing C; this set is called the affine hull of C. The relative interior of a convex set C, which is denoted by $\text{ri}(C)$, is

defined as the interior which results when C is regarded as a subset of its affine hull. Show now that the result (2) can be made more precise by proving that

$$\frac{X(b)-X(a)}{b-a} \in \text{ri } \left[\text{co}\left(X'(t) : t \in (a,b)\right)\right]. \tag{3}$$

As a consequence, following CARATHÉODORY's theorem, there are (at most) $n+1$ points $t_i \in (a,b)$ and $n+1$ nonnegative coefficients λ_i with sum 1 such that

$$\frac{X(b)-X(a)}{b-a} = \sum_{i=1}^{n+1} \lambda_i X'(t_i). \tag{4}$$

3°) Examples, limitations, extensions.

(*a*) Let $X : t \in [0,1] \mapsto X(t) = (t, t^2, t^3) \in \mathbb{R}^3$.
What is $D = \{X'(t) : t \in (0,1)\}$?

(*b*) Let $X : t \in [-1,1] \mapsto X(t) \in \mathbb{R}^2$ be defined as follows:

$$\begin{aligned} X(0) &= (0,0) \\ \text{and } X(t) &= \left(t^2 \cos\left(\frac{1}{t}\right), t^2 \sin\left(\frac{1}{t}\right)\right) \text{ if } t \neq 0. \end{aligned}$$

What is $D = \{X'(t) : t \in (-1,1)\}$?

(*c*) Suppose that X is continuously differentiable on (a,b). Show that there are (at most) n points $t_i \in (a,b)$ and n nonnegative coefficients λ_i with sum 1 such that:

$$\frac{X(b)-X(a)}{b-a} = \sum_{i=1}^{n} \lambda_i X'(t_i). \tag{5}$$

Hints. 2°) (*a*). The separation theorem (HAHN–BANACH) for convex sets in $\mathbb{R}^n$ can be translated analytically with the help of (the so-called) support functions $\sigma_S : d \in \mathbb{R}^n \mapsto \sigma_S(d) = \sup_{s \in S} s^T d$. Indeed, we have:

$$\begin{aligned} x &\in \overline{\text{co}}(S) \text{ if and only if } x^T d \leqslant \sigma_S(d) \text{ for all } d \in \mathbb{R}^n; \\ \text{if } S &= \cup_{i \in I} S_i, \text{ then } \sigma_{\overline{\text{co}}(\cup_{i \in I} S_i)} = \sup_{i \in I} \sigma_{S_i}. \end{aligned}$$

2°) (*b*). With $C = \overline{\text{co}}\left(X'(t) : t \in (a,b)\right)$, the aim is to eliminate the possibility that $\overline{X} = \frac{X(b)-X(a)}{b-a}$ lies on the relative boundary $\text{rbd}(C) = \overline{C} \setminus \text{ri}(C)$ of C. Supposing that $\overline{X} \in \text{rbd}(C)$, there exists a nontrivial

supporting hyperplane to $\overline{C}$ at $\overline{X}$, *i.e.,* one which does not contain C itself. In other words, there exists a $d_* \in \mathbb{R}^n$ such that:

$$\overline{X}^T d_* = \sigma_C(d_*);\ X^T d_* \leqslant \sigma_C(d_*) \text{ for all } X \in C; \tag{6}$$

$$X^T d_* < \sigma_C(d_*) \text{ for all } X \in \mathrm{ri}(C). \tag{7}$$

3°) Appeal to the Fenchel–Bunt theorem: If $S \subset \mathbb{R}^n$ has no more than n connected components (in particular, if S is connected), then any $x \in \mathrm{co}(S)$ can be expressed as a convex combination of n elements of S.

Answers. 2°) (*b*). Consider the direction d_* suggested in Hints, and let η_{d_*} be the function defined as $\eta_{d_*}(t) = [X(t)]^T d_* - \left(\overline{X}^T d_*\right) t$. Clearly, η_{d_*} is continuous on $[a,b]$ and differentiable on (a,b). Now, according to the definition of C and the properties of d_* displayed in (6) and (7), we have:

$$\eta'_{d_*}(t) \leqslant 0 \text{ for all } t \in (a,b);\ \eta'_{d_*}(t) < 0 \text{ for at least one } t \in (a,b).$$

Consequently, $\eta_{d_*}(b) < \eta_{d_*}(a)$; hence a contradiction is obtained.

3°) (*a*). We have that $D = \left\{(\alpha, \alpha, \frac{3\alpha^2}{4}) : \alpha \in (0,2)\right\}$. The mean value of X on $[0,1]$, that is, the vector $(1,1,1)$, lies in the convex hull of this curve D.

(*b*) We have: $X'(0) = (0,0)$ and for $t \neq 0$

$$X'(t) = \left(2t\cos\left(\frac{1}{t}\right) + \sin\left(\frac{1}{t}\right), 2t\sin\left(\frac{1}{t}\right) - \cos\left(\frac{1}{t}\right)\right).$$

The length of $X'(t), t \neq 0$, is $\sqrt{1+4t^2} \geqslant 1$. Thus, the image set D of the derivatives $X'(t)$ is made of two pieces: a curve which "lives" outside the unit circle of $\mathbb{R}^2$ and comes to wind round it, and the isolated point $(0,0)$. The mean value of X on $[-1,1]$, that is, the vector $(\sin 1, 0)$, lies in the convex hull of these two pieces.

In this example, D is not connected; it therefore serves to show that Darboux's theorem for the derivatives of real-valued functions (that is: the image of an interval by a derivative function is an interval) does not extend to vector-valued functions.

(*c*) When X is continuously differentiable on (a,b), the image-set $D = (X'(t) : t \in (a,b)) \subset \mathbb{R}^n$ is connected, so that any vector in its convex hull can be expressed as a convex combination of n elements in D (the Fenchel–Bunt theorem, *cf.* Hints in Tapa 138). Whence (5) is derived.

Comments. - The result expressed in (2) remains valid for a BANACH-valued mapping X; the proof follows the same path.

- DINI type results also hold, like

$$\overline{\text{co}}\,(X'(t) : t \in (a,b)) = \overline{\text{co}}\left(\frac{X(d)-X(c)}{d-c} : a \leqslant c < d \leqslant b\right). \qquad (8)$$

Determining the set of slopes $\left(\frac{X(d)-X(c)}{d-c} : a \leqslant c < d \leqslant b\right)$ can be fairly complicated; see, for example, Tapa 293 in

J.-B. HIRIART-URRUTY, *Mathematical tapas,* Volume 1 (for Undergraduates). Springer (2016).

In (7) one could even replace the derivatives $X'(t)$ with right-derivatives $X'_+(t)$ or left-derivatives $X'_-(t)$.

For more on mean value theorems in equality forms for vector-valued functions, see:

J.-B. HIRIART-URRUTY, *Théorèmes de valeur moyenne sous forme d'égalité pour les fonctions à valeurs vectorielles.* Revue de Mathématiques Spéciales n°7 (1983), 287–293.

---

129. ★★ *On the numerical range of a matrix* $A \in \mathcal{M}_n(\mathbb{C})$

Let $\mathbb{C}^n$ be equipped with the usual Hermitian inner product

$$(u = (u_1, ..., u_n) \in \mathbb{C}^n,\ v = (v_1, ..., v_n) \in \mathbb{C}^n) \mapsto \langle u, v\rangle = \sum_{i=1}^{n} \overline{u_i} v_i,$$

and the associated norm $u \mapsto \|u\| = \sqrt{\langle u, u\rangle}$.

For $A = [a_{ij}] \in \mathcal{M}_n(\mathbb{C})$, we set

$$\mathcal{H}(A) = \left\{ \langle u, Au\rangle = \sum_{i,j=1}^{n} a_{ij} u_i \overline{u_j} : \|u\| = 1 \right\}. \qquad (1)$$

This set is called the *hausdorffian set* or *numerical range* or *field of values* of A.

1°) First examples of computations.

(a) What is $\mathcal{H}(A)$ when A is the diagonal matrix $\text{diag}(\lambda_1, ..., \lambda_n)$ (where the λ_i's are complex numbers)?

(b) Let $U \in \mathcal{M}_n(\mathbb{C})$ be a unitary matrix. Check that

$$\mathcal{H}(U^* A U) = \mathcal{H}(A).$$

(c) Show that $\mathcal{H}(A)$ always contains the eigenvalues of A.

(d) Denote by $\overline{A}$ the matrix $[\overline{a_{ij}}]$ obtained from $A = [a_{ij}] \in \mathcal{M}_n(\mathbb{C})$ by taking the conjugates of the entries; similarly, let

$$\overline{\mathcal{H}(A)} = \{\overline{u} : u \in \mathcal{H}(A)\}.$$

Check that:

$$\mathcal{H}(\overline{A}) = \overline{\mathcal{H}(A)}. \tag{2}$$

As a consequence, if $A \in \mathcal{M}_n(\mathbb{R})$, the set $\mathcal{H}(A)$ in $\mathbb{C}$ is symmetric with respect to the real axis.

2°) The case $n = 2$.

Let $A \in \mathcal{M}_2(\mathbb{C})$. Prove that $\mathcal{H}(A)$ is a convex elliptic set whose foci are the eigenvalues of A.

3°) The case of normal matrices.

Let $A \in \mathcal{M}_n(\mathbb{C})$ be assumed normal. Show that $\mathcal{H}(A)$ is the convex polygon in $\mathbb{C}$ generated by the eigenvalues λ_i of A, that is to say, the convex hull of $\lambda_1, ..., \lambda_n$.

As a particular case, when A is Hermitian, $\mathcal{H}(A)$ is a line-segment on the real axis, whose endpoints are the extreme eigenvalues of A.

Hints. Since, by definition, $\mathcal{H}(A)$ is the image of the unit sphere in $\mathbb{C}^n$ under the continuous mapping $u \mapsto \langle u, Au\rangle$, $\mathcal{H}(A)$ is a compact connected subset of $\mathbb{C}$.

2°) Use the (important) result stating that A can be "reduced" to an upper triangular matrix: There is a unitary matrix $U \in \mathcal{M}_2(\mathbb{C})$ such that $U^*AU = \begin{bmatrix} \lambda_1 & a \\ 0 & \lambda_2 \end{bmatrix}$. Therefore, it is sufficient to carry out calculations in $\mathcal{H}(T)$ for upper triangular matrices T.

3°) Among the various definitions or characterizations of normal matrices, we choose the following one: A is normal if and only if there exists a unitary matrix U such that $U^*AU = \mathrm{diag}(\lambda_1, ..., \lambda_n)$, where the λ_i's are eigenvalues of A.

Answers. 1°) (a). $\mathcal{H}(A)$ is a convex polygon in $\mathbb{C}$, namely the convex hull of $\lambda_1, ..., \lambda_n$.

(b) We have, for all $u \in \mathbb{C}^n$,

$$\langle u, (U^*AU)\, u\rangle = \langle (Uu)\,, A\,(Uu)\rangle\,.$$

Make the substitution $v = Uu$. Since U is unitary, the unit sphere in $\mathbb{C}^n$ is invariant under this change of variable. Therefore,

$$\{\langle u, (U^*AU)\, u\rangle : \|u\| = 1\} = \{\langle v, Av\rangle : \|v\| = 1\}\,.$$

(c) If λ is an eigenvalue of A and if u is a unit eigenvector associated with it, we have: $\langle u, Au\rangle = \lambda \|u\|^2 = \lambda$.

2°) Let $T = U^*AU = \begin{bmatrix} \lambda_1 & a \\ 0 & \lambda_2 \end{bmatrix}$, where λ_1 and λ_2 are the eigenvalues of A. Since $\mathcal{H}(A) = \mathcal{H}(T)$, it suffices to determine $\mathcal{H}(T)$ in all particular cases.

- If $\lambda_1 = \lambda_2$ $(= \lambda)$, we have

$$\mathcal{H}(T) = \left\{ z \in \mathbb{C} : |z - \lambda| \leqslant \frac{|a|}{2} \right\},$$

a disk centered at λ and of radius $\frac{|a|}{2}$.

- If $\lambda_1 \neq \lambda_2$ and $a = 0$, T is a diagonal matrix and we have already proved that $\mathcal{H}(T)$ is the convex hull of λ_1, λ_2, that is to say, the line-segment joining λ_1 and λ_2.

- If $\lambda_1 \neq \lambda_2$ and $a \neq 0$. We set $\frac{\lambda_1 - \lambda_2}{2} = \rho e^{i\theta}$, so that:

$$T - \frac{\lambda_1 + \lambda_2}{2} = \rho e^{i\theta} \begin{bmatrix} \rho & ae^{-i\theta} \\ 0 & -\rho \end{bmatrix} (= \rho e^{i\theta} S).$$

$\mathcal{H}(S)$ is an elliptic set centered at the origin $(0, 0)$, with minor axis $|a|$, and with foci at $(\rho, 0)$ and $(-\rho, 0)$. Consequently, $\mathcal{H}(T)$ is an elliptic set with foci at λ_1 and λ_2, and the major axis making an angle θ with the real axis.

Comments. - A more advanced result, due to Toeplitz and Hausdorff (1918–1919) states that $\mathcal{H}(A)$ is always *convex;* therefore, $\mathcal{H}(A)$ is a compact convex set which always contains the convex polygon in $\mathbb{C}$ generated by the eigenvalues λ_i of A, that is to say, the convex hull of $\lambda_1, ..., \lambda_n$. However, the "gap" between $\mathcal{H}(A)$ and co $(\lambda_1, ..., \lambda_n)$ may be large, for non-normal matrices, as examples even for $n = 2$ show (see the result in 2°)).

- One proof of the convexity of $\mathcal{H}(A)$ consists in reducing the general situation to the case when $n = 2$ (discussed in Question 2°)).

- Another (easy to prove) general property showing what $\mathcal{H}(A)$ looks like concerns block-diagonal matrices. If $A_1 \in \mathcal{M}_{n_1}(\mathbb{C})$ and $A_2 \in$

$\mathcal{M}_{n_2}(\mathbb{C})$, and if $A \in \mathcal{M}_{n_1+n_2}(\mathbb{C})$ is constructed as follows

$$A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix},$$

then

$$\mathcal{H}(A) = \operatorname{co}\{\mathcal{H}(A_1) \cup \mathcal{H}(A_2)\}. \tag{3}$$

---

130. ★★ *Asymptotic behavior of the solutions of a linear differential system under the* LYAPUNOV *condition*

Consider the following linear differential system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Ax, \tag{$\mathcal{S}$}$$

where $A \in \mathcal{M}_n(\mathbb{R})$ satisfies the following so-called LYAPUNOV condition:

$$\left\{ \begin{array}{l} \text{There exists a positive definite } X \in \mathcal{S}_n(\mathbb{R}) \text{ such that the (symmetric)} \\ \text{matrix } A^T X + XA \text{ is negative definite.} \end{array} \right. \tag{$\mathcal{L}$}$$

Let $\varepsilon > 0$ be chosen so that $A^T X + XA + \varepsilon X$ is still negative definite.

1°) Show that any solution $x(\cdot)$ of $(\mathcal{S})$ satisfies:

$$\langle Xx(t), x(t)\rangle \leqslant \langle Xx(0), x(0)\rangle\, e^{-\varepsilon t} \text{ for all } t \geqslant 0. \tag{1}$$

(Here, $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $\mathbb{R}^n$.)

2°) Deduce that:

$$\|x(t)\| \leqslant \|x(0)\| \times \sqrt{\frac{\lambda_{\max}(X)}{\lambda_{\min}(X)}} \times e^{-\frac{\varepsilon}{2}t} \text{ for all } t \geqslant 0, \tag{2}$$

where $\lambda_{\max}(X)$ and $\lambda_{\min}(X)$ stand for, respectively, the largest and the smallest eigenvalues of X.

Hints. 1°) Consider the derivative of the function $t \mapsto u(t) = \langle Xx(t), x(t)\rangle$. With the help of the LYAPUNOV condition (the matrix $A^T X + XA + \varepsilon X$ is negative definite), show that

$$\frac{\mathrm{d}u}{\mathrm{d}t} + \varepsilon u(t) \leqslant 0 \text{ for all } t \in \mathbb{R},$$

or, equivalently,

$$e^{\varepsilon t}\frac{\mathrm{d}u}{\mathrm{d}t} + \varepsilon e^{\varepsilon t}u(t) \leqslant 0 \text{ for all } t \in \mathbb{R}.$$

We recognize in $e^{\varepsilon t}\frac{\mathrm{d}u}{\mathrm{d}t} + \varepsilon e^{\varepsilon t}u(t)$ the derivative of $t \mapsto e^{\varepsilon t}u(t)$.

2°) When $S \in \mathcal{S}_n(\mathbb{R})$, one has

$$\lambda_{\min}(S) \times \|v\|^2 \leqslant \langle Sv, v\rangle \leqslant \lambda_{\max}(S) \times \|v\|^2 \text{ for all } v \in \mathbb{R}^n.$$

Comment. Condition $(\mathcal{L})$ is clearly satisfied if the matrix A is symmetric negative definite. In that case, the differential system $(\mathcal{S})$ is nothing else than

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \nabla q(x),$$

where $x \in \mathbb{R}^n \mapsto q(x) = \frac{1}{2}x^T Ax$, a (strictly) concave quadratic form. Then, according to inequality (2), any solution $t \geqslant 0 \mapsto x(t)$ tends to 0 (the unique maximizer of q) as $t \to +\infty$.

---

131. ★★ *Measuring the shifts maintaining the same values in a continuous function*

Let $f : [0,1] \to \mathbb{R}$ be a continuous function satisfying $f(0) = f(1) = 0$. We consider the following two closed sets

$$\begin{aligned} S &= \{h \ : \ f(x+h) = f(x) \text{ for some } x \in [0,1]\}, \\ T &= \{h' \ : \ 1 - h' \in S\} \text{ (image of } S \text{ under the reflection } h \mapsto 1-h). \end{aligned}$$

1°) Prove that $S \cup T = [0,1]$, that is to say: any $h \in [0,1]$ is either in S or in T.

2°) Deduce that the common LEBESGUE measure of S and T must be at least $1/2$.

Hint. 1°) Consider the continuous function $g_h : [0,1] \to \mathbb{R}$ defined by:

$$g_h(x) = \begin{cases} f(x+h) - f(x) & \text{if } x + h \leqslant 1; \\ f(x+h-1) - f(x) & \text{if } x + h > 1. \end{cases}$$

Consider a point x_2, lying between a minimizer x_0 and a maximizer x_1 of f, where $g_h(x_2) = 0$.

---

132. ★★ *Dividing a set of positive* LEBESGUE *measure*

Let λ denote the LEBESGUE measure on $\mathbb{R}$. Consider a LEBESGUE measurable set $S \subset \mathbb{R}$ such that $\lambda(S) = 1$. Given $0 < \alpha < 1$, we intend to show that there is a LEBESGUE measurable set $A_\alpha \subset S$ such that $\lambda(A_\alpha) = \alpha$.

For that purpose, define the function $f : \mathbb{R} \to [0, 1]$ as follows:

$$f(x) = \lambda(S \cap (-\infty, x]),\ x \in \mathbb{R}.$$

1°) Properties of the function f. Check that:

$$\begin{array}{c} |f(x) - f(y)| \leqslant |x - y| \text{ for all } x, y \text{ in } \mathbb{R}; \\ f(x) \to 0 \text{ when } x \to -\infty; \\ f(x) \to 1 \text{ when } x \to +\infty. \end{array} \tag{1}$$

2°) Deduce from the properties above that there exists an $x_\alpha \in \mathbb{R}$ such that the set $A_\alpha = S \cap (-\infty, x_\alpha]$ has LEBESGUE measure α.

---

133. ★★ *A strict convexity or concavity criterion via the mean value theorem*

Let $f : I \to \mathbb{R}$ be assumed differentiable on the open interval I. We suppose the following: for all $a < b$ in I, there is only one $c \in (a, b)$ such that

$$\frac{f(b) - f(a)}{b - a} = f'(c). \tag{1}$$

In other words, the intermediate c appearing in the expression of the mean value theorem is unique.

Show that either f or $-f$ is strictly convex on I.

Hints. - A first approach. Prove firstly, by contradiction for example, that f' is injective. Then, remember that f' satisfies the intermediate value property (DARBOUX theorem): the image $f'(J)$ of any interval J under f' is an interval. As a result, f' should be strictly monotone and continuous.

- A second approach. Prove, as an intermediate step, that there is no (non-trivial) line-segment on the graph of f. As a consequence, the continuous mapping

$$(2)\ \left\{ \begin{array}{l} \varphi : (0,1) \times \{(x, y) \in I \times I : x < y\} \to \mathbb{R} \\ (\lambda, x, y) \mapsto \varphi(\lambda, x, y) = f\left[\lambda x + (1 - \lambda)y\right] - \lambda f(x) - (1 - \lambda) f(y) \end{array} \right.$$

transforms the connected (even convex) subset

$$(0,1) \times \{(x,y) \in I \times I : x < y\}$$

into an interval which does not contain 0.

Comments. - The converse of the proved property clearly holds true: If either f or $-f$ is strictly convex on I, there then exists only one intermediate c appearing in the expression (1) of the mean value theorem.

- If one strengthens the assumption of the Tapa by requiring that the unique c for which (1) holds true is always $(a+b)/2$, then f is necessarily quadratic, that is $f(x) = \alpha x^2 + \beta x + \gamma$, with $\alpha \neq 0$. This is therefore a characterization of quadratic functions via the mean value theorem.

In the same vein, if one requires that the unique c for which (1) holds true is always $pa + (1-p)b$, with the same $p \in (0,1)$ but $p \neq \frac{1}{2}$, then, surprisingly enough, one arrives at an impossibility...

---

134. ★★ *A convexity criterion via the mean value theorem*

Let $f : I \to \mathbb{R}$ be assumed differentiable on the open interval I. We suppose the following: for all $a < b$ in I, the set of $c \in (a,b)$ for which

$$\frac{f(b) - f(a)}{b - a} = f'(c) \tag{1}$$

is always an interval.

Show that either f or $-f$ is convex on I.

Hint. The following intermediate step could be helpful: Prove that, whenever there are 3 distinct points in the graph of f, then the part of the graph of f linking these 3 points is a line-segment.

Comments. - The converse of the proved property clearly holds true: If either f or $-f$ is convex on I, then the set of intermediate c for which the mean value property (1) holds is always an interval. We therefore have a characterization of convexity via the mean value theorem.

- One "extreme" case is when the set of intermediate c for which (1) holds is a singleton; in that case, either f or $-f$ is strictly convex on I (see the preceding Tapa). The second "extreme" case is when the set of c for which (1) holds is the whole of (a,b); in that case f is affine on I.

---

135. ★★ *Eigenvalues of AB with a positive definite A*

Let A and B be in $\mathcal{S}_n(\mathbb{R})$; we assume moreover that A is positive definite.

1°) Prove that all the eigenvalues of AB are real.

2°) Show that the largest eigenvalue $\lambda_{\max}(AB)$ of AB can be expressed as follows

$$\lambda_{\max}(AB) = \sup_{0 \neq x \in \mathbb{R}^n} \frac{x^T B x}{x^T A^{-1} x}. \tag{1}$$

3°) Show that AB can be diagonalized.

Hints. 1°) – 3°). Make use of the symmetric matrix $C = A^{1/2} B A^{1/2}$. Indeed: the eigenvalues of AB are those of C; if an orthogonal matrix U diagonalizes C, then $A^{1/2}U$ digonalizes AB.

2°) For $M \in \mathcal{S}_n(\mathbb{R})$, the variational formulation of its largest eigenvalue $\lambda_{\max}(M)$ is $\lambda_{\max}(M) = \sup_{0 \neq x \in \mathbb{R}^n} \frac{x^T M x}{\|x\|^2}$.

Comment. The results of this Tapa are just generalizations of properties which are known for $B \in \mathcal{S}_n(\mathbb{R})$, *i.e.*, the context of the Tapa with just $A = I_n$.

---

136. ★★ *A multilinear version of the* Cauchy–Schwarz *inequality*

Let $k \geqslant 2$ be a positive integer and let

$$H : \overbrace{\mathbb{R}^n \times \mathbb{R}^n \ldots \times \mathbb{R}^n}^{k \text{ times}} \to \mathbb{R}$$

be a symmetric multilinear (or k-linear) form. We suppose that there exists a (symmetric) positive semidefinite $n \times n$ matrix A such that

$$|H(x, x, ..., x)| \leqslant \left(x^T A x\right)^{\frac{k}{2}} \text{ for all } x \in \mathbb{R}^n. \tag{1}$$

Prove that

$$[H(x_1, x_2, ..., x_k)]^2 \leqslant \prod_{i=1}^{k} \left(x_i^T A x_i\right) \text{ for all } x_1, x_2, ..., x_k \text{ in } \mathbb{R}^n. \tag{2}$$

Comment. This result, due to A. NEMIROVSKI and Y. NESTEROV (1994), generalizes the symmetric bilinear case (*i.e.*, with $k = 2$) treated in Tapa 305 in

J.-B. HIRIART-URRUTY, *Mathematical tapas*, Volume 1 (for Undergraduates). Springer (2016).

---

137. ★★ *Mid-point convexity vs convexity of sets*

Let $S \subset \mathbb{R}^n$ satisfy the following property:

$$(x \in S, y \in S) \Rightarrow \left(\frac{x+y}{2} \in S \right).$$

1°) Does this imply that S is convex?

2°) Consider the same question if, moreover, we assume that S is closed.

Answers. 1°) No. Take, for example, the set $S \subset \mathbb{R}$ consisting of all rational numbers between 0 and 1.

2°) The answer is Yes now. With x_1 and x_2 taken in S, any x in the line-segment $[x_1, x_2]$ joining x_1 to x_2 lies in a sequence of sub-segments $\left[x_1^{(n)}, x_2^{(n)}\right]$ of $[x_1, x_2]$, where all the end-points $x_i^{(n)}$ are in S.

---

138. ★★ *Beyond* CARATHÉODORY*'s theorem for convex hulls of sets*

The convex hull co$S$ of a set S can be described as the set of all convex combinations of elements of S. If $S \subset \mathbb{R}^n$, any $x \in \text{co}S$ can be represented as a convex combination of (at most) $n + 1$ elements of S, this is CARATHÉODORY's theorem. Question: could we go beyond, that is to say, could we lower this upper bound $n + 1$ in some cases?

Prove that if $S \subset \mathbb{R}^n$ has no more than n connected components (in particular if S is connected), then any $x \in \text{co}S$ can be expressed as a convex combination of (at most) n elements of S.

Comments. This result, due to W. FENCHEL and L. BUNT, is geometrically very suggestive. For example, in the plane $\mathbb{R}^2$, the convex hull of a continuous curve can be obtained by just joining all pairs of points in it. In $\mathbb{R}$, the result simply says that connectedness and convexity are equivalent properties.

---

139. ★★ *More on the structure of extreme points of a convex set*

Let $C \subset \mathbb{R}^n$ be a convex set. We say that $x \in C$ is an *extreme point* of C if there are no two different points x_1 and x_2 in C such that $x = 1/2(x_1 + x_2)$.

1°) By maximizing the continuous function $x \mapsto \|x\|^2$ on C, show that the set ext$C$ of extreme points of C is nonempty.

2°) When $C \subset \mathbb{R}^2$, the set of extreme points of C is closed. We show with an example that this is no longer true in higher dimensions.

Let $C \subset \mathbb{R}^3$ be defined as the convex hull of the two points $(0, 0, 1)$, $(0, 0, -1)$ and the circle with equation

$$x^2 + (y - 1)^2 = 1 \text{ and } z = 0. \tag{1}$$

Draw this set C and determine its extreme points.

Hint. 1°) By using the strict convexity of the function $x \mapsto \|x\|^2$, show that any maximizer of this function on C is necessarily an extreme point of C.

Answer. 2°) C is shaped like a hard candy. Its extreme points are $(0, 0, 1), (0, 0, -1)$ and all the points of the circle described in (1) except the origin $(0, 0, 0)$.

---

140. ★★ *On extreme points of convex sets*

- Let C be defined as the convex hull of S. Show that an extreme point of C necessarily belongs to S.

- Let C be convex. Show that $x \in C$ is an extreme point of C if and only if the set $\{y \in C, y \neq x\}$ is convex.

Comment. These results are easily illustrated with the example of a compact convex polyhedron $C = \text{co}\{a_1, .., a_k\}$: the extreme points of C (also called vertices in the polyhedral case) are among the given a_i's, but not all the a_i's are extreme points of C.

---

141. ★★ *Projecting onto the epigraph of a convex function*

Let H be a HILBERT space and let $f : H \to \mathbb{R}$ be a differentiable convex function. The so-called *epigraph* C_f of f (literally "what is above the graph of f") is the closed convex subset of $H \times \mathbb{R}$ defined as:

$$C_f = \{(x, y) \in H \times \mathbb{R} : f(x) \leqslant y\}.$$

We want to orthogonally project $(a, b) \in H \times \mathbb{R}$ onto C_f; we therefore suppose that $(a, b) \notin C_f$. We call $(\overline{x}, \overline{y})$ the projection of (a, b) on C_f.

1°) Show that necessarily $\overline{y} = f(\overline{x})$.

2°) Prove that $\overline{x}$ is solution of the equation

$$a - \overline{x} + [b - f(\overline{x})]\, \nabla f(\overline{x}) = 0. \tag{1}$$

3°) Example 1. Let $H = \mathbb{R}$ and $f(x) = x^2$. Check that the projection $(\overline{x}, \overline{x}^2)$ of $(a, 0)$ onto the epigraph of f is made up from the real solution $\overline{x}$ of the cubic equation

$$2x^3 + x - a = 0. \tag{2}$$

4°) Example 2. Let $H = \mathbb{R}^2$ and $f(x, y) = x^2 + y^2$. Determine the projection of $\left(1, \sqrt{3}, \frac{1}{2}\right)$ onto the epigraph of f.

Hints. 2°) Use the following fact: the vector $(a - \overline{x}, b - f(\overline{x})) \in H \times \mathbb{R}$ should be "normal" to C_f at the point $(\overline{x}, f(\overline{x}))$, hence colinear to the vector $(\nabla f(\overline{x}), -1)$.

3°) Equation (2) has indeed only one real solution.

Answers. 4°) The projection is $\left(\frac{1}{2}, \frac{\sqrt{3}}{2}, 1\right)$.

Comment. The result of Question 2°) can be extended to arbitrary continuous convex functions $f : H \to \mathbb{R}$. In such a case, the relation (1) becomes

$$\frac{a - \overline{x}}{b - f(\overline{x})} \in -\partial f(\overline{x}), \tag{3}$$

where ∂f stands for the so-called subdifferential of f (a set-valued extension of the notion of gradient for convex functions).

142. ★★ *Weak convergence vs strong convergence of a sequence in a* HILBERT *space*

Let us consider a HILBERT space $(H, \langle \cdot, \cdot \rangle)$. The norm used in H is the one derived from the inner product, *i.e.,* $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$. We say that a sequence (u_n) of elements in H

- *converges strongly* towards $u \in H$ when $\|u_n - u\| \to 0$ as $n \to +\infty$. We then write, as usual, $u_n \to u$.

- *converges weakly* towards $u \in H$ when, for all $v \in H$, $\langle u_n, v \rangle \to \langle u, v \rangle$ as $n \to +\infty$. We then write $u_n \rightharpoonup u$.

Prove the next properties $1, 2, 3, 5, 7, 9, 10$. To prove property N, one can use properties $1, 2, .., N-1$.

1. If $u_n \rightharpoonup u$ and $u_n \rightharpoonup u'$, then $u = u'$ (*i.e.,* if the weak limit of (u_n) exists, it is unique).

2. Strong convergence implies weak convergence. The converse is true if H is finite-dimensional.

3. $(u_n \to u) \Leftrightarrow (u_n \rightharpoonup u \text{ and } \|u_n\| \to \|u\|)$.

4. Every weakly convergent sequence is (strongly) bounded.

5. If $u_n \to u$ and $v_n \rightharpoonup v$, then $\langle u_n, v_n \rangle \to \langle u, v \rangle$.

6. From every (strongly) bounded sequence one can extract a weakly convergent subsequence.

7. If A is a continuous linear mapping from H_1 into H_2 (H_1 and H_2 being two HILBERT spaces), and if $u_n \rightharpoonup u$ in H_1, then $Au_n \rightharpoonup Au$ in H_2.

8. If $u_n \rightharpoonup u$, there then exists a subsequence (u_{k_n}) extracted from the sequence (u_n) such that

$$\frac{u_{k_1} + u_{k_2} + ... + u_{k_n}}{n} \to u \text{ when } n \to +\infty.$$

9. If (u_n) is bounded in H, and if $\langle u_n, w \rangle \to \langle u_n, w \rangle$ for all w in a dense subset of H, then $u_n \rightharpoonup u$.

10. If $u_n \rightharpoonup u$, then $\|u\| \leqslant \liminf_{n \to +\infty} \|u_n\|$.

Hints. As usual in such a context, two mathematical tools play an instrumental role here:

- the basic expansion $\|a + b\|^2 = \|a\|^2 + \|b\|^2 + 2\langle a, b \rangle$;

- the CAUCHY–SCHWARZ inequality: $|\langle a, b \rangle| \leqslant \|a\| \times \|b\|$.

Comments. - Property 9 is of great practical interest: to have to prove that $\langle u_n - u, w\rangle \to 0$ only for some "nicer" elements w in a dense subset W in H may make life easier.

- According to Properties 10 and 3, what is missing for a weakly convergent sequence (u_n) to u, to be strongly convergent to u, is the inequality

$$\limsup_{n\to+\infty} \|u_n\| \leqslant \|u\| .$$

---

143. ★★ *Obstacles preventing a weakly convergent sequence to converge (strongly)*

With the help of the LEBESGUE space $L^2(I)$, I an interval in $\mathbb{R}$, structured as a HILBERT space with the scalar product

$$\langle f, g\rangle = \int_I f(x)g(x)\ \mathrm{d}x,$$

and the derived norm $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$, we illustrate three typical situations where $(u_n) \subset L^2(I)$ converges weakly to 0 but does not converge strongly to 0.

1°) *Oscillations.* Let $I = (0, \pi)$ and let $u_n \in L^2(I)$ be defined as

$$u_n(x) = \sqrt{\frac{2}{\pi}}\sin(nx).$$

2°) *Concentration.* Let $I = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and let $u_n \in L^2(I)$ be defined as

$$u_n(x) = \begin{cases} \sqrt{n} & \text{if } -\frac{1}{2n} \leqslant x \leqslant \frac{1}{2n}, \\ 0 & \text{otherwise.} \end{cases}$$

3°) *Evanescence.* Let $I = \mathbb{R}$ and let $u_n \in L^2(I)$ be defined as

$$u_n(x) = \begin{cases} \sqrt{n} & \text{if } n - \frac{1}{2n} \leqslant x \leqslant n + \frac{1}{2n}, \\ 0 & \text{otherwise.} \end{cases} .$$

Answers. 1°) The proposed sequence is an orthonormal sequence in $L^2(I)$; it then follows from the BESSEL inequality that it converges weakly to 0 (see Tapa 144 below, for example). At the same time, $\|u_n\| = 1$ for all $n \geqslant 1$.

2°) The sequence (u_n) is bounded since $\|u_n\| = 1$ for all $n \geqslant 1$. Let W be the set of continuous functions with a compact support contained in I. It is easy to see that, whenever $g \in W$,

$$\langle u_n, g\rangle = \int_I u_n(x)g(x)\ \mathrm{d}x \to 0 \text{ as } n \to +\infty.$$

Since W is dense in $L^2(I)$, we are therefore sure that the sequence (u_n) weakly converges to 0 (*cf.* Property 9 in Tapa 142).

This illustrates the phenomenon of "concentration at the origin 0" of the energy (or mass) of u_n.

3°) This is a situation very similar to the previous one. Here again, $u_n \rightharpoonup 0$ in $L^2(\mathbb{R})$ but $\|u_n\| \not\to 0$.

This illustrates the phenomenon of "evanescence at infinity" of the energy (or mass) of u_n.

---

144. ★★ *An orthonormal sequence always converges weakly to zero*

Let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal sequence in the (infinite-dimensional) HILBERT space $(H, \langle\cdot,\cdot\rangle)$. The norm used in H is the one derived from the inner product, that is $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$.

1°) Recall what the BESSEL inequality expresses.

2°) Show that the sequence $(e_n)_{n\in\mathbb{N}}$ necessarily converges weakly to 0, that is to say:

$$\lim_{n\to+\infty} \langle e_n, x\rangle = 0 \text{ for all } x \in H.$$

Answer. 1°) For all $x \in H$, the numerical series with general term $\langle e_n, x\rangle^2$ is convergent and

$$\sum_{n=0}^{+\infty} \langle e_n, x\rangle^2 \leqslant \|x\|^2 .$$

Comment. A strange situation indeed: the sequence $(e_n)_{n\in\mathbb{N}}$ "lives" on the unit-sphere of H (since $\|e_n\| = 1$ for all n), but it converges weakly towards the center 0 of this sphere!

An example is with $H = l^2(\mathbb{R})$ structured as a HILBERT space with the inner product

$$\langle x, y\rangle = \sum_{n=1}^{+\infty} x_n y_n, \text{ for } x = (x_n)_{n\geqslant 1} \text{ and } y = (y_n)_{n\geqslant 1} \text{ in } H.$$

Let $(e_n)_{n\geqslant 1}$ be the sequence in H defined as follows: for $n \geqslant 1$, $e_n = (0, ..., 0, 1, 0....)$ [1 at the n-th position, 0 everywhere else]. Then, $(e_n)_{n\in\mathbb{N}}$ is an orthonormal sequence in H.

---

145. ★★ *Weak limits of sequences lying on spheres*

Consider an infinite-dimensional HILBERT space $(H, \langle\cdot,\cdot\rangle)$; as usual, $\|\cdot\|$ denotes the norm in H derived from the inner product $\langle\cdot,\cdot\rangle$.

1°) Let (x_n) be a sequence in H such that:

$$\left\{\begin{array}{c}\text{The real sequence } \|x_n\| \text{ has a limit, denoted by } \ell, \text{ as } n\to+\infty;\\ (x_n) \text{ converges weakly towards } x \text{ as } n\to+\infty.\end{array}\right. \tag{1}$$

1°) Prove that $\|x\| \leqslant \ell$.

2°) A sort of converse to the previous result. Let $\ell > 0$ and let $x \in H$ be such that $\|x\| < \ell$. We intend to prove that there exists a sequence (x_n) in H such that:

$$\left\{\begin{array}{c}\|x_n\| = \ell \text{ for all } n;\\ (x_n) \text{ converges weakly towards } x \text{ as } n\to+\infty.\end{array}\right. \tag{2}$$

Consider $H_1 = x^\perp$, structured as a HILBERT space with the help of just the restriction of $\langle\cdot,\cdot\rangle$. Let now $(e_n)_{n\in\mathbb{N}}$ be an orthonormal sequence in the (infinite-dimensional) HILBERT space $(H_1, \langle\cdot,\cdot\rangle)$, and define

$$x_n = x + \sqrt{\ell^2 - \|x\|^2}\, e_n.$$

Prove that the sequence (x_n) above satisfies the requirements in (2).

Hints. 1°) Use the CAUCHY–SCHWARZ inequality $\langle x_n, x\rangle \leqslant \|x_n\| \times \|x\|$.

Without assuming that the sequence $(\|x_n\|)$ has a limit, a more general result is the following one: If $x_n \rightharpoonup x$, then $\|x\| \leqslant \liminf_{n\to+\infty} \|x_n\|$.

2°) Use the fact that the orthonormal sequence (e_n) necessarily converges weakly to 0 (see Tapa 144).

Comment. The result of Question 2°) is a generalization of what we saw in the previous Tapa. Strangely enough, any point in the ball can be obtained as a weak limit of a sequence "living" on the boundary of this ball.

---

146. ★★ *A further characterization of the projection onto a closed convex set*

Let $(H, \langle \cdot, \cdot \rangle)$ be a HILBERT space and let C be a closed convex set in H. The usual way of characterizing the orthogonal projection $p_C(x)$ of x onto C is as follows:

$$(\overline{x} = p_C(x)) \Leftrightarrow (\overline{x} \in C \text{ and } \langle x - \overline{x}, c - \overline{x} \rangle \leqslant 0 \text{ for all } c \in C)\,. \qquad (1)$$

Prove that a further characterization of $p_C(x)$ is as follows:

$$(\overline{x} = p_C(x)) \Leftrightarrow (\overline{x} \in C \text{ and } \langle c - \overline{x}, x - c \rangle \leqslant 0 \text{ for all } c \in C)\,. \qquad (2)$$

Hint. Show the equivalence between the assertions in the right-hand sides of (1) and (2).

Comment. It is of interest to visualize (in the plane) the meanings of the right-hand sides of (1) and (2) in terms of angles between vectors.

---

147. ★★ *The projection mapping onto a closed convex set is not weakly (sequentially) continuous*

Let $(H, \langle \cdot, \cdot \rangle)$ be an infinite-dimensional HILBERT space ($H = \ell^2(\mathbb{R})$, for example). The norm used in H is the one derived from the scalar product, *i.e.* $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$. When C is a closed convex subset in H, the projection mapping p_C enjoys a strong continuity property since

$$\|p_C(x) - p_C(y)\| \leqslant \|x - y\| \text{ for all } x, y \text{ in } H.$$

However, when the weak topology on H is considered, p_C loses the corresponding continuity property:

$$(u_n \rightharpoonup u) \text{ does not imply } (p_C(u_n) \rightharpoonup p_C(u))\,.$$

Here is a fairly simple counterexample. Let $(e_n)_{n \in \mathbb{N}}$ be an orthonormal sequence in H, let C be the closed unit ball in H. Consider the sequence (u_n) in H defined as:

$$u_n = e_1 + e_n.$$

1°) What is $p_C(u_n)$?

2°) Verify that the sequence (u_n) converges weakly to e_1 while the sequence $(p_C(u_n))$ does not converge weakly to $p_C(e_1) = e_1$.

Answers. 1°) We have: $p_C(u_n) = \frac{e_1+e_n}{\sqrt{2}}$.

2°) Since (e_n) converges weakly to 0 (see Tapa 144, for example), (u_n) converges weakly to e_1 and $(p_C(u_n))$ converges weakly to $e_1/\sqrt{2}$.

Comment. The weakly (sequential) continuity property of p_C, however, holds true when C is a closed vector subspace in H; in that case, p_C is a continuous linear operator. See Property 7 in Tapa 142.

---

148. ★★ *A primitive (or anti-gradient) function for the square of the distance function to a convex set*

Let $(H, \langle\cdot,\cdot\rangle)$ be a HILBERT space and let C be a closed convex set in H. We look for the primitive (or anti-gradient) functions of the orthogonal projection mapping, that is to say: differentiable functions $f : H \to \mathbb{R}$ such that

$$\nabla f(x) = p_C(x) \text{ for all } x \in H. \tag{1}$$

We consider the function $g = d_C^2$, *i.e.* the square of the distance function d_C to C.

1°) Show that g is differentiable on H with

$$\nabla g(x) = 2\,(x - p_C(x)) \text{ for all } x \in H. \tag{2}$$

2°) Deduce from the above all the differentiable functions $f : H \to \mathbb{R}$ satisfying (1).

Hint. 1°) The key-property to prove that

$$g(x+h) = g(x) + 2\,\langle x - p_C(x), h\rangle + \|h\|^2\,\varepsilon(h)$$

is the LIPSCHITZ property of the projection mapping:

$$\|p_C(u) - p_C(v)\| \leqslant \|u - v\| \text{ for all } u, v \text{ in } H.$$

Answer. 2°) The functions

$$f : x \mapsto f(x) = \frac{1}{2}\left[\|x\|^2 - d_C^2(x)\right] + r,$$

where r is a real number, are the primitive (or anti-gradient) functions of the projection mapping p_C.

---

149. ★★ *Projections onto moving convex sets*

Let $(H, \langle\cdot,\cdot\rangle)$ be a HILBERT space.

1°) Let C_1 and C_2 be two closed convex sets in H such that $C_1 \subset C_2$. Show that

$$\|p_{C_1}(x) - p_{C_2}(x)\|^2 \leqslant 2\left[d^2_{C_1}(x) - d^2_{C_2}(x)\right] \text{ for all } x \in H. \qquad (1)$$

2°) Let (C_n) be an increasing sequence (for the inclusion relation) of nonempty closed convex sets C_n in H, and let $C = \overline{\cup_n C_n}$.

(a) Check that C is convex.

(b) Prove that, for all $x \in H$,

$$\lim_{n\to+\infty} d_{C_n}(x) = d_C(x) \text{ and } \lim_{n\to+\infty} p_{C_n}(x) = p_C(x). \qquad (2)$$

3°) Let (C_n) be an decreasing sequence (for the inclusion relation) of nonempty closed convex sets C_n in H, and let $C = \overline{\cap_n C_n}$.

Assuming that C is nonempty, prove the results in (2) also hold true for all $x \in H$.

Hint. 1°) Start with the parallelogram identity applied to the vectors $x - p_{C_1}(x)$ and $x - p_{C_2}(x)$:

$$\begin{aligned} 2\left[d^2_{C_1}(x) + d^2_{C_2}(x)\right] &= 2\|p_{C_1}(x) - x\|^2 + 2\|p_{C_2}(x) - x\|^2 \\ &= \|p_{C_1}(x) - p_{C_2}(x)\|^2 + 4\left\|x - \frac{p_{C_1}(x) + p_{C_2}(x)}{2}\right\|^2. \end{aligned}$$

---

150. ★★ *Characterization of the projections onto an arbitrary closed set*

Let $(H, \langle\cdot,\cdot\rangle)$ be a HILBERT space and let S be a closed set in H. For $x \in H$, we denote by $P_S(x)$ the set of points (if any!) in S which are closest to x, *i.e.* those $\overline{x} \in S$ for which $\|x - \overline{x}\| = d_S(x)$. Surprisingly enough, these $\overline{x}$ can be *characterized* in various ways.

1°) Establish the equivalence between the three next statements:

(a) $\overline{x} \in P_S(x)$;
(b) $\overline{x} \in S$ and $\langle x - \overline{x}, c - \overline{x}\rangle \leqslant \frac{1}{2}\|c - \overline{x}\|^2$ for all $c \in S$;
(c) $\overline{x} \in P_S\left[\overline{x} + t(x - \overline{x})\right]$ for all $t \in [0,1]$.

2°) Check that if $\overline{x} \in P_S(x)$, then for all $t \in [0,1)$, $P_S\left[\overline{x} + t(x - \overline{x})\right]$ reduces to the singleton $\{\overline{x}\}$.

3°) Comment on the differences with the characterization of $\overline{x} = p_S(x)$ when S is convex.

Hint. 1°) – 2°). Drawing pictures in the plane aids our reasoning and helps us to understand the meaning of the characterization (b).

Answer. 3°) When S is convex, $P_S(x)$ reduces to the singleton $\{p_S(x)\}$ and the inequality of the characterization in (b) can be strengthened to:

$$\langle x - \overline{x}, c - \overline{x} \rangle \leqslant 0 \text{ for all } c \in S.$$

---

151. ★★ *Farthest points in an arbitrary bounded closed set*

Let $(H, \langle \cdot, \cdot \rangle)$ be a HILBERT space and let S be a bounded closed set in H. For $x \in H$, we define:

$$\Delta_S(x) = \sup_{s \in S} \|x - s\|, \tag{1}$$

$$Q_S(x) = \{s \in S : \|x - s\| = \Delta_S(x)\}. \tag{2}$$

In other words, $Q_S(x)$ is the set of points (if any!) in S which are farthest from x. We intend to prove some properties of the function Δ_S and of the set-valued mapping Q_S.

1°) (a) Show that $\Delta_S : H \to \mathbb{R}$ is convex and 1-LIPSCHITZ on H.

(b) - Check that Δ_S^2 is strongly convex on H, in the following sense: For all x_1, x_2 in H and all $\lambda \in [0,1]$,

$$\Delta_S^2\left[\lambda x_1 + (1-\lambda)x_2\right] \leqslant \lambda \Delta_S^2(x_1) + (1-\lambda)\Delta_S^2(x_2) - \lambda(1-\lambda)\left\|x_1 - x_2\right\|^2. \tag{3}$$

Put in an equivalent way, this means that $\Delta_S^2 - \|\cdot\|^2$ is still a convex function.

- Verify that Δ_S^2 is 1-coercive on H, that is to say:

$$\Delta_S^2(x)/\|x\| \to +\infty \text{ when } \|x\| \to +\infty. \tag{4}$$

2°) Variational characterization of farthest points.

Let $x \in H$ and $\overline{x} \in S$. Prove that $\overline{x} \in Q_S(x)$ if and only if:

$$\langle x - \overline{x}, s - \overline{x} \rangle \geqslant \frac{1}{2} \|\overline{x} - s\|^2 \text{ for all } s \in S. \tag{5}$$

Hints. Use the definitions of Δ_S and Q_S and apply basic calculus rules in the HILBERT space $(H, \langle \cdot, \cdot \rangle)$.

Comments. 1°) The convexity of Δ_S is obtained here for free, *i.e.* without assuming any additional property on S; this is a noticeable difference with the distance function d_S (which is convex only when the closed set S is convex).

2°) - It is a bit unexpected that one gets from (5) a *characterization* of farthest points in S from x, similar in spirit to that of projections onto S of x (see 1°) in Tapa 150).

- A consequence of the characterization of elements in $Q_S(x)$ is also that the set-valued mapping Q_S is *dissipative, i.e.*: For all x_1, x_2 in H and all $\overline{x}_1 \in Q_S(x_1)$, $\overline{x}_2 \in Q_S(x_2)$

$$\langle \overline{x}_1 - \overline{x}_2, x_1 - x_2 \rangle \leqslant 0. \tag{6}$$

- For the so-called "farthest point conjecture", see the comments following Tapa 119.

---

152. ★★ *Minimizing a quadratic function over* $\mathbb{R}^n$

Let $A \in \mathcal{S}_n(\mathbb{R})$, $b \in \mathbb{R}^n$. The question we address is that of minimizing the quadratic function

$$q : x \in \mathbb{R}^n \mapsto q(x) = \frac{1}{2} x^T A x - b^T x$$

over $\mathbb{R}^n$.

Prove the equivalence of the following three statements:

(*i*) The function is bounded from below on $\mathbb{R}^n$.

(*ii*) A is positive semidefinite and $b \in \operatorname{Im} A$.

(*iii*) The problem of minimizing q over $\mathbb{R}^n$ has a solution.

Hint. Warm-up: Begin with the easier case where A is invertible.

Comments. - When one of the three equivalent statements holds, it is possible to characterize the set of minimizers of q as solutions of the linear system $Ax = b$, or directly in terms of the pseudo-inverse of A.

- This Tapa is an example of a harmonious interplay between Matrix analysis and Optimization.

---

153. ★★ *An approximate* Rolle*'s theorem*

Let $f : \overline{B(0,r)} \subset \mathbb{R}^n \to \mathbb{R}$ be continuous on the closed ball $\overline{B(0,r)}$ and differentiable on the open ball $B(0,r)$. We suppose that there exists an $\varepsilon > 0$ such that:

$$|f(x)| \leqslant \varepsilon \text{ for all } x \text{ satisfying } \|x\| = r. \tag{1}$$

1°) Prove the following "approximate Rolle's theorem":

$$\text{There exists an } \overline{x} \in B(0,r) \text{ such that } \|\nabla f(\overline{x})\| \leqslant \frac{\varepsilon}{r}. \tag{2}$$

2°) Show, with the help of a counterexample, that the above result is sharp, *i.e.* it is not possible to do better than ε/r for the upper bound of $\|\nabla f(\overline{x})\|$ in (2).

Hints. 1°) Introduce the auxiliary function $g : \overline{B(0,r)} \to \mathbb{R}$ defined by:

$$g(x) = \frac{\|x\|^2}{r^2} - \left[\frac{f(x)}{\varepsilon}\right]^2.$$

Let $\mu = \min_{x \in \overline{B(0,r)}} g(x)$. Because $g(0) \leqslant 0$, we have that $\mu \leqslant 0$. Two different cases then have to be discussed: $\mu = 0$ and $\mu < 0$.

Answers. 1°) If $\mu = 0$, following the definitions of g, μ, and the calculation of $\nabla f(0)$, one easily obtains: $\|\nabla f(0)\| \leqslant \frac{\varepsilon}{r}$.

If $\mu < 0$, due to the assumption (1), the minimizers $\overline{x}$ of g on $\overline{B(0,r)}$ necessarily lie in $B(0,r)$. By writing the optimality condition $\nabla g(\overline{x}) = 0$, one checks that all these minimizers satisfy (2).

2°) Let $f(x) = a^T x$, with $a \in \mathbb{R}^n$, $\|a\| = \varepsilon/r$. It follows from the Cauchy–Schwarz inequality that (1) is indeed satisfied. But $\|\nabla f(\overline{x})\| = \|a\| = \varepsilon/r$ for all $\overline{x}$.

Comments. - If $\mathbb{R}^n$ is replaced with some general HILBERT space, it is still possible to have an approximate ROLLE inequality similar to (2), but with an upper bound $2\varepsilon/r$ instead of ε/r.

- In contrast to what happens in the finite-dimensional context, the "exact" ROLLE's theorem does not hold true in a HILBERT space context. Let $f : H \to \mathbb{R}$ be a $\mathcal{C}^\infty$ function on an infinite-dimensional Hilbert space H such that $f(x) = 0$ for all x in the unit-sphere of H; there does not necessarily exist a point $\overline{x}$ satisfying $\|\overline{x}\| < 1$ for which $\nabla f(\overline{x}) = 0$. An example with $H = L^2([0,1])$ is presented in detail in

D. AZÉ and J.-B. HIRIART-URRUTY, *Sur un air de* ROLLE *and* ROLLE. Revue de la Filière Mathématique (ex-Revue de Mathématiques Spéciales), n°3-4 (2000), 455–460.

---

154. ★★ *A kind of* ROLLE's *theorem for vector-valued functions*

1°) As an appetizer.

Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by:

$$f(x,y) = \begin{pmatrix} x(x^2+y^2-1) \\ y(x^2+y^2-1) \end{pmatrix}.$$

(*a*) Determine the Jacobian matrix $Jf(x,y)$ of f at (x,y) and check that there is no point in $\mathbb{R}^2$ where Jf vanishes.

(*b*) Let $\Omega = B(0,1)$ be the open unit ball in $\mathbb{R}^2$ for the usual Euclidean norm. What are the values taken by f on the boundary of Ω? What can you conclude about a possible ROLLE's theorem on $\overline{B(0,1)}$?

2°) (*a*) Let Ω be an open bounded connected subset in $\mathbb{R}^n$, let $(H, \langle\cdot,\cdot\rangle)$ be a HILBERT space and let $f : \mathbb{R}^n \to H$ be a differentiable mapping. We make the following assumption:

$$\left\{ \begin{array}{l} \text{There is a } d \in H \text{ such that the "scalarized" function} \\ x \mapsto \langle f(x), d\rangle \text{ is constant on the boundary of } \Omega. \end{array} \right. \qquad (\mathcal{H})$$

Prove that there exists a point c in H such that:

$$\langle Df(c)h, d\rangle = 0 \text{ for all } h \in \mathbb{R}^n. \qquad (\mathcal{C})$$

(Here, $Df(c) \in \mathcal{L}(\mathbb{R}^n, H)$ denotes the differential mapping of f at the point c).

(*b*) Back to the Question 1°): Given a unit vector d in $\mathbb{R}^2$, find a point $c \in \Omega$ such that $(\mathcal{C})$ holds true.

Hints. 2°) (a). Use the real-valued function $g : x \in \mathbb{R}^n \mapsto g(x) = \langle f(x), d\rangle$. The image $g(\overline{\Omega})$ is a compact interval $[m, M]$. The function g is constant on the boundary of Ω.

If $\overline{x} \in \Omega$ is a minimizer or a maximizer of g on $\overline{\Omega}$, then $Dg(\overline{x}) : h \in \mathbb{R}^n \mapsto \langle Df(\overline{x})h, d\rangle$ is null.

Answers. 1°) To impose $Jf(x, y) = 0$ amounts to having simultaneously:

$$xy = 0,\ 3x^2 + y^2 - 1 = 0,\ 3y^2 + x^2 - 1 = 0.$$

This is just impossible.

Besides that, $f(x, y) = 0$ for all (x, y) on the boundary of Ω. So, the immediate generalization of ROLLE's theorem to vector-valued functions ("f constant on the boundary of Ω implies that there is a point in Ω where the differential vanishes") is not valid.

2°) (b). For a given d, $(\mathcal{C})$ holds true for the point $c = d/\sqrt{3}$. But, according to 1°), there is no point c which would work for all d.

---

155. ★★ *The Jacobian matrix and determinant of two basic polynomial transformations*

1°) Elementary symmetric polynomials.

Given n real numbers $x_1, ..., x_n$, one denotes by $\sigma_1, ..., \sigma_n$ the associated (so-called) elementary symmetric polynomials, that is to say:

$$\text{For } k = 1, ..., n,\ \ \sigma_k = \sum_{1 \leqslant i_1 < ... < i_k \leqslant n} x_{i_1} x_{i_2} ... \ x_{i_k}. \tag{1}$$

For example, the two "extreme" cases are:

$$\sigma_1 = \sum_{i=1}^{n} x_i \text{ and } \sigma_n = x_1 x_2 ... x_n.$$

Consider the mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$F(x_1, ..., x_n) = (\sigma_1, ..., \sigma_n).$$

Determine the Jacobian matrix $JF(x_1, ..., x_n)$ $(\in \mathcal{M}_n(\mathbb{R}))$ and the Jacobian determinant $\det JF(x_1, ..., x_n)$ $(\in \mathbb{R})$ of F at $(x_1, ..., x_n)$.

2°) Coefficients of the inverse matrix.

Let $\Omega = \left\{(a,b,c,d) \in \mathbb{R}^4 : ad - bc \neq 0\right\}$ and let $F : \Omega \to \mathbb{R}^4$ be defined as

$$F(a,b,c,d) = \frac{1}{ad-bc}(d,-b,-c,a). \tag{2}$$

If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathcal{M}_2(\mathbb{R}) \equiv \mathbb{R}^4$ is invertible, one recognizes in (2) the entries of the inverse matrix $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \in \mathcal{M}_2(\mathbb{R}) \equiv \mathbb{R}^4$.

Determine the Jacobian matrix $JF(a,b,c,d)$ $(\in \mathcal{M}_4(\mathbb{R}))$ and the Jacobian determinant $\det JF(a,b,c,d)$ $(\in \mathbb{R})$ of F at (a,b,c,d).

Hints. 1°) - As a warm-up, begin with the cases $n = 2$ and $n = 3$.

- We have, for all $(x_1,...,x_n)$ and all $(X_1,...,X_n)$ in $\mathbb{R}^n$,

$$\begin{gathered} X^n - \sigma_1 X^{n-1} + ... + (-1)^k \sigma_k X^{n-k} + \ (-1)^n \sigma_n = \\ (X - x_1)...\ (X - x_k)...\ (X - x_n). \end{gathered} \tag{3}$$

Since the partial derivatives $\partial\sigma_k/\partial x_i$ have to be computed, begin by differentiating the identity (3) above with respect to x_i.

- By intuition, you may think that the result involves the VANDERMONDE determinant... you are right!

2°) Indeed, there will be a simple relationship between $\det JF(a,b,c,d)$ and

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

Answers. 1°) We obtain:

$$\det JF(x_1,...,x_n) = \prod_{1 \leqslant i < j \leqslant n} (x_i - x_j). \tag{4}$$

2°) Let $\Delta = ad - bc$. For $\Delta \neq 0$, we have

$$JF(a,b,c,d) = \frac{1}{\Delta^2}\begin{bmatrix} -d^2 & cd & bd & -bc \\ bd & -ad & -b^2 & ab \\ cd & -c^2 & -ad & ac \\ -bc & ac & ab & -a^2 \end{bmatrix}.$$

Trick: $-bc = -ad + \Delta$, $-ad = -bc - \Delta$, $ad = bc + \Delta$, $bc = ad - \Delta$.

Thus, denoting by u the column vector $\begin{pmatrix} d \\ -b \\ -c \\ a \end{pmatrix}$, we have

$$\begin{bmatrix} -d^2 & cd & bd & bc \\ bd & -ad & -b^2 & ab \\ cd & -c^2 & ad & ac \\ -bc & ac & ab & -a^2 \end{bmatrix} =$$

$$\begin{bmatrix} & 0 & & 0 & & 0 & & \Delta \\ -d\ u + & 0 & , c\ u + & -\Delta & , b\ u + & 0 & , -a\ u + & 0 \\ & 0 & & 0 & & -\Delta & & 0 \\ & \Delta & & 0 & & 0 & & 0 \end{bmatrix}.$$

The multilinearity of the determinant (as a function of column vectors) gives the rest. Finally,

$$\det JF(a,b,c,d) = \frac{1}{\Delta^4} = \frac{1}{(ad-bc)^4}. \tag{5}$$

This is confirmed by a computation with MAPLE.

Comment. As expected, the result in Question 2°) can be generalized to the case where $n > 2$. For invertible $A = [a_{i,j}] \in \mathcal{M}_n(\mathbb{R}) \equiv \mathbb{R}^{n^2}$, let $F(A)$ be defined by

$$F(a_{1,1}, ..., a_{n,n}) = (b_{1,1}, ..., b_{n,n}),$$

where the $b_{i,j}$'s are the entries of A^{-1} ; by doing so, one defines a mapping F from an open subset of $\mathbb{R}^{n^2}$ into $\mathbb{R}^{n^2}$. Then, it turns out that:

$$\det JF(a_{1,1}, ..., a_{n,n}) = \frac{(-1)^n}{(\det A)^{2n}}. \tag{6}$$

---

156. ★★ *Differentiating* $t \mapsto \exp[(1-t)A]\exp(tB)$.

Let A and B be two matrices in $\mathcal{M}_n(\mathbb{R})$. We define $\theta : \mathbb{R} \to \mathcal{M}_n(\mathbb{R})$ as follows:

$$t \in \mathbb{R} \mapsto \theta(t) = \exp[(1-t)A]\exp(tB).$$

(Here, $\exp M$ denotes the exponential of the matrix M; also denoted by e^M.)

1°) (*a*) Show that θ is continuously differentiable on $\mathbb{R}$, and determine $\theta'(t), t \in \mathbb{R}$.

(*b*) Deduce that

$$e^A - e^B = \int_0^1 e^{(1-t)A}(A-B)e^{tB}\,\mathrm{d}t. \tag{1}$$

How does this formula simplify when A and B commute?

2°) Using the result above, check that the exponential mapping

$$M \in \mathcal{M}_n(\mathbb{R}) \mapsto \exp M \in \mathcal{M}_n(\mathbb{R})$$

is locally LIPSCHITZ.

Hints. (*a*) It helps to note that the matrices M and $\exp M$ commute.

2°) The mapping $(s, M) \in \mathbb{R} \times \mathcal{M}_n(\mathbb{R}) \mapsto \exp(sM)$ is continuous, hence transforms bounded sets into bounded sets.

Answers. 1°) We have that:

$$\theta'(t) = e^{(1-t)A}(B-A)e^{tB}.$$

Consequently, (1) just comes from writing

$$\theta(1) - \theta(0) = \int_0^1 \theta'(t)\,\mathrm{d}t.$$

When A and B commute, (1) is simplified into

$$e^A - e^B = e^A \int_0^1 (A-B)e^{t(B-A)}\,\mathrm{d}t. \tag{2}$$

2°) Let $\|\cdot\|$ be a norm on $\mathcal{M}_n(\mathbb{R})$; due to the continuity of the symmetric bilinear form $(U, V) \mapsto UV$, there exists an $\alpha > 0$ such that $\|UV\| \leqslant \alpha \|U\| \times \|V\|$ for all U, V in $\mathcal{M}_n(\mathbb{R})$. If $\mathcal{B}$ is a bounded subset in $\mathcal{M}_n(\mathbb{R})$, then there exists a $\beta > 0$ such that

$$\left\|e^{sM}\right\| \leqslant \beta \text{ whenever } s \in [0,1] \text{ and } M \in \mathcal{B}.$$

Therefore, we deduce from (1): with $L = \alpha^2\beta^2$ for example,

$$\left\|e^A - e^B\right\| \leqslant L\,\|A-B\| \text{ for all } A, B \text{ in } \mathcal{B}.$$

Comment. With the results displayed above, one easily obtains the following "first-order sensitivity" result on the exponential of a matrix: The function $x \in \mathbb{R} \mapsto \sigma(x) = \exp(A + xH)$ is differentiable at 0 with

$$\sigma'(0) = \int_0^1 e^{(1-t)A} H e^{tA} \, \mathrm{d}t. \tag{3}$$

---

157. ★★ *Squeezing a function with* LIPSCHITZ *gradients between two quadratic functions*

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function whose gradient mapping ∇f satisfies a LIPSCHITZ property on $\mathbb{R}^n$: there exists an $L > 0$ such that

$$\|\nabla f(x) - \nabla f(x')\| \leqslant L \, \|x - x'\| \text{ for all } x, x' \text{ in } \mathbb{R}^n,$$

where $\|\cdot\|$ stands for the usual Euclidean norm.

Show that, for all $x \in \mathbb{R}^n$:

$$f(x+d) \leqslant f(x) + \nabla f(x)^T d + \frac{L}{2} \|d\|^2 \text{ for all } d \in \mathbb{R}^n;$$

$$f(x) + \nabla f(x)^T d - \frac{L}{2} \|d\|^2 \leqslant f(x+d) \text{ for all } d \in \mathbb{R}^n.$$

Hint. Use the elementary first-order TAYLOR expansion in an integral form. Write

$$f(x+d) = f(x) + \int_0^1 \nabla f(x+td)^T d \, \mathrm{d}t; \; \nabla f(x)^T d = \int_0^1 \nabla f(x)^T d \, \mathrm{d}t.$$

Comments. - If f is assumed twice continuously differentiable on $\mathbb{R}^n$, the required LIPSCHITZ property on ∇f would certainly be secured if

$$-L \, I_n \preccurlyeq \nabla^2 f(u) \preccurlyeq L \, I_n \text{ for all } u \in \mathbb{R}^n,$$

or, equivalently, if all the eigenvalues of $\nabla^2 f(u)$ lie in the interval $[-L, L]$ for all $u \in \mathbb{R}^n$.

- Squeezing the increment $f(x_k + d) - f(x_k)$ around an iterate x_k, with the help of quadratic functions built up from $\nabla f(x_k)^T d \pm \frac{L}{2} \|d\|^2$, turns out to be helpful in designing numerical optimization procedures for minimizing f. See Tapa 219 for an example.

---

158. ★★ *Characterizing convex functions with* LIPSCHITZ *gradients*

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function. We know that

$$f(x') - f(x) - \nabla f(x)^T (x' - x) \geqslant 0 \text{ for all } x, x' \text{ in } \mathbb{R}^n; \quad (1)$$
$$\left[\nabla f(x') - \nabla f(x)\right]^T (x' - x) \geqslant 0 \text{ for all } x, x' \text{ in } \mathbb{R}^n. \quad (2)$$

We intend to make these inequalities stronger if we assume that the gradient mapping ∇f satisfies a LIPSCHITZ property on $\mathbb{R}^n$, *i.e.,* when there exists an $L > 0$ such that

$$\|\nabla f(x) - \nabla f(x')\| \leqslant L \|x - x'\| \text{ for all } x, x' \text{ in } \mathbb{R}^n, \quad (3)$$

where $\|\cdot\|$ stands for the usual Euclidean norm.

Prove that the following three statements are equivalent:

(*i*) $\|\nabla f(x) - \nabla f(x')\| \leqslant L \|x - x'\|$ for all x, x' in $\mathbb{R}^n$;

(*ii*) $f(x') - f(x) - \nabla f(x)^T (x' - x) \geqslant \frac{1}{2L} \|\nabla f(x) - \nabla f(x')\|^2$ for all x, x' in $\mathbb{R}^n$;

(*iii*) $\left[\nabla f(x') - \nabla f(x)\right]^T (x' - x) \geqslant \frac{1}{L} \|\nabla f(x) - \nabla f(x')\|^2$ for all x, x' in $\mathbb{R}^n$.

Hints. The implications $[(ii) \Rightarrow (iii)]$ and $[(iii) \Rightarrow (i)]$ are easy to prove. For the proof of $[(i) \Rightarrow (ii)]$, it is advisable to consider the function

$$f_x : y \mapsto f_x(y) = f(y) - f(x) - \left[\nabla f(x)\right]^T (y - x).$$

This new function is convex and its gradient is LIPSCHITZ on $\mathbb{R}^n$ with L again as a LIPSCHITZ constant. It is moreover minimized on $\mathbb{R}^n$ at $\overline{y} = x$.

---

159. ★★ *Determining and characterizing critical points of a function*

Let H be a HILBERT space, in which $\langle \cdot, \cdot \rangle$ denotes the scalar product and $\|\cdot\|$ the associated norm. For a given $a \neq 0$ in H, consider the function $f : H \to \mathbb{R}$ defined as follows:

$$f(x) = \langle a, x \rangle \times e^{-\|x\|^2}.$$

1°) (*a*) Calculate the gradient $\nabla f(x)$ of f at x.

(*b*) Determine the critical (or stationary) points of f, that is to say, the points $x \in H$ where the gradient of f equals the zero vector.

2°) (*a*) Calculate the second-order differential $D^2 f(x) : H \times H \to \mathbb{R}$ of f at x.

(*b*) Determine the nature of the critical points of f (local minimizer, local maximizer, saddle-point).

Answers. 1°) (*a*). We have:

$$\nabla f(x) = [a - 2\,\langle a, x\rangle\, x]\, e^{-\|x\|^2}. \tag{1}$$

(*b*) There are exactly two critical points:

$$u = \frac{\sqrt{2}}{2}\frac{a}{\|a\|} \text{ and } v = -\frac{\sqrt{2}}{2}\frac{a}{\|a\|}. \tag{2}$$

2°) (*a*). We have:

$$D^2 f(x)(h,k) = \left[\begin{array}{c} 4\,\langle a, x\rangle \times \langle x, h\rangle \times \langle x, k\rangle - 2\,\langle a, k\rangle \times \langle x, h\rangle \\ -2\,\langle a, h\rangle \times \langle x, k\rangle - 2\,\langle a, x\rangle \times \langle h, k\rangle \end{array}\right] \times e^{-\|x\|^2}.$$

(*b*) Consider the critical point u (described in (2)). We have

$$D^2 f(u)(h,h) = -\frac{\sqrt{2} \times e^{-1/2}}{\|a\|}\left[\langle a, h\rangle^2 + \|a\|^2 \times \|h\|^2\right],$$

so that

$$D^2 f(u)(h,h) \leqslant -\alpha\,\|h\|^2$$

for some $\alpha > 0$. Therefore, u is a strict local maximizer of f.

Since $f(-x) = -f(x)$ for all $x \in H$, the point $v = -u$ is a strict local minimizer of f.

Comment. Actually, u (resp. v) is a global maximizer (resp. a global minimizer) of f on H.

---

160. ★★ *Convexity properties and differentials of the* ln det *function*

Let $\Omega \subset \mathcal{S}_n(\mathbb{R})$ denote the open convex cone of positive definite $n \times n$ matrices. We consider the function $f : \Omega \to \mathbb{R}$ defined as follows:

$$A \in \Omega \mapsto f(A) = \ln\det(A^{-1}) = -\ln\det(A).$$

1°) Show that f is strictly convex.

2°) Determine, as "economically" as possible, the first differential $Df(A)$ and the second differential $D^2f(A)$ of f at $A \in \Omega$.

3°) (a) Determine $D^3f(A)(H,H,H)$ for the third differential $D^3f(A)$ of f at $A \in \Omega$.

(b) Check that there exists a constant $\alpha > 0$ such that

$$\left|D^3f(A)(H,H,H)\right| \leqslant \alpha \left[D^2f(A)(H,H)\right]^{3/2} \text{ for all } H. \qquad (SC)$$

Answers. 2°) We have, for the first differential $Df(A)$ and the second differential $D^2f(A)$ of f at A:

$$\begin{aligned} Df(A) &: H \in \mathcal{S}_n(\mathbb{R}) \mapsto -\operatorname{tr}(A^{-1}H) = -\operatorname{tr}(A^{-1/2}HA^{-1/2}); \quad (1) \\ D^2f(A) &: (H,K) \in \mathcal{S}_n(\mathbb{R}) \times \mathcal{S}_n(\mathbb{R}) \mapsto \operatorname{tr}(A^{-1}HA^{-1}K). \quad (2) \end{aligned}$$

As a particular case of (2), we note that

$$D^2f(A)(H,H) = \operatorname{tr}\left[(A^{-1/2}HA^{-1/2})^2\right]. \qquad (3)$$

As a result, the second-order Taylor–Young development of f at $A \in \Omega$ is: for $H \in \mathcal{S}_n(\mathbb{R})$,

$$\begin{aligned} \ln\det(A+H) &= \ln\det(A) + \operatorname{tr}(A^{-1}H) \\ &\quad - \operatorname{tr}(A^{-1}HA^{-1}H) + \|H\|^2 \varepsilon(H), \qquad (4) \end{aligned}$$

with $\varepsilon(H) \to 0$ when $H \to 0$.

3°) We have: for all $A \in \Omega$ and $H \in \mathcal{S}_n(\mathbb{R})$,

$$D^3f(A)(H,H,H) = -2\operatorname{tr}\left[(A^{-1/2}HA^{-1/2})^3\right].$$

Consequently,

$$\left|D^3f(A)(H,H,H)\right| \leqslant 2\left[D^2f(A)(H,H)\right]^{3/2}.$$

Comments. - The result of the first question could be translated into the following: if A and B are two different positive definite matrices and $\lambda \in (0,1)$,

$$\det\left[\lambda A + (1-\lambda)B\right] > (\det A)^{\lambda}\,(\det B)^{1-\lambda}.$$

A consequence of this inequality is, for example, the convexity of the upperlevel-set $\{A \in \Omega : \det A \geqslant 1\}$, which is not obvious at all...

- If $\mathcal{S}_n(\mathbb{R})$ is equipped with the inner product $(U, V) \mapsto \langle\langle U, V\rangle\rangle = \mathrm{tr}(UV)$, one can read (1) again by saying that the gradient (matrix) of the function f at $A \in \Omega$ is $\nabla f(A) = -A^{-1}$. Due to the convexity of the function f, its gradient mapping is "increasing" (one also says "monotone") in the following sense:

$$-\left\langle\left\langle B^{-1} - A^{-1}, B - A\right\rangle\right\rangle = \mathrm{tr}(A^{-1}B) + \mathrm{tr}(B^{-1}A) - 2n \geqslant 0.$$

- The function f plays the role of a "barrier" for Ω : $f(A) \to +\infty$ when A approaches the boundary of Ω. The inequality (SC), called "self-concordance", is a technical property which turns out to be very important in Convex optimization.

---

161. ★★ *Inverting a negative definite matrix via convex optimization*

Let $\mathcal{S}_n(\mathbb{R})$ be equipped with the standard inner product $\langle\langle \cdot, \cdot \rangle\rangle$ (recall that $\langle\langle U, V\rangle\rangle = \mathrm{tr}(UV)$). Let $\Omega \subset \mathcal{S}_n(\mathbb{R})$ denote the open convex cone of positive definite $n \times n$ matrices. Given a (symmetric) negative definite $n \times n$ matrix, we consider

$$\begin{array}{rcl} f_B : \Omega & \to & \mathbb{R} \\ A & \mapsto & f_B(A) = \langle\langle A, B\rangle\rangle + \ln\det(A). \end{array}$$

According to the previous Tapa, f_B is a differentiable strictly concave function.

Calculate $\sup_{A\in\Omega} f_B(A)$ and determine the unique A^* where this supremum is attained.

Answer. The gradient (matrix) of the function f_B at $A \in \Omega$ is $\nabla f_B(A) = A^{-1}$. So, the (strictly) concave function f_B is maximized at the (unique) matrix A^* satisfying:

$$B + (A^*)^{-1} = 0, \tag{1}$$

that is to say, $A^* = -B^{-1}$.

Consequently,

$$\sup_{A\in\Omega} f_B(A) = -\ln\det(-B) - n. \tag{2}$$

Comments. - The so-called LEGENDRE–FENCHEL transform of the (strictly convex) function $A \succ 0 \mapsto f(A) = -\ln\det(A) - \frac{n}{2}$ is, by definition,

$$B \prec 0 \mapsto f^*(B) = \sup_{A \succ 0} \left[\langle\langle A, B\rangle\rangle - f(A)\right].$$

According to the result proved in this Tapa,

$$f^*(B) = -\ln\det(-B) - \frac{n}{2}. \tag{3}$$

So, f and its LEGENDRE–FENCHEL transform f^* look very similar.

- See Tapa 13 and Tapa 73 for two further examples of LEGENDRE transforms.

---

162. ★★ *Higher-order* TAYLOR–YOUNG *developments of the determinant function*

Let $E = \mathcal{M}_n(\mathbb{R})$ be structured as a Euclidean space with the help of the inner product $\langle\langle U, V\rangle\rangle = \text{trace}\,(U^T V)$. The associated norm is denoted by $\|\cdot\|$. We consider the determinant function $f : E \to \mathbb{R}$, *i.e.,* the one that associates $f(M) = \det M$ with $M \in E$.

1°) (*a*) Show that, for all M and N in E :

$$f(M+N) = f(M) + \sum_{k=1}^{n} \frac{1}{k!} D^k f(M) \underbrace{(N, ..., N)}_{k \text{ times}}, \tag{1}$$

where $D^k f(M)$ denotes the $k-$th order differential of f at M.

(*b*) Deduce that, for A and B in E :

$$\frac{1}{(n-p)!} D^{n-p} f(B) \underbrace{(A, ..., A)}_{n-p \text{ times}} = \frac{1}{p!} D^p f(A) \underbrace{(B, ..., B)}_{p \text{ times}} \tag{2}$$

if $p \in \{1, ..., n-1\}$, and

$$f(A) = \frac{1}{n!} D^n f(B) \underbrace{(A, ..., A)}_{n \text{ times}}. \tag{3}$$

Deduce that $(\det A)/\left\|A\right\|^{n-1} \to 0$ as $A \to 0$.

2°) Applications. Throughout, cof$A$ denotes the matrix of cofactors of A.

(*a*) Let $n = 2$. Show that, for A and B in $\mathcal{M}_2(\mathbb{R})$,

$$\langle\langle \text{cof}A, B\rangle\rangle = \langle\langle \text{cof}B, A\rangle\rangle\,; \tag{4}$$

$$\det(A+B) = \det A + \langle\langle \mathrm{cof}A, B\rangle\rangle + \det B. \tag{5}$$

(b) Let $n = 3$. Show that, for A and B in $\mathcal{M}_3(\mathbb{R})$,

$$\det(A+B) = \det A + \langle\langle \mathrm{cof}A, B\rangle\rangle + \langle\langle \mathrm{cof}B, A\rangle\rangle + \det B. \tag{6}$$

(c) Determine $D^k f$ for all integers $k \geqslant 1$ in the cases where $n = 2$ or $n = 3$.

Hints. 1°) (a). The considered function f is (continuous) multilinear on $E \equiv \underbrace{\mathbb{R}^n \times ... \times \mathbb{R}^n}_{n \text{ times}}$. Therefore, $D^{n+1} f = 0$.

Moreover, $D^k f(A)$ is multilinear on $\underbrace{E \times ... \times E}_{k \text{ times}}$ and homogeneous of degree $n - k$.

Answers. 2°) (c). Case $n = 2$. We have:

$$\begin{aligned} Df(A) &: H \mapsto Df(A)(H) = \langle\langle \mathrm{cof}A, H\rangle\rangle\,; \\ D^2 f(A) &: (H,K) \mapsto D^2 f(A)(H,K) = \frac{1}{2}\left[\det(H+K) - \det(H-K)\right] \\ &= \det(H+K) - \det H - \det K. \end{aligned}$$

Case $n = 3$. We have:

$$\begin{aligned} Df(A)(H) &= \langle\langle \mathrm{cof}A, H\rangle\rangle\,; \\ D^2 f(A)(H,K) &= \frac{1}{2}\,\langle\langle \mathrm{cof}(H+K) - \mathrm{cof}(H-K), A\rangle\rangle \\ &= \langle\langle \mathrm{cof}(H+K) - \mathrm{cof}H - \mathrm{cof}K, A\rangle\rangle\,; \\ D^3 f(A)(H,K,L) &= \frac{1}{4}[\det(H+K+L) - \det(H+K-L) + \\ &\quad \det(H-K-L) - \det(H-K+L)]. \end{aligned}$$

Comment. - The determinant function is of the utmost importance in Matrix analysis. Note the elegance of the formulas (5) and (6) allowing the exact expansion of $\det(A+B)$.

- An example of application of formula (5) is: for A and B in $\mathcal{M}_2(\mathbb{R})$,

$$\det(A+B) + \det(A-B) = 2\det A + 2\det B.$$

Similarly, an example of application of formula (6) is: for A and B in $\mathcal{M}_3(\mathbb{R})$,

$$\det(A+B)+\det(A-B)=2\det A+2\left\langle\langle A,\mathrm{cof}B\rangle\right\rangle.$$

---

163. ★★ *Convexity of entropy-like functions*

Let $\varphi:(0,+\infty)\to\mathbb{R}$ be a convex function. We define $I_\varphi:(0,+\infty)^n\times(0,+\infty)^n\to\mathbb{R}$ as follows:

$$(p=(p_1,...,p_n),q=(q_1,...,q_n))\mapsto I_\varphi(p,q)=\sum_{i=1}^n q_i\times\varphi\left(\frac{p_i}{q_i}\right). \quad (1)$$

1°) (a) Show that I_φ is a convex function.

(b) Deduce the inequality below:

$$I_\varphi(p,q)\geqslant\left(\sum\nolimits_{i=1}^n q_i\right)\times\varphi\left(\frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i}\right). \quad (2)$$

2°) One defines $\varphi^\diamondsuit:(0,+\infty)\to\mathbb{R}$ by $\varphi^\diamondsuit(u)=u\times\varphi\left(\frac{1}{u}\right)$.

(a) Check that $\varphi^\diamondsuit$ is convex.

(b) How do I_φ and $I_{\varphi^\diamondsuit}$ compare?

3°) Derive the expressions of $\varphi^\diamondsuit$ and I_φ in the following cases:

$$\varphi(t)=t\ln t,\ (1-\sqrt{t})^2,\ t^\alpha \text{ with } \alpha>1,\ (t-1)^2,\ |t-1|.$$

Hints. 1°) (a). The key-point is to prove the convexity of the function

$$(x,y)\in(0,+\infty)\times(0,+\infty)\mapsto y\times\varphi\left(\frac{x}{y}\right).$$

Answers. 2°) (a). This is a consequence of previous results.

(b) We have $I_{\varphi^\diamondsuit}(p,q)=I_\varphi(q,p)$.

3°) If, for example, $\varphi(t)=(1-\sqrt{t})^2$, then $\varphi^\diamondsuit=\varphi$ and $I_\varphi(p,q)=\sum\limits_{i=1}^n\left(\sqrt{p_i}-\sqrt{q_i}\right)^2$.

Comment. A convexity result in a more general context is as follows. Let $f : \mathbb{R}^n \to \mathbb{R}$; the so-called *perspective function* $\widetilde{f} : \mathbb{R}^n \times (0, +\infty) \to \mathbb{R}$ associated with f is defined as:

$$\widetilde{f}(x, u) = u \times f\left(\frac{x}{u}\right).$$

It is not difficult to prove that $\widetilde{f}$ is convex on $\mathbb{R}^n \times (0, +\infty)$ if and only if f is convex on $\mathbb{R}^n$. The function $f^{\diamondsuit}$ considered in 2°) is nothing else than $\widetilde{f}(1, \cdot)$.

The function $\widetilde{f}$, which is clearly positively homogeneous ($\widetilde{f}(\alpha x, \alpha u) = \alpha \widetilde{f}(x, u)$ for all $\alpha > 0$), is a sort of "homogenized dilated" version of f.

---

164. ★★ *Maximizing the determinant over a unit ball of matrices*

Let $\mathcal{M}_n(\mathbb{R})$ be structured as a Euclidean space with the help of the scalar product $\langle\langle A, B \rangle\rangle = \mathrm{tr}(A^T B)$. Let $\mathcal{N}$ be an arbitrary norm on $\mathcal{M}_n(\mathbb{R})$, for which one sets:

$$\begin{aligned} \mathcal{S} &= \{M \in \mathcal{M}_n(\mathbb{R}) : \mathcal{N}(M) = 1\} \quad \text{(the unit sphere for } \mathcal{N}), \\ \mathcal{B} &= \{M \in \mathcal{M}_n(\mathbb{R}) : \mathcal{N}(M) \leqslant 1\} \quad \text{(the closed unit ball for } \mathcal{N}). \end{aligned}$$

1°) Show that the maximum of the determinant function $f : X \mapsto f(X) = \det X$ on $\mathcal{S}$ is also the maximum of f on $\mathcal{B}$.

Let $\mathcal{N}_* : A \in \mathcal{M}_n(\mathbb{R}) \mapsto \mathcal{N}_*(A) = \max_{M \in \mathcal{S}} \langle\langle A, M \rangle\rangle$. This new function $\mathcal{N}_*$ is also a norm on $\mathcal{M}_n(\mathbb{R})$, called the dual norm of $\mathcal{N}$.

2°) Let X be a matrix of $\mathcal{S}$ where f attains its maximal value on $\mathcal{S}$.

(a) Prove that X is invertible and

$$\mathcal{N}_*(X^{-T}) \geqslant n. \tag{1}$$

Here and below, X^{-T} denotes the transpose of X^{-1} (or, which amounts to the same, the inverse of X^T).

(b) By using the fact that f is maximized on $\mathcal{S}$ (or on $\mathcal{B}$) at X, show that

$$\left\langle\left\langle X^{-T}, M - X \right\rangle\right\rangle \leqslant 0 \text{ for all } M \in \mathcal{B}. \tag{2}$$

(c) Deduce finally that

$$\mathcal{N}_*(X^{-T}) = n.$$

Hints. 2°) (b). A necessary condition for X to maximize f on $\mathcal{B}$ is:

$$\langle\langle \nabla f(X), M - X \rangle\rangle \leqslant 0 \text{ for all } M \in \mathcal{B}.$$

For the considered (determinant function) f, we have

$$\nabla f(X) = (\det X\ )\ X^{-T}.$$

---

165. ★★ *Closedness of the sets of matrices of bounded rank*

Consider in $\mathcal{M}_{m,n}(\mathbb{R})$ the set S_r of matrices M such that $\text{rank}(M) \leqslant r$.

1°) Show that S_r is closed (or that its complementary set S_r^c in $\mathcal{M}_{m,n}(\mathbb{R})$ is open).

2°) What kind of continuity property can be expected for the rank function?

Hint. 1°) If M is of rank r, there exists a minor $\Delta_r(M)$ of order r extracted from M which is different from 0. If $M_k \to M$, then $\Delta_r(M_k) \to \Delta_r(M)$ by the continuity property of the determinant function. Thus, $\Delta_r(M_k) \neq 0$ for k large enough. In showing this, one proves that

$$S_r^c = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) : \text{rank}(M) > r\}$$

is an open set.

Answer. 2°) The rank function may take only integer values $0, 1, ..., p = \min(m, n)$; therefore, no continuity property is expected... However, the rank function is lower-semicontinuous, that is:

$$\liminf_{k \to +\infty} \text{rank}(M_k) \geqslant \text{rank}(M) \text{ whenever } M_k \to M.$$

This is the "functional" translation of the property that the sublevel-sets

$$S_r = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) : \text{rank}(M) \leqslant r\}, r \in \mathbb{R},$$

are all closed.

Comment. We can complete this by adding some topological properties of

$$\Sigma_k = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) : \text{rank}(M) = k\},$$

where $k = 0, 1, ..., p$. Indeed,

$$\begin{cases} \Sigma_p \text{ is an open dense subset of } S_p = \mathcal{M}_{m,n}(\mathbb{R}); \\ \text{If } k < p, \text{ the interior of } \Sigma_k \text{ is empty while its closure is } S_k. \end{cases}$$

Nevertheless, the sets S_k remain fairly complicated from the geometrical viewpoint.

---

166. ★★ *A convexity criterion using mean values in integral forms*

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function. For a (symmetric) positive definite matrix A ($A \succ 0$ in short), $x \in \mathbb{R}^n$ and $r > 0$, the elliptic convex set associated with these data is

$$\mathcal{E}(A, x, r) = \left\{ y \in \mathbb{R}^n : (y - x)^T A (y - x) \leqslant r \right\}.$$

$\mathcal{E}(A, x, r)$ is centered at x, its "radius" is measured by r, and its shape is governed by the eigenvalues of A. We denote by $\lambda[\mathcal{E}(A, x, r)]$ the volume (= the LEBESGUE measure) of $\mathcal{E}(A, x, r)$. The mean value of f on $\mathcal{E}(A, x, r)$, in an integral form, is

$$\mu_{\mathcal{E}(A,x,r)}(f) = \frac{1}{\lambda[\mathcal{E}(A, x, r)]} \int_{\mathcal{E}(A,x,r)} f(y) \, \mathrm{d}y.$$

Prove that f is convex if and only if

$$f(x) \leqslant \mu_{\mathcal{E}(A,x,r)}(f) \text{ for all } A \succ 0,\ x \in \mathbb{R}^n \text{ and } r > 0.$$

Hint. The key-fact is that f is convex if and only if its Hessian matrix (of second partial derivatives) $Hf(x)$ is positive semidefinite for all $x \in \mathbb{R}^n$.

To obtain the desired property, use second-order TAYLOR expansions with remainders in integral forms, for example.

---

167. ★★ *Logarithmically convex functions*

Let Ω be an open convex set in $\mathbb{R}^n$. A function $f : \Omega \to (0, +\infty)$ is said to be *logarithmically convex* (on Ω) when the function $\log f : \Omega \to \mathbb{R}$ is convex.

1°) Prove that a logarithmically convex function is necessarily convex.

2°) Assume that f is twice differentiable on Ω. Then prove the equivalence of the following three statements:

$$f \text{ is logarithmically convex;} \tag{1}$$

$$f(x)\nabla^2 f(x) \succcurlyeq \nabla f(x) \left[\nabla f(x)\right]^T \text{ for all } x \in \Omega; \tag{2}$$

$$x \mapsto g_a(x) = e^{a^T x} \times f(x) \text{ is convex for all } a \in \mathbb{R}^n. \tag{3}$$

3°) Deduce from the above that the sum of two logarithmically convex functions is logarithmically convex.

Hints. 2°) The key-ingredient is the characterization of a twice differentiable convex function h on Ω by the property " $\nabla^2 h(x) \succcurlyeq 0$ for all $x \in \Omega$ ", as well as the calculation of $\nabla^2 g_a(x)$ in terms of a, $\nabla f(x)$ and $\nabla^2 f(x)$.

Comment. It is possible to derive the equivalence [(1) $\Leftrightarrow$ (3)] in 2°) without assuming that f is differentiable.

---

168. ★★ *Convex functions vs quasi-convex functions*

A weakening of the notion of convexity is quasi-convexity. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *quasi-convex* if the following property holds true: for all x, y in $\mathbb{R}^n$ and all $t \in (0, 1)$,

$$f\left[tx + (1-t)y\right] \leqslant \max\left[f(x), f(y)\right]. \tag{1}$$

It is easy to check that this is equivalent to the following geometrical property: for all $r \in \mathbb{R}$, the sublevel-sets

$$\{x \in \mathbb{R}^n : f(x) \leqslant r\}$$

are convex.

All convex functions are quasi-convex, but quasi-convex functions do not share all of the nice properties of convex functions; for example: a quasi-convex function is not necessarily continuous (counterexample: $f(x) = 1$ for $x \neq 0, f(0) = 0$), a sum of quasi-convex functions is not necessarily quasi-convex (counterexample: add the linear function x to the preceding one). We intend here to prove how to characterize quasi-convex functions which are indeed convex.

Prove that $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the "tilted" version of f

$$x \in \mathbb{R}^n \mapsto f(x) + s^T x$$

is quasi-convex for any "slope" $s \in \mathbb{R}^n$.

Hints. The implication [$\Rightarrow$] is clear (the sum of convex functions is convex, hence quasi-convex). For the converse, use the analytical definition (1) of quasi-convexity.

Answer. Let $x \neq y$ in $\mathbb{R}^n$ and $t \in (0,1)$; set $z = (1-t)x + ty$. Choose a slope s such that $f(y) - f(x) = s^T(x-y)$ (or $f(y) + s^T y = f(x) + s^T x$). Now, since $x \in \mathbb{R}^n \mapsto f(x) + s^T x$ is quasi-convex, we have that

$$\begin{aligned} f(z) + s^T z &\leqslant \max\left[f(x) + s^T x, f(y) + s^T y\right] \\ &= (1-t)\left[f(x) + s^T x\right] + t\left[f(y) + s^T y\right], \end{aligned}$$

whence $f\left[(1-t)x + ty\right] \leqslant (1-t)f(x) + tf(y)$.

Comments. - This characterization of convex functions among quasi-convex functions is due to J.-P. Crouzeix (1977).

- Here are two additional results of interest in quasi-convex optimization. Let $f : \mathbb{R}^n \to \mathbb{R}$ be quasi-convex and $C \subset \mathbb{R}^n$ be convex; then:

(*i*) A strict local minimizer of f (on C) is actually a global minimizer;

(*ii*) A strict local maximizer of f (on C) is always an extremal point of C.

---

169. ★★ *Building $\mathcal{C}^1$-diffeomorphisms on $\mathbb{R}^n$ from convex functions*

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^2$ convex function. Prove that the mapping

$$\begin{aligned} F : \mathbb{R}^n &\to \mathbb{R}^n \\ p &\mapsto F(p) = p + \nabla f(p) \end{aligned}$$

is a $\mathcal{C}^1$-diffeomorphism from $\mathbb{R}^n$ onto $\mathbb{R}^n$.

Hints. To prove that F is bijective, for given $x \in \mathbb{R}^n$, consider the function

$$g_x : u \mapsto g_x(u) = f(u) + \frac{1}{2}\|u - x\|^2.$$

This function is minimized on $\mathbb{R}^n$ at a unique point $p_f(x)$; that point is characterized by the optimality condition

$$\nabla f[p_f(x)] + p_f(x) - x = 0;$$

it is therefore the right candidate for being $F^{-1}(x)$.

To prove that F^{-1} is $\mathcal{C}^1$, use the inverse function theorem.

Comment. For various f, one can determine explicitly the mapping $x \mapsto p_f(x)$. For example, if $f(x) = \frac{1}{2}x^T Ax$ with A positive semidefinite, then $p_f(x) = (I_n + A)^{-1}(x)$.

170. ★★ *Dilations in* $\mathbb{R}^n$

Let $\mathbb{R}^n$ be equipped with the usual Euclidean norm $\|\cdot\|$. A mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be a *dilation* (or anti-Lipschitzian) when:

$$\|F(x) - F(y)\| \geqslant \|x - y\| \text{ for all } x, y \text{ in } \mathbb{R}^n. \tag{1}$$

Prove that a dilation F of class $\mathcal{C}^1$ is a $\mathcal{C}^1$-diffeomorphism from $\mathbb{R}^n$ onto $\mathbb{R}^n$.

Hints. One has to prove that F is injective (easy), surjective, and apply the inversion theorems.

- For all $x \in \mathbb{R}^n$, the differential mapping $DF(x)$ is a linear isomorphism from $\mathbb{R}^n$ onto $\mathbb{R}^n$. To show that the linear mapping $DF(x) : \mathbb{R}^n \to \mathbb{R}^n$ is injective, derive from inequality (1):

$$\|DF(x) \cdot h\| \geqslant \|h\| \text{ for all } h \text{ in } \mathbb{R}^n. \tag{2}$$

- Inversion theorems allow us to claim that F is a $\mathcal{C}^1$-diffeomorphism from $\mathbb{R}^n$ onto the open set $F(\mathbb{R}^n)$.

- Surjectivity of F. At least two paths could be followed. The first one consists in proving that $F(\mathbb{R}^n)$ is closed (use for that the characterization of closedness of a set via converging sequences). The second one is via optimization techniques: given $z \in \mathbb{R}^n$, minimize the function $g(x) = \frac{1}{2}\|F(x) - z\|^2$ over $\mathbb{R}^n$; the infimum of g is indeed attained (because $g(x) \to +\infty$ whenever $\|x\| \to +\infty$); a solution x_0 satisfies $\nabla g(x_0) = DF(x_0) \cdot (F(x_0) - z) = 0$, whence $F(x_0) = z$.

Comment. Tapa 169 could be viewed as a particular instance of Tapa 170 since

$$\begin{aligned} \|F(p) - F(q)\|^2 &= \|(p + \nabla f(p)) - (q + \nabla f(q))\|^2 \\ &\geqslant \|p - q\|^2 + (\nabla f(p) - \nabla f(q))^T (p - q), \end{aligned}$$

and the convexity of f ensures that $(\nabla f(p) - \nabla f(q))^T (p - q) \geqslant 0$.

171. ★★ *The closest ordered sample from a given sample of real numbers*

Let $u = (u_1, ..., u_n) \in \mathbb{R}^n$. We are looking for $x = (x_1, ..., x_n)$ with $x_1 \leqslant x_1 \leqslant ... \leqslant x_n$, as close as possible to x (with respect to the usual Euclidean distance).

1°) Formalize this problem as the orthogonal projection of u onto a closed convex cone K in $\mathbb{R}^n$. Hence, there is one and only one $\overline{x} = (\overline{x}_1, ..., \overline{x}_n)$ answering the question.

2°) Provide a necessary and sufficient condition characterizing $\overline{x}$.

3°) An example in $\mathbb{R}^4$. Let $u = (2, 1, 5, 4)$. Find the unique $\overline{x} = (\overline{x}_1, \overline{x}_2, \overline{x}_3, \overline{x}_4) \in \mathbb{R}^4$ satisfying $\overline{x}_1 \leqslant \overline{x}_2 \leqslant \overline{x}_3 \leqslant \overline{x}_4$ which is closest to $(2, 1, 5, 4)$.

Hints. - Beware: reordering the coordinates $u_1, ..., u_n$ increasingly is not the correct answer.

- 2°) $\overline{x}$ minimizes $f(x) = \frac{1}{2} \|u - x\|^2$ under the $n - 1$ linear inequality constraints

$$g_i(x) = x_i - x_{i+1} \leqslant 0,\ i = 1, ..., n - 1.$$

It is suggested to write optimality conditions in this convex minimization problem (with a quadratic objective function and simple linear constraints).

Answers. 1°) Let $K = \{x = (x_1, ..., x_n) \in \mathbb{R}^n : x_1 \leqslant x_1 \leqslant ... \leqslant x_n\}$. Then, K is a closed convex cone, even polyhedral, in $\mathbb{R}^n$. The problem in question is to find the orthogonal projection of $u = (u_1, ..., u_n)$ onto K.

2°) $\overline{x} = (\overline{x}_1, \overline{x}_2, ... , \overline{x}_n) \in K$ is the orthogonal projection of $u = (u_1, ..., u_n)$ onto K if and only if:

$$(\mathcal{C}) \quad \begin{cases} \sum_{j=1}^{i} \overline{x}_j \leqslant \sum_{j=1}^{i} u_j \text{ for all } i = 1, ..., n-1; \\ \sum_{j=1}^{n} \overline{x}_j = \sum_{j=1}^{n} u_j; \\ \left(\sum_{j=1}^{i} u_j - \sum_{j=1}^{i} \overline{x}_j \right) (\overline{x}_i - \overline{x}_{i+1}) = 0 \text{ for all } i = 1, ..., n-1; \\ (u_n - \overline{x}_n)(\overline{x}_{n-1} - \overline{x}_n) = 0. \end{cases}$$

3°) The answer is $\overline{x} = \left(\frac{3}{2}, \frac{3}{2}, \frac{9}{2}, \frac{9}{2}\right)$. Its distance from $u = (2, 1, 5, 4)$ is 1, while the distance from the reordered sample $(1, 2, 4, 5)$ to $(2, 1, 5, 4)$

is 2.

---

172. ★★ *Expressing the sum of the m largest real numbers among n*

Let $x_1, .., x_n$ be n real numbers. Consider n real coefficients $\alpha_1, ..., \alpha_n$ satisfying

$$0 \leqslant \alpha_i \leqslant 1 \text{ for all } i = 1, ..., n \text{ and } \sum_{i=1}^{n} \alpha_i = m, \qquad (1_m)$$

where m is an integer between 1 and n.

We denote by $x_{i_1}, ..., x_{i_m}$ the m largest real numbers among the x_i' s.

1°) Show, via a direct calculation, that

$$\sum_{i=1}^{n} \alpha_i x_i \leqslant x_{i_1} + ... + \ x_{i_m}. \qquad (\mathcal{I})$$

2°) A second approach. Let Π_m denote the convex compact polyhedron in $\mathbb{R}^n$ consisting of all the $\alpha = (\alpha_1, ..., \alpha_n)$ satisfying (1_m).

(a) Example visualization. Let $n = 3$. Draw in the same picture Π_1, Π_2 and Π_3.

(b) Determine all the vertices of Π_m.

(c) Deduce from this the following evaluation:

$$x_{i_1} + ... + \ x_{i_m} = \max_{(\alpha_1, ..., \alpha_n) \in \Pi_m} \sum_{i=1}^{n} \alpha_i x_i. \qquad (\mathcal{J})$$

Hints. 1°) It helps to relabel indices so that

$$x_1 \geqslant x_2 \geqslant ... \geqslant x_m \geqslant x_{m+1} \geqslant \ ... \ \geqslant x_n.$$

2°) (c). Since every point in Π_m is a convex combination of vertices of Π_m, maximizing a linear form on Π_m amounts to maximizing it on the vertices of Π_m.

Answer. 2°) (b). The vertices of Π_m are those $(\overline{\alpha}_1, ..., \overline{\alpha}_m)$, m of which are equal to 1 and the remaining are equal to 0.

Comments. When $m = n$, there is nothing to prove: Π_n reduces to a singleton $\{(1, 1, .., 1)\}$ and $(\mathcal{I})$ (as well as $(\mathcal{J})$) trivially holds true.

When $m = 1$, the polyhedron Π_1 can also be described as

$$\Lambda_n = \left\{ (\alpha_1, ..., \alpha_n) : \ \alpha_i \geqslant 0 \text{ for all } i, \text{ and } \sum_{i=1}^{n} \alpha_i = 1 \right\},$$

called the unit-simplex of $\mathbb{R}^n$. Consequently, $(\mathcal{J})$ takes the following form in that case:

$$\max_{i=1,...,n} x_i = \max_{(\alpha_1,...,\alpha_n)\in\Lambda_n} \sum_{i=1}^{n} \alpha_i x_i.$$

---

173. ★★ *Unit spectraplex of positive semidefinite matrices*

Let $\mathcal{S}_n(\mathbb{R})$ be structured as a Euclidean space with the help of the scalar product $\langle\langle A, B\rangle\rangle = \mathrm{tr}(AB)$. We consider the following set

$$\Omega_1 = \{A \succcurlyeq 0 : \mathrm{tr}(A) = 1\}.$$

1°) Check that Ω_1 is a compact convex subset of $\mathcal{S}_n(\mathbb{R})$.

2°) Prove that the extreme points of Ω_1 are exactly the positive semidefinite matrices of the form xx^T, where x is a unit vector of $\mathbb{R}^n$.

3°) Calculate the support function of Ω_1, that is to say

$$M \in \mathcal{S}_n(\mathbb{R}) \mapsto \sigma_1(M) = \max_{A\in\Omega_1} \langle\langle M, A\rangle\rangle.$$

Hints. 2°) Use a spectral decomposition of A as $\sum_{i=1}^{n} \lambda_i(x_i x_i^T)$, where the λ_i's are eigenvalues of A and the x_i's are unit vectors in $\mathbb{R}^n$.

3°) The sought result involves one particular eigenvalue of M.

Answer. 3°) Since $\left\langle\left\langle M, xx^T\right\rangle\right\rangle = \langle Mx, x\rangle$, we have

$$\sigma_1(M) = \max_{\|x\|=1} \langle Mx, x\rangle = \lambda_{\max}(M) \ \text{ (the largest eigenvalue of } M).$$

Comments. - A consequence of 2°) is that

$$\Omega_1 = \mathrm{co}\left\{xx^T : \|x\| = 1\right\}.$$

- One can also check that $A \in \Omega_1$ is necessarily a matrix of the form

$$A = U\mathrm{diag}(\alpha_1, .., \alpha_n)U^T,$$

where U is an orthogonal matrix and $(\alpha_1, .., \alpha_n)$ is a vector in the unit-simplex of $\mathbb{R}^n$. This explains why Ω_1 is sometimes called the *unit spectraplex* or *spectraplex of order 1* (or even *1-spectrahedron*) of positive semidefinite matrices.

---

174. ★★ *Spectraplex of order m of positive semidefinite matrices*

Let $\mathcal{S}_n(\mathbb{R})$ be structured as a Euclidean space with the help of the scalar product $\langle\langle A, B\rangle\rangle = \mathrm{tr}(AB)$. Let m be an integer between 1 and n. We consider the following set

$$\Omega_m = \{A \succcurlyeq 0 : \mathrm{tr}(A) = m \text{ and } \lambda_{\max}(A) \leqslant 1\}. \tag{1}$$

1°) What is Ω_1? What is Ω_n?

2°) Check that Ω_m is a compact convex subset of $\mathcal{S}_n(\mathbb{R})$.

3°) Prove that the extreme points of Ω_m are exactly the positive semidefinite matrices of the form XX^T, where $X \in \mathcal{M}_{n,m}(\mathbb{R})$ satisfies $X^TX = I_m$.

4°) Calculate the support function of Ω_m, that is to say,

$$M \in \mathcal{S}_n(\mathbb{R}) \mapsto \sigma_m(M) = \max_{A \in \Omega_m} \langle\langle M, A\rangle\rangle.$$

Hints. 3°) Show $A \in \Omega_m$ is necessarily a matrix of the form

$$U\mathrm{diag}(\alpha_1, .., \alpha_n)U^T,$$

where U is an orthogonal matrix and $(\alpha_1, .., \alpha_n)$ is a vector in the following polyhedron Π_m of $\mathbb{R}^n$:

$$\Pi_m = \left\{(\alpha_1, .., \alpha_n) : 0 \leqslant \alpha_i \leqslant 1 \text{ for all } i = 1, ..., n \text{ and } \sum_{i=1}^{n} \alpha_i = m\right\}.$$

The extreme points (or vertices) of this convex polyhedron have been determined in Tapa 172. This explains why Ω_m is sometimes called the

spectraplex of order m (or even *m-spectrahedron*) of positive semidefinite matrices.

4°) The sought expression of this support function involves some eigenvalues of M.

Answers. 1°) We have

$\Omega_1 = \{A \succcurlyeq 0 : \mathrm{tr}(A) = 1\}$
(the additional constraint $\lambda_{\max}(A) \leqslant 1$ in (1) is not necessary here)
$= \mathrm{co}\left\{xx^T : \|x\| = 1\right\}$ (see Tapa 173),

while Ω_n just reduces to $\{I_n\}$.

4°) We have

$$\sigma_m(M) = \max_{A\in\Omega_m} \langle\langle M, A\rangle\rangle = \max_{X^TX=I_m} \langle\langle M, A\rangle\rangle \quad (2)$$

$$= \text{ sum of the } m \text{ largest eigenvalues of } M. \quad (3)$$

In particular, for $m = n$, $\sigma_n(M) =$ sum of all the eigenvalues of $M = \mathrm{tr}(M)$.

Comments. - This Tapa 174 can be seen as the m-version of the 1-version displayed in Tapa 173, and also the "matricial cousin" of Tapa 172. The variational formulation (2)–(3) of the sum of the m largest eigenvalues of $M \in \mathcal{S}_n(\mathbb{R})$ is due to Ky Fan (1949).

- A consequence of 2°) is that

$$\Omega_m = \mathrm{co}\left\{XX^T,\, X \in \mathcal{M}_{n,m}(\mathbb{R}) \text{ satisfying } X^TX = I_m\right\}.$$

---

175. ★★ *Local minimizers vs global minimizers*

Let $n \geqslant 2$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be the polynomial function of degree 5 defined by

$$f(x_1, ..., x_n) = (1 + x_n)^3 \sum_{i=1}^{n-1} x_i^2 + x_n^2.$$

1°) Show that 0 ($\in \mathbb{R}^n$) is the only critical (or stationary) point of f.

2°) Check that 0 is a strict local minimizer but not a global minimizer of f.

Answers. 1°) Via easy calculations of partial derivatives of f, one easily checks that $\overline{x} = 0$ is the only point $\overline{x}$ satisfying $\nabla f(\overline{x}) = 0$.

2°) Since, moreover, the Hessian matrix $\nabla^2 f(\overline{x}) = \mathrm{diag}(2, ..., 2)$ is positive definite, $\overline{x}$ is a strict local minimizer of f. But

$$f(1, ..., 1, x_n) = (n-1)\,(1 + x_n)^3 + x_n^2$$

is a polynomial function of degree 3 of the real variable x_n, whose range is therefore the whole of $\mathbb{R}$. Thus, $\overline{x}$ is not a global minimizer of f.

---

176. ★★ *Orthogonal projection over a convex elliptic set in* $\mathbb{R}^n$

Let $a_1, ..., a_n$ be n positive real numbers, and let $\mathcal{E}$ be the convex elliptic set in $\mathbb{R}^n$ defined as follows:

$$\mathcal{E} = \left\{ u = (u_1, ..., u_n) \in \mathbb{R}^n : \sum_{i=1}^{n} \left(\frac{u_i}{a_i} \right)^2 \leqslant 1 \right\}.$$

Let $x \notin \mathcal{E}$ and let $\overline{x}$ denote the orthogonal projection of x onto $\mathcal{E}$. Prove that

$$\overline{x}_i = \frac{a_i^2 x_i}{a_i^2 + \overline{\mu}} \quad \text{for all } i = 1, .., n, \tag{1}$$

where $\overline{\mu} > 0$ is the unique solution of the equation (in μ):

$$\sum_{i=1}^{n} \frac{a_i^2 x_i^2}{(a_i^2 + \mu)^2} = 1. \tag{2}$$

Hints. - The function $\mu \mapsto d(\mu) = \sum\limits_{i=1}^{n} \frac{a_i^2 x_i^2}{(a_i^2+\mu)^2}$ is continuous and strictly decreasing on $(0, +\infty)$; it decreases from $d(0) > 1$ (because $x \notin \mathcal{E}$) to $0 = \lim_{\mu \to +\infty} d(\mu)$.

- The sought $\overline{x}$ clearly should satisfy $\sum\limits_{i=1}^{n} \left(\frac{\overline{x}_i}{a_i} \right)^2 = 1$. Moreover, the optimality condition reads as follows: there exists a Lagrange multiplier $\overline{\mu} \geqslant 0$ such that

$$\overline{x}_i - x_i + \overline{\mu}\, \frac{\overline{x}_i}{a_i^2} = 0 \text{ for all } i = 1, .., n.$$

---

177. ★★ *Unit partition in the set of positive definite matrices*

Given $k+1$ nonzero vectors $x^1, ..., x^k, y$ in $\mathbb{R}^n$, we ask when it is possible to find positive definite matrices $M_1, .., M_k$ such that:

$$\left\{ \begin{array}{c} M_i y = x^i \text{ for all } i = 1, ..., k, \\ \text{and} \\ M_1 + ... + M_k = I_n. \end{array} \right. \qquad (\mathcal{C})$$

The first equation in $(\mathcal{C})$ is of the so-called *quasi*-NEWTON *type* in numerical optimization; the second gives the name of the problem.

1°) Check that a necessary condition for the existence of such matrices is:

$$\left\langle x^i, y \right\rangle > 0 \text{ for all } i = 1, ..., k \text{ and } \sum_{i=1}^{k} x^i = y. \qquad (1)$$

Unfortunately, this condition is not sufficient...

2°) Prove that a necessary and sufficient condition for the existence of positive definite matrices M_i satisfying the two conditions in $(\mathcal{C})$ is that:

$$I_n - \sum_{i=1}^{k} \frac{x^i \left(x^i\right)^T}{\left\langle x^i, y \right\rangle} \text{ is positive definite on the hyperplane } y^{\perp}.$$

Hint. The main tools to be used are the CAUCHY–SCHWARZ inequality and the basic properties of positive definite matrices.

Comment. For more on this problem and on positive semidefinite or positive definite matrices, see the survey

J.-B. HIRIART-URRUTY and J. MALICK, *A fresh variational-analysis look at the positive semidefinite matrices world.* J. of Optimization Theory and Applications 153 (3), 551–577 (2012).

---

178. ★★ *A minimization problem on positive definite matrices*

Let Ω_n denote the set of $n \times n$ symmetric positive definite matrices. For A and B in Ω_n, consider the following minimization problem:

$$(\mathcal{P}) \quad \left\{ \begin{array}{c} \text{Minimize } f(X) = \text{tr}(AX) + \text{tr}(BX^{-1}) \\ \text{subject to } X \in \Omega_n. \end{array} \right.$$

1°) Show that $(\mathcal{P})$ has one and only one solution.

2°) (*a*) Check that the solution of $(\mathcal{P})$ is the unique $\overline{X} \in \Omega_n$ satisfying the matricial equation

$$\overline{X}A\overline{X} = B. \tag{1}$$

(*b*) Describe a process which, starting from BA (or AB), allows one to obtain $\overline{X}$.

(*c*) Deduce from the above the optimal value $\overline{f}$ in $(\mathcal{P})$.

3°) We choose $n = 2$ here. Show that the optimal value $\overline{f}$ in $(\mathcal{P})$ is

$$\overline{f} = 2\sqrt{\mathrm{tr}(AB) + 2\sqrt{\det(AB)}}. \tag{2}$$

Hints. The case $n = 1$ allows us to guide the approach and to control the results.

1°) Show that f is differentiable and strictly convex on Ω_n, with $f(X) \to +\infty$ when X tends to a matrix on the boundary of Ω_n.

2°) (*a*). In the space $\mathcal{S}_n(\mathbb{R})$, made Euclidean with the inner product $\langle U, V\rangle = \mathrm{tr}(UV)$, calculate the gradient of f at X.

Answers. 2°) (*b*). As a product of two positive definite matrices, BA is diagonalizable and its eigenvalues are all positive (*cf.* Tapa 135 if necessary). If P diagonalizes BA, then the matrix

$$M = P\mathrm{diag}(\sqrt{\lambda_1}, ..., \sqrt{\lambda_n})P^{-1} \text{ (where } \lambda_1, .., \lambda_n \text{ are eigenvalues of } BA)$$

satisfies $M^2 = BA$, and $\overline{X} = M^{-1}B$ is the sought solution.

(*c*) The optimal value in $(\mathcal{P})$ is twice the sum of the square roots of the eigenvalues λ_i of BA (or AB).

3°) Write $\sqrt{\lambda_1} + \sqrt{\lambda_2}$ as $\sqrt{\lambda_1 + \lambda_2 + 2\sqrt{\lambda_1\lambda_2}}$.

---

179. ★★ *Maximizing the area of a triangle of given perimeter*

Among the triangles having the same perimeter, determine those which have the largest area.

Hints. The work consists in formalizing the posed question as an optimization problem and in studying its essential points: existence of

solutions, detecting points satisfying the first-order optimality conditions, analyzing and synthesizing the obtained results to answer the posed question.

If a, b, c denote the lengths of the sides of a triangle ABC, so that its perimeter is $a+b+c=2p$, the area of ABC is $\sqrt{p(p-a)(p-b)(p-c)}$ (according to the ARCHIMEDES–HERON formula). The positive real numbers a, b, c are the lengths of the sides of a triangle if and only if

$$a \leqslant b+c,\ b \leqslant a+c,\ c \leqslant a+b.$$

Answer. Equilateral triangles ABC are the solutions of the posed problem. With $2p$ denoting its perimeter, the maximal obtained area is then $\frac{p^2}{3\sqrt{3}}$.

---

180. ★★ *Maximizing the area of a hinged quadrilateral of given perimeter*

In the plane, consider a mechanical system of four linked metal rods forming a convex quadrilateral (or quadrangle); the junctions of the rods pivot freely, hence the system is a hinged one – it can be articulated. The objective is to determine the configuration of the system so that the formed convex quadrilateral is of maximal area.

Consider a numerical example with $a = 4\text{cm}$, $b = 5\text{cm}$, $c = 5\text{cm}$, $d = 3\text{cm}$.

Hints. Let $ABCD$ be a convex quadrilateral whose sides have lengths: $AB = c$, $BC = d$, $CD = a$, $AD = d$. Here it is supposed that $a+b+c+d = 2p$ is fixed. There is no formula giving the area of $ABCD$ in terms of a, b, c, d (or of p) alone; this is easy to understand since a mechanical system with the same lengths of sides (or the same perimeter $2p$) can be deformed by articulations and the areas of the resulting convex quadrilaterals vary. Let α and β denote the angles at the vertices D and B. After some laborious calculations starting from the areas of triangles ADC and ABC (having the side AC in common) and exploiting some trigonometrical identities, one arrives at the following expressions for the area $\mathcal{A}$ of $ABCD$:

$$\mathcal{A}^2 = \frac{1}{16}[4(a^2b^2+c^2d^2)-(a^2+b^2-c^2-d^2)^2-8\ abcd\cos(\alpha+\beta)];\ (1)$$

$$\mathcal{A}^2 = (p-a)(p-b)(p-c)(p-d) - abcd\cos^2\left(\frac{\alpha+\beta}{2}\right). \qquad (2)$$

$$\mathcal{A}^2 = (p-a)(p-b)(p-c)(p-d) - abcd\cos^2\left(\frac{\alpha+\beta}{2}\right). \qquad (3)$$

However, when the vertices are cocyclical (*i.e.,* lie on a single circle), there is a formula giving the area $\mathcal{A}$ of $ABCD$ in terms of p alone, this is BRAHMAGUPTA's formula (Indian mathematician, 598–668):

$$\mathcal{A} = \sqrt{(p-a)(p-b)(p-c)(p-d)}. \qquad (4)$$

Note that the limiting case of BRAHMAGUPTA's formula, when $d \to 0$ for example, is the ARCHIMEDES–HERON formula for the area of a triangle (see the previous Tapa 179); in some sense, a triangle is a degenerate quadrilateral which is always inscribable.

Answer. The maximal area is obtained in (1) (or in (2) and (3)) when $\cos(\alpha+\beta) = -1$ (or when $\cos\left(\frac{\alpha+\beta}{2}\right) = 0$), that is to say, when $\alpha+\beta = \pi$; this corresponds to the situation where the four vertices A, B, C, D are put on the same circle (a so-called inscribable quadrilateral), which is always possible.

Numerical example: the maximal area is $\simeq 17.41\text{cm}^2$.

Comments. - The German mathematicians C.A. BRETSCHNEIDER and K.G. VON STAUT discovered in the same year (1842) the formulas (2) and (3) giving the area of a convex quadrilateral in terms of its four sides and two opposite angles.

- Using BRAHMAGUPTA's formula, one could solve the following optimization problem, echoing the one posed in the preceding Tapa 179: among the inscribable quadrilaterals of given perimeter $2p$, the square (of sides with common length $\frac{2}{p}$) is the one maximizing the area.

- Generalization. Let $P = A_1A_2....A_n$ be a varying convex polygon in the plane (the rigid bars A_iA_{i+1} pivoting at their ends), with fixed lengths of sides A_1A_2, A_2A_3, ..., $A_{n-1}A_n$, A_nA_1. Then it can also be proved that the maximal value of the area of P is achieved when the polygon is inscribed in a circle. However, we do not know of any (simple) formula like BRAHMAGUPTA's in that case.

A preliminary question however could be: given n lengths $l_1, l_2, ..., l_n$, when are they the lengths of sides of a convex polygon? The answer, easy to check, is: if, and only if, the largest length l_{i_0} is less than the sum of the $n-1$ other lengths $\left(i.e.,\ \sum_{i\neq i_0} l_i\right)$.

181. ★★ *An isoperimetric inequality for polygons with n vertices*

Let $P = A_1A_2....A_n$ be a varying convex polygon, with n vertices and fixed perimeter L. Let $\mathcal{A}$ denote the area of P.

Show that

$$\frac{\mathcal{A}}{L^2} \leqslant \frac{1}{4n\tan(\pi/n)}, \tag{1}$$

with equality if and only if P is a regular convex polygon.

Hints. - A first step consists in proving that those convex polygons, with n vertices and perimeter L, that maximize $\frac{\mathcal{A}}{L^2}$ should be inscribed in a circle.

- Indeed, for a regular convex polygon with n vertices inscribed in a circle of radius 1, we have:

$$L = 2n\sin\left(\frac{\pi}{n}\right),\ \mathcal{A} = \frac{n}{2}\sin\left(\frac{2\pi}{n}\right). \tag{2}$$

---

182. ★★ *Maximizing the area of lateral parts of a tetrahedron*

Let us consider tetrahedrons (or pyramids) with fixed basis ABC and fixed height h. Find (if any!) tetrahedrons of this type whose sum of areas of the three lateral triangles is minimal.

Hint. Formalize the posed question as an optimization problem with constraints. Consider the orthogonal projection H of the top O of the pyramid $OABC$ on the plane containing the triangle ABC. The relevant variables in the optimization problem are the distances h_1, h_2, h_3 of H to the sides of the triangle ABC.

Answer. The optimal pyramid is the one where O is projected onto the center of the inscribed circle of the triangle ABC.

Comment. The posed question is a particular case of a more general one, called the problem of A. Pleijel (1955), for which some unanswered questions remain. Here it is. Let C be a convex compact set contained in the horizontal plane (P). Given $h > 0$ and a point x_h in

C, one forms the piece of a convex cone whose basis is C and whose apex is situated on the perpendicular line to (P) at x_h, at height h. One then defines $\mathcal{A}(x_h, h)$ as the area of the lateral surface of this piece of a convex cone. Questions:

- For a given $h > 0$, does there exist a unique $\overline{x}_h$ minimizing $\mathcal{A}(x_h, h)$? The answer is Yes. However it is unknown how to characterize and how to determine such a $\overline{x}_h$ for general C.

- What is the behavior of $\overline{x}_h$ when $h \to 0$? The answer is known only for some particular cases of C.

---

183. ★★ *Minimizing the area of a plate positioned on the three axes of coordinates*

In the usual three-dimensional affine Euclidean space, marked with an orthonormal basis $(O, \overrightarrow{i}, \overrightarrow{j}, \overrightarrow{k})$, one considers triangular plates ABC with vertices $A = (x, 0, 0)$, $B = (0, y, 0)$, $C = (0, 0, z)$, where $x > 0$, $y > 0$, $z > 0$. We assume that these plates are forced to pass through a fixed point (a, b, c), where $a > 0$, $b > 0$, $c > 0$ are given coordinates. Question: among all these triangular plates, are there any minimizing the area? If so, how do we find such "minimal" plates?

Numerical example with $a = 1$, $b = 4$, $c = 6$.

Hints. The posed question can be formalized as a constrained optimization problem, which opens the way to existence and uniqueness results, characterization of solutions via the LAGRANGE theorem, and so on. The square of the area of the triangular plate ABC is $\frac{1}{4}(x^2y^2 + x^2z^2 + y^2z^2)$; the constraints imposed on the plate are:

$$\begin{cases} \frac{a}{x} + \frac{b}{y} + \frac{c}{z} = 1, \\ x > 0,\ y > 0,\ z > 0. \end{cases}$$

Answers. There is one and only one configuration minimizing the area of the triangular plate ABC. Here is a way to find it. Consider the function of the real variable

$$\theta(t) = \frac{a}{a + \sqrt{t + a^2}} + \frac{b}{b + \sqrt{t + b^2}} + \frac{c}{c + \sqrt{t + c^2}}. \qquad (1)$$

The θ function is continuous, strictly decreasing on $[0, +\infty)$, with $\theta(0) = 3/2$ and $\lim_{t \to +\infty} \theta(t) = 0$. Hence, the equation $\theta(t) = 1$ has a unique solution, denoted by t^*. The triangular plate with minimal area has

$$x^* = a + \sqrt{t^* + a^2},\ y^* = b + \sqrt{t^* + b^2},\ z^* = c + \sqrt{t^* + c^2}. \qquad (2)$$

Numerical example. Here, $t^* = 22.82$, so that $x^* = 5.88$, $y^* = 10.23$, $z^* = 13.67$ (numerical approximations, of course).

Comment. The problem can be generalized to $\mathbb{R}^n$ with $n > 3$ and solved in a very similar way.

---

184. ★★ *Minimizing the energy in a problem of* COULOMB *type*

Consider N distinct points in the three-dimensional space $\mathbb{R}^3$, $N \geqslant 3$. By choosing groups of 3 points, one builds $\left(\begin{array}{c} N \\ 3 \end{array} \right) = \frac{N(N-1)(N-2)}{6}$ triangles; since each triangle has 3 angles, one has finally $\frac{N(N-1)(N-2)}{2}$ angles θ_i built from a configuration of N points. The so-called "local energy" function (of COULOMB type) to be minimized has the following expression:

$$E_N = -\frac{1}{4}\left[\frac{N(N-1)}{2} + F_N\right],$$

where

$$F_N = \sum_{\text{all the angles } \theta_i} \cos(\theta_i).$$

The posed problem is to find configurations of N distinct points minimizing the energy function E_N, hence maximizing F_N.

Objective in this Tapa: solve the problem for $N = 3$ and $N = 4$.

Hints. The posed question can be formalized as a constrained optimization problem, which opens the way to existence results, necessary conditions for optimality via the LAGRANGE theorem, etc. For example, for $N = 3$, one has to pass through maximizing $\cos(\theta_1) + \cos(\theta_2) + \cos(\theta_3)$ under the constraints

$$\left\{ \begin{array}{c} \theta_1 + \theta_2 + \theta_3 = \pi, \\ \theta_1 \geqslant 0,\ \theta_2 \geqslant 0,\ \theta_3 \geqslant 0. \end{array} \right.$$

Answers. For $N = 3$, the optimal configurations are provided by equilateral triangles.

For $N = 4$, the optimal configurations are provided by regular tertrahedrons.

Comment. For $N \geqslant 5$, the problem of finding the optimal configurations is still not completely solved.

185. ★★ *An approximation problem on matrices*

Let $\mathcal{M}_n(\mathbb{R})$ be structured as a Euclidean space with the help of the inner product $\langle\langle A, B\rangle\rangle = \operatorname{tr}(A^T B)$; we denote by $\|\cdot\|$ the resulting matricial norm. Let S denote the following closed set of $\mathcal{M}_n(\mathbb{R})$:

$$S = \{M \in \mathcal{M}_n(\mathbb{R}) : \det(M) = 1\}.$$

Question: Which matrices in S are closest to the null matrix? In other words, we have to solve the following approximation or projection problem:

$$\begin{cases} \text{Minimize } \|M\| \\ \text{subject to } M \in S. \end{cases}$$

Hints. The posed question is formalized as a constrained optimization problem, with $1/2\,\|M\|^2$ as an objective function, S as a constraint set. The existence of solutions does not offer any difficulty; necessary conditions for optimality via the LAGRANGE theorem easily help to delineate the solutions.

Answer. All the orthogonal matrices whose determinant equals 1 are solutions of the posed problem.

186. ★★ *A maximization problem on the unit sphere*

1°) Determine the maximal value of $\prod_{i=1}^{n} x_i^2$ subject to the constraint $\sum_{i=1}^{n} x_i^2 = 1$.

2°) Deduce from this the following inequality: for all $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$,

$$\left|\prod_{i=1}^{n} x_i\right| \leqslant \left(\frac{\|x\|}{\sqrt{n}}\right)^n, \tag{1}$$

where $\|\cdot\|$ denotes the usual Euclidean norm on $\mathbb{R}^n$.

Hints. The posed question can be formalized as a constrained maximization problem, with $f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} x_i^2$ as an objective function, and

$$h(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} x_i^2 - 1 = 0$$

as an equality constraint. Necessary conditions for optimality via the LAGRANGE theorem help to determine the solutions, hence to obtain the asked optimal value.

Answers. 1°) We have that

$$\max_{\sum_{i=1}^{n} x_i^2 = 1} \left(\prod_{i=1}^{n} x_i^2 \right) = \frac{1}{n^n}. \tag{2}$$

2°) If $x = (x_1, x_2, ..., x_n)$ is a non-null vector in $\mathbb{R}^n$, the normalized vector $y = \frac{x}{\|x\|}$ satisfies $\sum_{i=1}^{n} y_i^2 = 1$. Hence, the result (1) leads to the inequality (2).

---

187. ★★ *Minimization of an electrostatic energy*

For $n \geqslant 2$, let $\mathcal{O}$ denote the set of $x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ such that $x_i \neq x_j$ for $i \neq j$. One defines

$$x = (x_1, x_2, ..., x_n) \in \mathcal{O} \mapsto f(x) = \sum_{i=1}^{n} x_i^2 - \sum_{1 \leqslant i < j \leqslant n} \log(|x_i - x_j|). \tag{1}$$

The objective here is to study the problem of minimizing f on $\mathcal{O}$.

1°) After having checked that minimizers of f on $\mathcal{O}$ exist, consider such a minimizer $(a_1, a_2, ..., a_n)$. One sets

$$t \in \mathbb{R} \mapsto H(t) = (t - a_1)(t - a_2)... (t - a_n). \tag{2}$$

(*i*) Show that:

$$H''(a_k) - 4a_k H'(a_k) = 0 \text{ for all } k = 1, 2, ..., n. \tag{3}$$

(*ii*) Deduce from the above that:

$$H'' - 4tH' = -4nH. \tag{4}$$

(*iii*) Imagine now a procedure to determine the minimizers of f on $\mathcal{O}$ as well as the optimal value $\overline{f}$.

2°) Use the procedure resulting from the previous question to calculate $\overline{f}$ in the case where $n = 2$ and $n = 3$.

Hints. 1°) (i). Start with the relation $H(t) = (t - a_k)Q_k(t)$, where

$$Q_k(t) = \prod_{j \neq k} (t - a_j) = \frac{H(t)}{t - a_k},$$

and calculate $H'(a_k)$ and $Q'_k(a_k)$ in terms of $Q_k(a_k)$. Then, make use of the optimality condition $\nabla f(a_1, a_2, ..., a_n) = 0$.

(ii) The polynomial functions H and $H'' - 4tH'$ are both of degree n and both have n distinct real roots $a_1, a_2, ..., a_n$; therefore, there exists a constant c such that $H'' - 4tH' = cH$.

Answers. 1°) (iii). The expression $H(t) = c_0 + c_1 t + ... + c_{n-1}t^{n-1} + t^n$, substituted into (4), allows us to determine the coefficients c_i in H. Then, solving the equation $H(t) = 0$ gives rise to n real distinct roots $a_1, a_2, ..., a_n$. Finally, the $n!$ points with coordinates $(a_{i_1}, a_{i_2}, ..., a_{i_n})$, where $(i_1, i_2, ..., i_n)$ is a permutation of $(1, 2, ..., n)$, giving rise to the same value $\overline{f}$, are the minimizers of f on $\mathcal{O}$.

2°) Case $n = 2$. The two minimizers are $\left(-\frac{1}{2}, \frac{1}{2}\right)$ and $\left(\frac{1}{2}, -\frac{1}{2}\right)$; the optimal value $\overline{f}$ is $\frac{1}{2}$.

Case $n = 3$. The six minimizers are, within a permutation of coordinates, of the form $\left(-\frac{\sqrt{3}}{2}, 0, \frac{\sqrt{3}}{2}\right)$; the optimal value $\overline{f}$ is $\frac{3}{2} - \log\left(\frac{3\sqrt{3}}{4}\right)$.

---

188. ★★ *Characterizing global minimizers in quadratic optimization*

Let us consider the following optimization problem:

$$(\mathcal{P}) \quad \left\{ \begin{array}{c} \text{Minimize } f(x) = \frac{1}{2}x^T A_0 x + b_0^T x + c_0 \\ \text{subject to } h(x) = \frac{1}{2}x^T A_1 x + b_1^T x + c_1 = 0, \end{array} \right.$$

where A_0 and A_1 are $n \times n$ symmetric matrices, b_0 and b_1 are vectors in $\mathbb{R}^n$, and c_0 and c_1 are real numbers. The objective function f in $(\mathcal{P})$ is not assumed to be convex (*i.e.,* A_0 is not necessarily positive semidefinite); also the constraint set, defined as $\{x \in \mathbb{R}^n : h(x) = 0\}$, could be disconnected. We assume, without loss of generality, that $A_1 \neq 0$ and:

There is an x^* such that $h(x^*) > 0$; there is an x_* such that $h(x_*) < 0$.

The usual LAGRANGE necessary condition for optimality asserts the following: If $\overline{x}$ satisfying $h(\overline{x}) = 0$ is a local minimizer in $(\mathcal{P})$, then there exists a $\lambda \in \mathbb{R}$ (called a multiplier) such that $\nabla f(\overline{x}) + \lambda \nabla h(\overline{x}) = 0$. This raises the question: what additional condition would ensure that $\overline{x}$ is a global minimizer in $(\mathcal{P})$? We intend to answer this question.

1°) Prove that $\overline{x}$ satisfying $h(\overline{x}) = 0$ is a global minimizer in $(\mathcal{P})$ if and only if there exists a $\lambda \in \mathbb{R}$ such that:

$$\left\{ \begin{array}{l} \nabla f(\overline{x}) + \lambda \nabla h(\overline{x}) = (A_0 + \lambda A_1)\,\overline{x} + b_0 + \lambda b_1 = 0; \\ A_0 + \lambda A_1 \text{ is positive semidefinite.} \end{array} \right. \qquad (\mathcal{C})$$

So, just adding the condition on the positive semidefiniteness of $A_0 + \lambda A_1$ to the classical LAGRANGE condition gives us a way of *characterizing global solutions* in $(\mathcal{P})$.

2°) An example. Apply the results above to the following optimization problem in $\mathbb{R}^2$: Find the closest points to $(1, 0)$ in the hyperbola with equation $x_1^2 - x_2^2 = 4$.

Hints. - The key-ingredient is that the second-order TAYLOR developments of the involved functions are exact (because they are quadratic); for example, for $d \in \mathbb{R}^n$ and $t \in \mathbb{R}$,

$$f(\overline{x} + td) = f(\overline{x}) + t\,(A_0\overline{x} + b_0)^T\, d + \frac{t^2}{2} d^T A_0 d.$$

- Begin by proving the result in the simpler case where

$$h(x) = \frac{1}{2}\left(\|x\|^2 - r^2\right).$$

Answer. 2°) In this example, $f(x_1, x_2) = (x_1 - 1)^2 + x_2^2$, $h(x_1, x_2) = x_1^2 - x_2^2 - 4$. Two points satisfy the LAGRANGE condition:

$$\begin{array}{rcl} (2, 0) \text{ with the multiplier } \lambda & = & -\dfrac{1}{2}; \\ (-2, 0) \text{ with the multiplier } \lambda & = & -\dfrac{3}{2}. \end{array}$$

But the condition "$A_0 + \lambda A_1$ is positive semidefinite" is satisfied only in the first case. Thus, $(2, 0)$ is the (global) solution in our problem. Of course, in this very simple orthogonal projection of a point onto the hyperbola, we knew which point was going to be the solution...

Comments. - An instance of problem $(\mathcal{P})$ is that of determining the points in the hypersurface with equation $h(x) = 0$ which are closest to a given point $a \in \mathbb{R}^n$.

- The result of this Tapa is surprising... We do not expect similar results with two or more quadratic equality constraints $h_i(x) = 0$, or when the function h defining the sole equality constraint $h(x) = 0$ is not quadratic.

---

189. ★★ *An integral vs a sum of a series involving similar terms*

Let the function $x > 0 \mapsto x^x$ be extended by continuity at 0 by defining $0^0 = 1$. In the same vein, the function $x > 0 \mapsto x \ln(x)$ is extended by continuity at 0 by defining $0 \ln(0) = 0$. Thus, the relation $x^x = e^{x \ln(x)}$ is in force for all $x \in [0, 1]$.

We intend to prove the following remarkable identity:

$$\int_0^1 \frac{1}{x^x} \, \mathrm{d}x = \sum_{n=1}^{+\infty} \frac{1}{n^n}. \tag{1}$$

The suggested steps to follow are:

(i) For $x > 0$, develop $\frac{1}{x^x} = e^{-x \ln(x)}$ as a power series and deduce from that:

$$\int_0^1 \frac{1}{x^x} \, \mathrm{d}x = \sum_{n=0}^{+\infty} \frac{1}{n!} \int_0^1 [-x \ln(x)]^n \, \mathrm{d}x. \tag{2}$$

(ii) For positive integers p and q, let $I(p, q) = \int_0^1 x^p [\ln(x)]^q \, \mathrm{d}x$. Justify the convergence of such an integral.

(iii) Calculate $I(m, 0)$ and, via an integration by parts, $I(p, q)$.

(iv) Conclude the relation (1).

Answers. (ii) For $x > 0$ small enough, we have

$$|x^p [\ln(x)]^q| \leqslant \frac{1}{\sqrt{x}},$$

and the integral $\int_0^1 \frac{1}{\sqrt{x}} \, \mathrm{d}x$ is convergent.

(iii) We have:

$$I(m, 0) = \frac{1}{m+1}; \tag{3}$$

$$I(p, q) = -\frac{q}{p+1} \, I(p, q-1). \tag{4}$$

We thus get at

$$I(p,q) = \frac{(-1)^q\ q!}{(p+1)^q}\ I(p,0) = (-1)^q \frac{q!}{(p+1)^{q+1}}. \tag{5}$$

(*iv*) With (2) and (5), one obtains:

$$\int_0^1 \frac{1}{x^x}\ \mathrm{d}x = \sum_{n=0}^{+\infty} \frac{1}{n!}(-1)^n I(n,n) = \sum_{n=0}^{+\infty} \frac{1}{(n+1)^{n+1}}.$$

Comment. A formula similar to (1), proved in the same way, is the following:

$$\int_0^1 x^x\ \mathrm{d}x = -\sum_{n=1}^{+\infty} \frac{(-1)^n}{n^n}. \tag{6}$$

Part 3. Three-starred tapas

"A great discovery solves a great problem, but there is a grain of discovery in the solution of any problem"

G. Polya (1887–1985)

190. ★★★ *Sets of continuous linear mappings vs their images*

Let E and F be two Banach spaces and let $\mathcal{L}(E,F)$ denote the set of continuous linear mappings from E to F. Given $\mathcal{A} \subset \mathcal{L}(E,F)$, we are interested in answering the following question: does $B \in \mathcal{L}(E,F)$ satisfying

$$Bx \in \mathcal{A}x \text{ for all } x \in E \tag{1}$$

necessarily belong to $\mathcal{A}$? [In (1), $\mathcal{A}x$ stands for the set $\{Ax : A \in \mathcal{A}\}$.]

Prove that the answer is Yes when $\mathcal{A}$ is a countable set in $\mathcal{L}(E,F)$.

Hint. Here is an occasion where we can use a form of the Baire category theorem: If (the Banach space) E is written as a countable union of closed sets, then at least one of those closed sets has a nonempty interior.

Answer. The kernels of $B - A_i$ for all $A_i \in \mathcal{A}$ are closed subspaces of E (because $A_i \in \mathcal{L}(E,F)$) whose union is E (since for every x in E,

$Bx \in \mathcal{A}x$ implies that $Bx = A_{i(x)}x$ for some $A_{i(x)}$ in $\mathcal{A}$). But then, by the BAIRE category theorem, for some A_{i_0} the kernel of $B - A_{i_0}$ must have a nonempty interior, hence must be all of E, so that $B = A_{i_0}$.

Comment. It is a bit surprising that the answer to the posed question is Yes. Even with E and F finite-dimensional, the answer could be No for sets $\mathcal{A}$ which are not countable. The problem is discussed in Tapa 323 in:

J.-B. HIRIART-URRUTY, *Mathematical tapas.* Volume 1 (for Undergraduates). Springer (2016).

---

191. ★★★ *Surprising characterizations of the exponential and cosine functions*

We consider here only functions $f : \mathbb{R} \to \mathbb{R}$ which are of class $\mathcal{C}^\infty$.

1°) Prove that the exponential function $f : x \mapsto f(x) = e^x$ is the only one satisfying:

$$\begin{aligned} f^{(n)}(x) &\geqslant 0 \text{ for all } x \in \mathbb{R} \text{ and all integers } n \geqslant 0; \\ f(0) &= f'(0) = f''(0) = 1. \end{aligned}$$

2°) Prove that the cosine function $f : x \mapsto f(x) = \cos(x)$ is the only one satisfying:

$$\begin{aligned} \left|f^{(n)}(x)\right| &\leqslant 1 \text{ for all } x \in \mathbb{R} \text{ and all integers } n \geqslant 0; \\ f(0) &= 1 \text{ and } f''(0) = -1. \end{aligned}$$

Comments. - This surprising characterization of the exponential function is due to S. BERNSTEIN (1928); the equally surprising characterization of the cosine function is due to H. DELANGE (1967).

- In the same vein: the sine function $f : x \mapsto f(x) = \sin(x)$ is the only one satisfying:

$$\begin{aligned} \left|f^{(n)}(x)\right| &\leqslant 1 \text{ for all } x \in \mathbb{R} \text{ and all integers } n \geqslant 0; \\ f'(0) &= 1 \text{ (no condition on } f(0)). \end{aligned}$$

---

192. ★★★ *The convex polyhedron of bistochastic matrices*

In the space of $n \times n$ real matrices, consider the affine subspace

$$V_n = \left\{ M = [m_{ij}] : \sum_{j=1}^{n} m_{ij} = 1 \text{ for all } i,\ \sum_{i=1}^{n} m_{ij} = 1 \text{ for all } j \right\}$$

and the convex compact set (of so-called *bistochastic matrices* of size n)

$$\Sigma_n = \{ M \in V_n : m_{ij} \geqslant 0 \text{ for all } i, j = 1, ..., n \}.$$

1°) Show that the affine hull of Σ_n (that is, the smallest affine space containing Σ_n) is V_n, and that V_n has dimension $(n-1)^2$.

2°) Set $J_n = \left[\frac{1}{n}\right]$ (the $n \times n$ bistochastic matrix all of whose elements are equal; it is a "central" matrix in Σ_n). Check that J_n lies in the relative interior of V_n (*i.e.*, the interior considered in the topological space V_n, not in $\mathbb{R}^n$), and that

$$J_n M = M J_n = J_n \text{ for all } M \in \Sigma_n$$
(J_n "absorbs" the whole of Σ_n).

3°) Now equip the space of matrices with the standard inner product $\langle M, P \rangle = \operatorname{tr}(M^T P)$. Denoting by $e \in \mathbb{R}^n$ the vector whose components are all 1, show that

$$V_n^{\perp} = \left\{ ue^T + ev^T : u \in \mathbb{R}^n,\ v \in \mathbb{R}^n \right\}. \tag{1}$$

4°) Deduce that the orthogonal projection of a matrix M onto V_n is

$$M_n' = J_n + K_n M K_n, \tag{2}$$

where $K_n = I_n - J_n$.

Comment. One could therefore think of projecting a matrix M onto Σ_n in two steps: first project M onto V_n (with the help of (2)) and then, within V_n, project M_n' onto Σ_n.

---

193. ★★★ *Vertices of the convex polyhedron of bistochastic matrices*

Let P be a convex polyhedron of $\mathbb{R}^n$ described as

$$P = \{ x \in \mathbb{R}^n : Ax = b, x \geqslant 0 \}, \tag{1}$$

where A is an $m \times n$ matrix and $b \in \mathbb{R}^m$.

1°) For $0 \neq x = (\xi_1, ..., \xi_n) \in P$, show that x is an extreme point (or vertex) in P if and only if the columns of A corresponding to indices i for which $\xi_i > 0$ are linearly independent vectors.

2°) Use this result to prove the following theorem (G. BIRKHOFF (1946)): the extreme points (or vertices) of the compact convex set Σ_n of bistochastic matrices are the permutation matrices (a permutation matrix is a square matrix that has exactly one entry of 1 in each row and each column and 0's elsewhere).

Hint. Describe $\Sigma_n \subset \mathcal{M}_n(\mathbb{R})$ as a polyhedron in $\mathbb{R}^{n \times n}$ of the form (1), with appropriate A and b.

Comment. Permutation matrices can be characterized in various ways; here is one which is fairly original. Let A be an invertible matrix such that the entries of both A and A^{-1} are nonnegative integers; then A is a permutation matrix.

---

194. ★★★ *An integral defined from the distance to a convex polyhedron in the plane*

Let C be a convex compact polyhedron in the plane; we denote by $d_C(x, y)$ the usual Euclidean distance of a point (x, y) in the plane to C.

Calculate the integral

$$I = \int_{\mathbb{R}^2} e^{-d_C(x,y)} \, \mathrm{d}x\mathrm{d}y. \tag{1}$$

Hints. - Distinguish three zones of integration: when (x, y) lies in C; when (x, y) is orthogonally projected onto a vertex of C; when (x, y) is orthogonally projected onto a point lying on an edge of C.

- It helps to know that $\int_0^{+\infty} e^{-u^2} \, \mathrm{d}u = \frac{\sqrt{\pi}}{2}$.

- The result of the integration (1) has to do with the area of C and the perimeter of C (= length of its boundary).

Answer. We have:

$$I = 2\pi + \mathcal{A}(C) + \ell(C), \tag{2}$$

where $\mathcal{A}(C)$ and $\ell(C)$ denote the area and the perimeter of C, respectively.

Comment. A result in the same vein is

$$\int_{\mathbb{R}^2} e^{-d_C^2(x,y)}\, \mathrm{d}x\mathrm{d}y = \pi + \mathcal{A}(C) + \frac{\sqrt{\pi}}{2}\ell(C). \tag{3}$$

---

195. ★★★ *The cancellation law for compact convex sets*

Let $\mathcal{K}(\mathbb{R}^n)$ denote the collection of nonempty compact sets in $\mathbb{R}^n$. We denote by co$S$ the convex hull of a set S. Hence, if A belongs to $\mathcal{K}(\mathbb{R}^n)$, so does co$A$.

1°) Let A, B, C be taken in $\mathcal{K}(\mathbb{R}^n)$. Show that

$$(A + B = A + C) \Rightarrow (\mathrm{co}B = \mathrm{co}C). \tag{1}$$

2°) Let $B \in \mathcal{K}(\mathbb{R}^n)$. Show that

$$n(\mathrm{co}B) + B = n(\mathrm{co}B) + coB. \tag{2}$$

Check, with the help of a counterexample, that (2) does not necessarily hold true with, instead of n, an integer $k < n$.

3°) Deduce from the above the following: If B and C are in $\mathcal{K}(\mathbb{R}^n)$, then

$$\left(\begin{array}{c} \text{For all } A \in \mathcal{K}(\mathbb{R}^n), \\ (A + B = A + C) \Rightarrow (B = C) \end{array} \right) \Leftrightarrow (\text{Both } B \text{ and } C \text{ are convex}). \tag{3}$$

Hints. 2°) Use the following two facts: any $x \in \mathrm{co}(B)$ can be written as a convex combination of $n+1$ extreme points of co(B); any extreme point of co(B) necessarily belongs to B.

A counterexample could be $B = \{0, e_1, .., e_n\}$, where $e_1, .., e_n$ are the canonical basis elements in $\mathbb{R}^n$.

3°) [$\Rightarrow$] implication. Suppose, for example, that B is not convex. By choosing as a specific set $A = n\,\mathrm{co}(B)$, one sees with (2) that the cancellation law expressed in the left-hand side of (3) does not hold true.

Answer. 2°) counterexample. With the set B suggested in Hints, we have

$$\mathrm{co}B = \left\{ (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n : \sum_{i=1}^{n} \alpha_i \leqslant 1 \text{ and } \alpha_i \geqslant 0 \text{ for all } i \right\},$$

and $k\,(\mathrm{co}B) + B$ is strictly contained in $(k+1)\,\mathrm{co}B$ whenever $k < n$.

---

196. ★★★ *Differentiability of the convex hull function*

Given a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ bounded from below on $\mathbb{R}^n$ by some affine function, we denote by $\mathrm{co}f : \mathbb{R}^n \to \mathbb{R}$ the *convex hull function*, defined as follows:

$$\mathrm{co}f = \text{supremum of all affine functions } a \text{ such that } a \leqslant f \text{ on } \mathbb{R}^n.$$

We intend to answer the following question: is the function $\mathrm{co}f$ differentiable on $\mathbb{R}^n$?

1°) Case $n = 1$.

(a) Two examples. What is $\mathrm{co}f$ when $f(x) = \exp(-x^2)$? when $f(x) = \big(x^2 - 1\big)^2$?

(b) A general result. Prove that $\mathrm{co}f$ is differentiable on $\mathbb{R}$.

2°) Case $n \geqslant 2$.

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the $\mathcal{C}^\infty$ function defined by $f(x, y) = \sqrt{x^2 + \exp(-y^2)}$. What is $\mathrm{co}f$ for this f? Conclusion?

Answers. 1°) (a). First example: We have that $\mathrm{co}f$ is the null function on $\mathbb{R}$. In this example, f and $\mathrm{co}f$ never coincide, $(\mathrm{co}f)(x) < f(x)$ for all $x \in \mathbb{R}$.

Second example: We have that $\mathrm{co}f(x) = \Big[\big(x^2 - 1\big)^+\Big]^2$, $x \in \mathbb{R}$.

Nevertheless, $\mathrm{co}f$ is differentiable on $\mathbb{R}$ whenever f is differentiable on $\mathbb{R}$.

2°) Here, $(\mathrm{co}f)(x, y) = |x|$. Thus, for $n \geqslant 2$, the convex hull function of a differentiable (even $\mathcal{C}^\infty$) function on $\mathbb{R}^n$ is not necessarily differentiable on $\mathbb{R}^n$.

---

197. ★★★ *When is a point x satisfying $\nabla f(x) = 0$ a global minimizer of f?*

Given a function $f : \mathbb{R}^n \to \mathbb{R}$ bounded from below on $\mathbb{R}^n$ by some affine function, we denote by $\mathrm{co}f : \mathbb{R}^n \to \mathbb{R}$ the convex hull function, whose possible definition is as follows:

$$\mathrm{co}f = \text{supremum of all affine functions } a \text{ such that } a \leqslant f \text{ on } \mathbb{R}^n.$$

It is also the largest convex function g satisfying $g \leqslant f$.

We begin by considering only differentiable functions. Our objective is to decide when a critical (or stationary) point of f is a global minimizer of f.

1°) Let $\overline{x}$ be a global minimizer of f. Show that:

$$(i)\ \nabla f(\overline{x}) = 0; \tag{1}$$

$$(ii)\ (\mathrm{co}f)(\overline{x}) = f(\overline{x}). \tag{2}$$

2°) Converse of the previous statement. Let $\overline{x}$ be a critical point of f, *i.e.* satisfying $\nabla f(\overline{x}) = 0$. Assume, moreover, that $\mathrm{co}f$ and f coincide at $\overline{x}$, *i.e.* $(\mathrm{co}f)(\overline{x}) = f(\overline{x})$.

(a) Prove that $\mathrm{co}f$ is differentiable at $\overline{x}$ with $\nabla(\mathrm{co}f)(\overline{x}) = 0$.

(b) Check that $\overline{x}$ is indeed a global minimizer of f.

Hints. 1°) Proof of (ii). When $\overline{x}$ is a global minimizer of f, the constant function $x \mapsto f(\overline{x})$ is an affine function bounding f from below on $\mathbb{R}^n$.

2°) (a). The function $\mathrm{co}f$ is convex; it therefore admits directional derivatives at any point x and in all directions d:

$$(\mathrm{co}f)'(x,d) = \lim_{t>0\to 0} \frac{(\mathrm{co}f)(x+td) - (\mathrm{co}f)(x)}{t}.$$

Moreover, a positively homogeneous convex function bounded from above by a linear function is itself linear.

Comments. - As a general rule, differentiable functions f whose critical points are also global minimizers are those for which

$$\{x : \nabla f(x) = 0\} \subset \{x : f(x) = (\mathrm{co}f)(x)\}.$$

- The necessary and sufficient conditions for global minimality exhibited in this Tapa (combination of conditions (1) and (2)) no longer hold

for non-differentiable functions, when condition (2) is replaced, for example, by "$\overline{x}$ is a local minimizer of f" or by "$\overline{x}$ is a kink-point of f"; consider the following example:

$$f(x) \quad = \quad \begin{cases} |x| & \text{if } |x| \leqslant 1; \\ 2-|x| & \text{if } 1 \leqslant |x| \leqslant 3/2; \\ |x|-1 & \text{if } |x| \geqslant 3/2. \end{cases}$$

For more on this question, see

J.-B. Hiriart-Urruty, *When is a point x satisfying $\nabla f(x) = 0$ a global minimum of f?* American Math. Monthly 93 (1986), 556–558.

---

198. ★★★ *Differentiability of the distance function to a set vs convexity of this set*

Let $\mathbb{R}^n$, $n \geqslant 2$, be equipped with the usual inner product and the associated Euclidean norm $\|\cdot\|$. Let S be a nonempty closed set in $\mathbb{R}^n$; we denote by d_S the distance function to S and by $P_S(x)$, $x \in \mathbb{R}^n$, the nonempty set of orthogonal projections of x onto S :

$$P_S(x) = \{c \in S : \|x-c\| = d_S(x)\}\,.$$

When $P_S(x)$ is single-valued, we denote by $p_S(x)$ the unique element in $P_S(x)$ (which is certainly the case when $x \in S$, with $P_S(x) = \{x\}$).

1°) Differentiability of d_S *vs* that of d_S^2.

(a) Let x be outside the boundary of S. Check that d_S is differentiable at x if and only if d_S^2 is differentiable at x.

(b) Verify that:

$$(d_S \text{ is differentiable on } \mathbb{R}^n \setminus S) \Longleftrightarrow \left(d_S^2 \text{ is differentiable on } \mathbb{R}^n\right).$$

(c) Let x be on the boundary of S. Could d_S be differentiable at x?

2°) Differentiability of d_S at x *vs* single-valuedness of $P_S(x)$.

(a) Let $x \notin S$ be a point where d_S is differentiable. Prove that $P_S(x)$ contains only one element, say $P_S(x) = \{p_S(x)\}$, and:

$$\nabla d_S(x) = \frac{x - p_S(x)}{\|x - p_S(x)\|}. \tag{1}$$

(b) Let $x \notin S$ be a point with $P_S(x)$ single-valued, say $P_S(x) = \{p_S(x)\}$. Show that d_S is differentiable at x with (1) as an expression of $\nabla d_S(x)$.

3°) When convexity enters into the picture.

Prove the equivalence of the following statements:

$$\begin{array}{l}(i)\ d_S^2 \text{ is differentiable on } \mathbb{R}^n;\\ (ii)\ P_S(x) \text{ is single-valued for all } x \in \mathbb{R}^n;\\ (iii)\ S \text{ is convex.}\end{array} \tag{2}$$

Hints. 2°) (b). Prove that, for any neighborhood V of $p_S(x)$, one can find a neighborhood U of x such that:

$$P_S(x') \subset V \text{ for all } x' \text{ in } U.$$

3°) $[(iii) \Rightarrow (ii)]$. This is a classical result on orthogonal projection onto convex sets. Moreover, in this case, the mapping p_S is LIPSCHITZ on $\mathbb{R}^n$:

$$\|p_S(x) - p_S(x')\| \leqslant \|x - x'\| \text{ for all } x, x' \text{ in } \mathbb{R}^n.$$

Answers. 1°) (c). The answer is Yes. Consider, for example, in $\mathbb{R}^2$,

$$S = \overline{B(0,1)} \cup \{(x,y) : y \leqslant 0\},$$

and the point $(0,0)$. Then d_S is differentiable at $(0,0)$.

When this happens, $\nabla d_S(x)$ is the null vector (because d_S is minimized at this point).

Comments. - When S is convex, a consequence of the results proved in this Tapa is: d_S is differentiable on the whole of $\mathbb{R}^n$ if and only if $S = \mathbb{R}^n$.

- The equivalence between (ii) and (iii) in (2) (Question 3°)) is known as the BUNT–MOTZKIN theorem (1934–1935). The possible equivalence to a HILBERT space context is still an open question; for more on these questions, see:

A. POMMELLET, *Différentiabilité de la fonction distance à une partie.* Revue de Mathématiques Spéciales n°7 (1993), 377–382.

J.-B. HIRIART-URRUTY, *Ensembles de Tchebychev vs ensembles convexes : l'état de la situation vu via l'analyse convexe non lisse.* Ann. Sci. Math. Québec 22, n°1 (1998), 47–62.

- When S is convex, the set of points where d_S is not differentiable is exactly the boundary of S. It is then a set of LEBESGUE measure zero.

199. ★★★ *Convexity of the quotient of a quadratic form by a norm*

Let $M \in \mathcal{S}_n(\mathbb{R})$, let $\|\cdot\|$ denote the usual Euclidean norm on $\mathbb{R}^n$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be the positively homogeneous function defined as follows:

$$f(x) = \begin{cases} \frac{x^T M x}{\|x\|} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases} \tag{1}$$

Under what conditions on the matrix M is this function f convex on $\mathbb{R}^n$?

Hints. - Show that M has to be positive semidefinite.

- The required conditions involve extreme eigenvalues of M.

Answer. The function f defined in (1) is convex on $\mathbb{R}^n$ if and only if:

$$\lambda_{\max}(M) \leqslant 2\ \lambda_{\min}(M), \tag{2}$$

where $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$ denote, respectively, the largest and the smallest eigenvalues of M.

Comments. Note that, apart from the null matrix, all the matrices M satisfying (2) are positive definite. Moreover, if a non-null matrix M satisfies (2), so does its inverse M^{-1}.

---

200. ★★★ *Positive homogeneity, biconvexity and convexity*

1°) Consider a positively homogeneous (of degree 1) function $f : \mathbb{R}^n \to \mathbb{R}$, *i.e.,* satisfying:

$$f(tx) = tf(x) \text{ for all } x \in \mathbb{R}^n \text{ and all } t > 0. \tag{1}$$

(a) Check that f is convex if and only if it is subadditive, that is to say:

$$f(x + y) \leqslant f(x) + f(y) \text{ for all } x \text{ and } y \text{ in } \mathbb{R}^n. \tag{2}$$

(b) Show that f is continuous on $\mathbb{R}^n$ if and only if f is continuous at each point x on the unit sphere (*i.e.,* satisfying $\|x\| = 1$).

2°) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be *biconvex* (one also says *separately convex*), that is: $f(x, \cdot)$ is convex for all $x \in \mathbb{R}$, and $f(\cdot, y)$ is convex for all $y \in \mathbb{R}$.

(a) Prove that f is necessarily continuous on $\mathbb{R}^2$.

(b) Prove that if, in addition, f is positively homogeneous, then f is convex on $\mathbb{R}^2$.

Hints. 2°) (b). One may make use of the following results:

- A continuous function $f : \mathbb{R}^2 \to \mathbb{R}$ which is convex on each of the four open half-spaces $\mathbb{R} \times (0, +\infty), \mathbb{R} \times (-\infty, 0), (0, +\infty) \times \mathbb{R}$ and $(-\infty, 0) \times \mathbb{R}$ is necessarily convex on $\mathbb{R}^2$. To see why, show that f is convex along any line in $\mathbb{R}^2$.

- The convexity of $x \mapsto f(x, 1)$ induces that of $(x, y) \mapsto y f\left(\frac{x}{y}, 1\right)$ on $\mathbb{R} \times (0, +\infty)$ (the associated perspective function, see the Comment at the end of Tapa 163). Repeat the process with the functions $f(x, -1), f(1, y)$ and $f(-1, y)$.

Comments. - A related tapa involving positively homogeneous and convex functions is Tapa 70.

- The tricky results in 2°) are due to B. Dacorogna (1989). Following a more involved path, one could bypass (a) and show that $f : \mathbb{R}^2 \to \mathbb{R}$ is convex; hence f is continuous. The method suggested in the Hints is due to P. Maréchal (2007).

- It is somewhat surprising that the result in 2°) does not hold true if we accept functions taking the value $+\infty$. The definition of convexity for $f : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ is the same as for finite-valued functions, except that the inequality of definition $(f[\lambda x + (1-\lambda)y] \leqslant \lambda f(x) + (1-\lambda) f(y))$ is considered in $\mathbb{R} \cup \{+\infty\}$. Likewise for the definition of positive homogeneity (equality (1) is taken in $\mathbb{R} \cup \{+\infty\}$). For example, $f : (x, y) \mapsto f(x, y) = 0$ if $xy \geqslant 0$, $+\infty$ if not, is biconvex, positively homogeneous, but not convex on $\mathbb{R}^2$.

- The result in 2°) does not hold in higher dimensions ($n \geqslant 3$).

---

201. ★★★ *When 0 is the sole zero of two quadratic forms*

Consider an integer $n \geqslant 3$ and $\mathbb{R}^n$ equipped with the usual Euclidean structure. Let A and B be two real symmetric $n \times n$ matrices.

1°) Prove that the following statements are equivalent:

$$\left(\begin{array}{c} x^T A x = 0 \\ \text{and} \\ x^T B x = 0 \end{array} \right) \Rightarrow (x = 0)\,; \tag{1}$$

$$\text{There exists } \mu_1 \text{ and } \mu_2 \in \mathbb{R} \text{ such that } \mu_1 A + \mu_2 B \succ 0. \tag{2}$$

2°) Verify with the help of a counterexample that the equivalence above does not hold if $n = 2$.

Hints. 1°) - According to BRICKMAN's theorem, when $n \geqslant 3$,

$$K = \left\{ \left(x^T Ax, x^T Bx \right) : \|x\| = 1 \right\} \tag{3}$$

is a compact convex set in $\mathbb{R}^2$.

- The so-called "separation theorem" (of a point from a compact convex set) asserts that if a point $(a, b) \in \mathbb{R}^2$ does not belong to K, there is a line strictly separating (a, b) from K.

2°) Try with

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{4}$$

Answers. 1°) When $C \in \mathcal{S}_n(\mathbb{R})$ is positive definite, $x^T Cx = 0$ only for $x = 0$; thus, the implication [(2) $\Rightarrow$ (1)] is immediate.

Converse implication [(1) $\Rightarrow$ (2)]. We have that the set K is convex (a result due to L. BRICKMAN (1961)); what (1) says is that $(0, 0) \notin K$; we can therefore "separate" $(0, 0)$ from K: there exists a line strictly separating $(0, 0)$ from K, *i.e.,* there exists $(\mu_1, \mu_2) \in \mathbb{R}^2$ and $r \in \mathbb{R}$ such that

$$\mu_1 u + \mu_2 v > r > 0 \text{ for all } (u, v) \in K. \tag{5}$$

Now, (5) is nothing else than

$$\mu_1 x^T Ax + \mu_2 x^T Bx > 0 \text{ for all } x \in \mathbb{R}^n, \|x\| = 1,$$

that is (2).

2°) With the proposed matrices A and B in $\mathcal{S}_2(\mathbb{R})$ in (4) (see Hints): (1) holds true but there is no way to have

$$\mu_1 A + \mu_2 B = \begin{bmatrix} \mu_1 & \mu_2 \\ \mu_2 & -\mu_1 \end{bmatrix} \succ 0.$$

Comments. - As already said, the convexity result of K, for $n \geqslant 3$, is due to L. BRICKMAN (1961). There is a result of the same vein, which is due to L. DINES (1941): the set

$$\mathcal{K} = \left\{ \left(x^T Ax, x^T Bx \right) : \|x\| \in \mathbb{R}^n \right\}$$

is a convex cone in $\mathbb{R}^2$. When the implication (1) holds, $\mathcal{K}$ is closed... but it could be the whole $\mathbb{R}^2$; that is what happened in the counterexample in 2°); in that case, there is no way to "separate" $(0,0)$ from $\mathcal{K}$.

- The result proposed in this Tapa was proved by P. Finsler (1936–37) and rediscovered by E. Calabi (1964).

---

202. ★★★ *When is the maximum of two quadratic forms nonnegative?*

Let A and B be two real symmetric $n \times n$ matrices. Prove that the following statements are equivalent:

$$\begin{cases} \max\left(x^T Ax, x^T Bx\right) \geqslant 0 \text{ for all } x \text{ in } \mathbb{R}^n \\ (\text{resp.} > 0 \text{ for all } x \neq 0 \text{ in } \mathbb{R}^n); \end{cases} \tag{1}$$

$$\begin{cases} \text{There exists } \mu_1 \geqslant 0 \text{ and } \mu_2 \geqslant 0,\ \mu_1 + \ \mu_1 = 1, \text{ such that} \\ \mu_1 A + \mu_2 B \succcurlyeq 0 \text{ (resp. } \mu_1 A + \mu_2 B \succ 0). \end{cases} \tag{2}$$

Hints. - According to Dines' theorem,

$$\mathcal{K} = \left\{\left(x^T Ax, x^T Bx\right) : x \in \mathbb{R}^n\right\} \tag{3}$$

is a convex cone in $\mathbb{R}^2$.

- A version of the "separation theorem" (of two disjoint convex cones) asserts that if $\mathcal{K}$ does not meet the open convex cone $\mathcal{N} = (-\infty, 0) \times (-\infty, 0)$, then there is a line passing through the origin just separating them.

Answer. Implication [(1) ⇒ (2)] (first case). The image set $\mathcal{K}$ is a convex cone (a result due to L. Dines (1941)); what (1) says is that $\mathcal{K}$ and the open convex cone $\mathcal{N} = (-\infty, 0) \times (-\infty, 0)$ are disjoint; we can therefore "separate" them: there exists a line passing through the origin just separating them, *i.e.,* there exists a non-null $(\mu_1, \mu_2) \in \mathbb{R}^2$ for which:

$$\mu_1 u + \mu_2 v \geqslant 0 \text{ for all } (u, v) \in \mathcal{K}, \tag{4}$$

$$\mu_1 \alpha + \mu_2 \beta \geqslant 0 \text{ for all } (\alpha, \beta) \in \mathcal{N}. \tag{5}$$

Then, (4) yields that $\mu_1 A + \mu_2 B$ is positive semidefinite, while (5) ensures that $\mu_1 \geqslant 0$ and $\mu_2 \geqslant 0$.

Comment. The result proposed in this Tapa is due to Y. YUAN (1990). In some sense, it is the "unilateral" version of the equivalence proposed in the previous Tapa 201.

---

203. ★★★ *On differentiable functions with* LIPSCHITZ *gradients*

Let $(H, \langle\cdot,\cdot\rangle)$ be a HILBERT space; we denote by $\|\cdot\|$ the norm associated with $\langle\cdot,\cdot\rangle$. Let $f : H \to \mathbb{R}$ be a differentiable function and let $\alpha > 0$. Prove that the next two statements are equivalent:

$$|\langle \nabla f(x) - \nabla f(y), x - y\rangle| \leqslant L\,\|x-y\|^2 \text{ for all } x, y \text{ in } H; \quad (1)$$
$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\,\|x-y\| \text{ for all } x, y \text{ in } H. \quad (2)$$

Hints. - [(2) ⇒ (1)] is just an application of the CAUCHY–SCHWARZ inequality.

- For [(1) ⇒ (2)], it is helpful to observe that both

$$\frac{L}{2}\,\|\cdot\|^2 - f \text{ and } \frac{L}{2}\,\|\cdot\|^2 + f$$

are convex functions on H.

Comments. - The GATEAUX-differentiability of f on H would be sufficient, *i.e.*: For all $x \in H$,

$$\lim_{t\to 0} \frac{f(x+td) - f(x)}{t} \text{ exists and is a continuous linear form of } d$$
$$(d \mapsto \lim_{t\to 0} \frac{f(x+td) - f(x)}{t} = \langle \nabla f(x), d\rangle).$$

- Although it was clear and known for twice continuously differentiable functions f, the equivalence between (1) and (2) is rather surprising... Clearly (1), which involves f on line-segments, is easier to check.

- For related results, *cf.* Tapas 157 and 158.

---

204. ★★★ *On convex functions summing up to* $\|\cdot\|^2$

Let $(H, \langle\cdot,\cdot\rangle)$ be a HILBERT space; we denote by $\|\cdot\|$ the norm derived from $\langle\cdot,\cdot\rangle$. Let $\alpha > 0$ and let $f, g : H \to \mathbb{R}$ be two convex functions on H satisfying:

$$f(x) + g(x) = \alpha\,\|x\|^2 \text{ for all } x \in H. \quad (1)$$

1°) Show that both f and g are GATEAUX-differentiable on H.

2°) Prove that:

$$\begin{array}{c}\langle \nabla f(x) - \nabla f(y), \nabla g(x) - \nabla g(y)\rangle \geqslant 0 \text{ for all } x, y \text{ in } H; \\ \nabla f \text{ and } \nabla g \text{ are } 2\alpha - \text{LIPSCHITZ mappings on } H.\end{array} \tag{2}$$

Hint. 1°) Since both f and g are convex, they admit directional derivatives $f'(x,\cdot)$ and $g'(x,\cdot)$ at all $x \in H$. These directional derivatives are positively homogeneous functions on H. According to (1), their sum $f'(x,\cdot) + g'(x,\cdot)$ is the continuous linear form $2\alpha\,\langle x, \cdot\rangle$, which does not leave much room for $f'(x,\cdot)$ and $g'(x,\cdot)$...

Comments. - 2°) Although $\nabla f(x) + \nabla g(x) = 2\alpha x$ for all $x \in H$, one cannot do better than 2α for LIPSCHITZ constants of ∇f and ∇g.

- There obviously are connections between the results in this Tapa and those in the previous one (Tapa 203).

---

205. ★★★ MOREAU*'s decomposition following a closed convex cone and its polar*

Let $(H, \langle\cdot,\cdot\rangle)$ be a HILBERT space and let K be a closed convex cone with apex 0 in H. We denote by K° the so-called (negative) polar cone of K; it is defined as follows:

$$K^\circ = \{y \in H : \langle y, x\rangle \leqslant 0 \text{ for all } x \in K\}\,.$$

K° is a closed convex cone with apex 0 in H, and it turns out that $(K^\circ)^\circ = K$.

1°) Characterization of the projection $p_K(x)$ of $x \in H$ onto K.

Prove that:

$$(\overline{x} = p_K(x)) \Leftrightarrow \left(\begin{array}{c}\overline{x} \in K,\ x - \overline{x} \in K^\circ \\ \text{and } \langle x - \overline{x}, \overline{x}\rangle = 0\end{array}\right). \tag{1}$$

2°) Prove that the following two assertions (concerning $x \in H$) are equivalent:

$$(i)\ \ x = x_1 + x_2, \text{ with } x_1 \in K,\ x_2 \in K^\circ,\ \langle x_1, x_2\rangle = 0; \tag{2-1}$$
$$(ii)\ \ x_1 = p_K(x) \text{ and } \ x_1 = p_{K^\circ}(x). \tag{2-2}$$

This result is known as MOREAU's decomposition of x.

3°) "Optimality" of the decomposition in (2-1)–(2-2).

Let $x \in H$ be decomposed as

$$x = x_1 + x_2, \text{ with } x_1 \in K,\ x_2 \in K^\circ. \tag{3}$$

Show that necessarily

$$\|x_1\| \geqslant \|p_K(x)\| \text{ and } \|x_2\| \geqslant \|p_{K^\circ}(x)\|\,. \tag{4}$$

4°) A first illustration. Let $H = \mathcal{S}_n(\mathbb{R})$ be structured as a Euclidean space with the help of the scalar product $\langle\langle A, B\rangle\rangle = \operatorname{tr}(AB)$. Let us consider as K the closed convex cone of positive semidefinite matrices

$$K = \{A \in \mathcal{S}_n(\mathbb{R}) : A \succcurlyeq 0\}\,.$$

(*a*) Show that K° is $-K$, *i.e.* the closed convex cone of negative semidefinite matrices.

(*b*) Find Moreau's decomposition of $A \in \mathcal{S}_n(\mathbb{R})$, hence the projection of A onto K and K°.

5°) Another illustration, in the infinite-dimensional context this time. Let I be the real interval (a, b), let H be the Lebesgue space $L^2(I, \mathbb{R})$ (denoted simply by $L^2(I)$) equipped with the scalar product $\langle f, g\rangle_{L^2} = \int_I f(t)g(t)\ \mathrm{d}\lambda(t)$.

We denote by K the set of functions $g \in H$ of the form $g = f'$ with $f : I \to \mathbb{R}$ a convex function (K could be called "the set of increasing functions in $L^2(I)$").

(*a*) Check that K is a closed convex cone in H.

Let S denote the set of functions $h : I \to I$ which preserve the Lebesgue measure (in the sense that the image of the Lebesgue measure λ under h is λ itself). Obviously, the identity function id_I on I belongs to S. We do not need to know more about S, we just admit the following result:

$$(g \in K) \Leftrightarrow (\langle g, h - \mathrm{id}_I\rangle_{L^2} \leqslant 0 \text{ for all } h \in S)\,. \tag{5}$$

(*b*) Deduce from the above that K° is the closure (in $L^2(I)$) of the convex cone

$$L = \left\{ \sum_{i=1}^{k} \alpha_i (h_i - \mathrm{id}_I) : k \text{ positive integer, } \alpha_i \geqslant 0 \text{ and } h_i \in S \text{ for all } i \right\}. \tag{6}$$

(*c*) Let $\varphi \in L^2(I)$. Establish that there exist a convex function $f : I \to \mathbb{R}$ (unique within an additive constant), $h \in K^\circ$, such that:

$$\varphi = f' + h \text{ and } \langle f', h\rangle_{L^2} = 0. \tag{7}$$

Hints. 1°) Adapt to the convex conical context K the characterization of the projection $p_C(x)$ of $x \in H$ onto a closed convex set $C \subset H$:

$$(\overline{x} = p_C(x)) \Leftrightarrow (\overline{x} \in C \text{ and } \langle x - \overline{x}, c - \overline{x} \rangle \leqslant 0 \text{ for all } c \in C)\,.$$

The key-observation is: If $x \in K$, then $\alpha x \in K$ for all $\alpha \geqslant 0$.

2°) Relations (2-1) and (2-2) are geometrically very expressive; some drawings in the plane will help to grasp and understand them.

4°) Use a spectral decomposition of $A \in \mathcal{S}_n(\mathbb{R})$: there exists an orthogonal matrix U such that

$$U^T A U = \mathrm{diag}(\lambda_1, .., \lambda_n)$$

(the λ_i' s are the eigenvalues of A, all real numbers).

Next, observe that

$$\mathbb{R}^n \ni (\lambda_1, .., \lambda_n) = (\lambda_1^+, .., \lambda_n^+) + (\lambda_1^-, .., \lambda_n^-),$$

where $\lambda_i^+ = \max(\lambda_i, 0)$ and $\lambda_i^- = \min(\lambda_i, 0)$ is MOREAU's decomposition along the cone $(\mathbb{R}^n)^+$ and its polar $(\mathbb{R}^n)^- = -(\mathbb{R}^n)^+$.

5°) (b). By definition, L is the convex cone generated by the set $S - \mathrm{id}_I$.

Answers. 4°) (b). We have

$$\begin{aligned} A_1 &= p_K(A) = U \mathrm{diag}(\lambda_1^+, .., \lambda_n^+) U^T, \\ A_2 &= p_{K^\circ}(A) = U \mathrm{diag}(\lambda_1^-, .., \lambda_n^-) U^T, \end{aligned}$$

where $\lambda_i^+ = \max(\lambda_i, 0)$ and $\lambda_i^- = \min(\lambda_i, 0)$.

5°) (b). This easily follows from the admitted results (6) and $(K^\circ)^\circ = K$.

Comments. - 2°) This result is due to J.J. MOREAU (1965). It generalizes what is known when K is a closed vector space V in H. In that case, K° is the orthogonal space $V^\perp$ and

$$x = p_V(x) + p_{V^\perp}(x), \quad \langle p_V(x), p_{V^\perp}(x) \rangle = 0.$$

- 3°) But, in contrast to the "linear case" (K is a closed vector space V in H), there may be several decompositions like (3); the result in (4) says that the one provided by MOREAU's decomposition is the "best".

206. ★★★ *Directional derivative of the projection mapping onto a closed convex set in a* HILBERT *space.*

Let C be a closed subset in a HILBERT space $(H, \langle \cdot, \cdot \rangle)$. We denote by p_C the projection mapping onto C. We intend to prove that, at least when $x \in C$, this projection mapping has a directional derivative $p'_C(x, d)$ at x in every direction $d \in H$, and that this directional derivative can be expressed as the projection of d onto a closed convex cone associated with C and x.

Let $x \in C$. We denote by $T_C(x)$ the so-called *tangent cone* to C at x; it can be defined as follows:

$$d \in T_C(x) \Leftrightarrow \left(\begin{array}{l} \text{There exist sequences } t_n > 0 \to 0,\ (d_n) \to d \\ \text{such that } x + t_n d_n \in C \text{ for all } n \end{array} \right). \quad (1)$$

$T_C(x)$ turns out to be a closed convex cone with apex 0.

Prove that:

$$p'_C(x, d) = \lim_{t > 0 \to 0} \frac{p_C(x + td) - p_C(x)}{t}$$
$$\text{exists and equals } p_{T_C(x)}(d). \quad (2)$$

(Here, $p_{T_C(x)}(d)$ denotes the projection of the direction d onto the closed convex cone $T_C(x)$).

Hints. - Due to the convexity of C, an alternative definition of $T_C(x)$ is as follows: $T_C(x)$ is the closure of the cone generated by $C - x$, *i.e.,*

$$T_C(x) = \overline{\{\lambda(c - x) : \lambda \geqslant 0 \text{ and } c \in C\}}. \quad (3)$$

- A definition which goes with $T_C(x)$, $x \in C$, is that of its polar cone $[T_C(x)]^\circ = N_C(x)$, called the *normal cone* to C at x, and whose alternative definition, derived from (3), is as follows:

$$(v \in N_C(x)) \Leftrightarrow (\langle v, c - x \rangle \leqslant 0 \text{ for all } c \in C). \quad (4)$$

- The formula (2) expressing $p'_C(x, d)$ as $p_{T_C(x)}(d)$ can easily be illustrated by drawing some pictures in the plane.

- The characterizations of $p_C(x)$ and $p_{T_C(x)}(d)$ as solutions of variational inequalities, as well as MOREAU's decomposition (*cf.* the previous Tapa 205), are used constantly in the proofs.

- The main difficulty here comes from the infinite-dimensional setting of the problem: the sequences exhibited are (easily) made weakly convergent, but their strong convergence is required. Start therefore with a

sequence (u_n) aimed at converging to the directional derivative $p'_C(x,d)$ of p_C at x in the d direction:

$$u_n = \frac{p_C(x+t_n d) - x}{t_n}, \text{ with } t_n > 0. \tag{5}$$

Answers (partial). Three points of proof are in order:

- The sequence (u_n) defined in (5) (*cf.* Hints) is bounded by $\|d\|$; therefore there exists a subsequence $(u_{n_k})_k$ of $(u_n)_n$ which converges weakly to some $u \in H$ when $k \to +\infty$.

- The found limit u is indeed the projection of d onto the closed convex cone $T_C(x)$.

- The whole sequence $(u_n)_n$ converges strongly towards u when $n \to +\infty$.

Comments. - The result of this Tapa 206 is important in Solid mechanics (contact problems, friction problems, etc.); at each point x of the solid C, a force or a velocity can be decomposed (in MOREAU's sense) into a tangential component and a normal component.

- The point of reference x has been chosen in C, so that $p_C(x) = x$. Curiously enough, when x lies outside C, the directional derivative $\lim_{t>0\to 0} [p_C(x+td) - p_C(x)]/t = p'_C(x,d)$ may fail to exist; there are counterexamples even in $H = \mathbb{R}^3$.

---

207. ★★★ VON NEUMANN's *algorithm of alternating projections onto two closed vector subspaces in a* HILBERT *space*

Let V_1 and V_2 be two closed vector subspaces in the Hilbert space $(H, \langle\cdot,\cdot\rangle)$. Given an arbitrary $x \in H$, one designs a sequence $(x_n)_n$ in H by projecting alternatively onto V_1 and onto V_2 :

$$\begin{cases} \text{Initial point: } x_0 = x; \\ \text{Iteration: For all } n \geqslant 1,\ x_{2n-1} = p_{V_1}(x_{2n-2}) \\ \text{and } x_{2n} = p_{V_2}(x_{2n-1}). \end{cases} \tag{1}$$

Prove that the sequence (x_n) defined above in (1) converges towards the projection of x onto $V_1 \cap V_2$.

Hints. - The algorithm (1) is easily illustrated by drawing some pictures in the plane.

- The main difficulty comes from the infinite-dimensional setting of the problem: the sequences exhibited are made weakly convergent, but their strong convergence is required. A helpful idea could be to firstly prove the proposed result in a finite-dimensional space H.

- Since V_1 (or V_2) is closed and H is complete, V_1 (or V_2) is complete; therefore any CAUCHY sequence in V_1 (resp. in V_2) is convergent to an element in V_1 (resp. in V_2).

Answers (partial). Four points of proof are in order:

- The sequence of norms $(\|x_n\|)_n$ is decreasing.

- The subsequence $(x_{2n})_n$ is CAUCHY in V_2.

- The whole sequence $(x_n)_n$ converges strongly towards an element $y \in V_1 \cap V_2$.

- The exhibited limit point y is indeed the projection of x onto $V_1 \cap V_2$.

Comment. The algorithm of alternating projections onto two closed convex sets C_1 and C_2, defined as in (1), does produce a sequence converging towards *some* point in $C_1 \cap C_2$, but not necessarily towards *the projection* of x onto $C_1 \cap C_2$. Counterexamples exist even in $H = \mathbb{R}^2$ with C_1 a vector subspace and C_2 a vector half-space.

---

208. ★★★ HERMITE *polynomials and the* FOURIER *transform in* $L^2(\mathbb{R})$

The n-th HERMITE function is the product of the n-th HERMITE polynomial (after some change of scale in the variable) by the function $e^{-\pi x^2}$. As we shall see, the HERMITE functions enjoy two remarkable properties:

- They form an orthogonal basis for $L^2(\mathbb{R})$;

- They are eigenfunctions for the FOURIER transform in $L^2(\mathbb{R})$.

($L^2(\mathbb{R})$ is equipped with the usual inner product $\langle f, g\rangle = \int_{\mathbb{R}} f(x)g(x)\mathrm{d}x$; the associated norm is just denoted $\|\cdot\|$; recall that $(L^2(\mathbb{R}), \langle\cdot,\cdot\rangle)$ is a HILBERT space).

In fact, the HERMITE functions allow an elegant alternative to the classical approaches to defining and studying the FOURIER transform in $L^2(\mathbb{R})$.

1°) The HERMITE polynomials.

For a nonnegative integer n, define the polynomial H_n as follows:

$$H_n(x) = (-1)^n\, e^{x^2/2} \frac{\mathrm{d}^n}{\mathrm{d}x^n}\left(e^{-x^2/2}\right). \qquad (1)$$

(a) Prove the following recursion formula:

$$\text{For all } n \geqslant 1,\ H_n(x) = x\ H_{n-1}(x) - H'_{n-1}(x). \tag{2}$$

(b) Using this recursion formula, determine, for example, H_6.

2°) The HERMITE functions.

For a nonnegative integer n, define the HERMITE function h_n as follows:

$$\begin{aligned} h_n(x) &= e^{\pi x^2}\left[\frac{\mathrm{d}^n}{\mathrm{d}x^n}(e^{-2\pi x^2})\right] &(3\text{-}1)\\ &= \left(-2\sqrt{\pi}\right)^n e^{-\pi x^2} H_n(2\sqrt{\pi}x). &(3\text{-}2)\end{aligned}$$

Thus, h_n is obtained by multiplying the n-th HERMITE polynomial H_n (after some dilation in the variable x) by a function of Gaussian type.

(a) What is the function h_2, for example?

(b) Show that h_n and h_{n-1} satisfy the following differential relation:

$$\text{For all } n \geqslant 1,\ h_n(x) = h'_{n-1}(x) - 2\pi x\ h_{n-1}(x). \tag{4}$$

3°) The FOURIER transforms of HERMITE functions.

Recall that the FOURIER transform $\widehat{f}$ of a function $f \in L^1(\mathbb{R})$ is defined as follows:

$$\xi \mapsto \widehat{f}(\xi) = \int_{\mathbb{R}} e^{-2i\pi x\xi} f(x)\ \mathrm{d}x.$$

(a) Prove the following fairly simple and elegant formula for the FOURIER transforms of the h_n's:

$$\text{For all } n,\ \widehat{h_n} = (-i)^n\, h_n. \tag{5}$$

(b) Show that the proposed HERMITE functions form an orthogonal basis of the HILBERT space $L^2(\mathbb{R})$.

4°) The FOURIER transform in $L^2(\mathbb{R})$.

The family $\left(e_n = \frac{h_n}{\|h_n\|}\right)_n$ is an orthonormal basis of $L^2(\mathbb{R})$. Let us define

$\mathcal{H}_0$ the subspace of $L^2(\mathbb{R})$ whose orthonormal basis is $(e_0, e_4, e_8, ...)$;
$\mathcal{H}_1$ the subspace of $L^2(\mathbb{R})$ whose orthonormal basis is $(e_1, e_5, e_9, ...)$;
$\mathcal{H}_2$ the subspace of $L^2(\mathbb{R})$ whose orthonormal basis is $(e_2, e_6, e_{10}, ...)$;
$\mathcal{H}_3$ the subspace of $L^2(\mathbb{R})$ whose orthonormal basis is $(e_3, e_7, e_{11}, ...)$.

Thus, $L^2(\mathbb{R}) = \mathcal{H}_0 \oplus \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \mathcal{H}_3$. Now, with $f \in L^2(\mathbb{R})$ uniquely decomposed on these four subspaces as $f = f_0 + f_1 + f_2 + f_3$, one follows (5) to define the FOURIER transform $\mathcal{F}f$ of f as follows:

$$\mathcal{F}f = f_0 - if_1 - f_2 + if_3. \tag{6}$$

(a) Check that $\mathcal{F}f$ coincides with $\widehat{f}$ whenever $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.

(b) Verify that the FOURIER transform $\mathcal{F}$ defines an isometry in $\left(L^2(\mathbb{R}), \|\cdot\|\right)$.

5°) Beyond $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$: the FOURIER transform on $L^p(\mathbb{R})$, $1 \leqslant p \leqslant 2$.

(a) Show that the FOURIER transform is defined without any ambiguity on $L^1(\mathbb{R}) + L^2(\mathbb{R})$.

(b) - Prove that all the $L^p(\mathbb{R})$, $1 \leqslant p \leqslant 2$, are contained in $L^1(\mathbb{R}) + L^2(\mathbb{R})$.

- Propose therefore a definition of the FOURIER transform on $L^p(\mathbb{R})$ when $1 \leqslant p \leqslant 2$.

Hints. 3°) (a). Proof by induction, for example. Recall two classical properties of FOURIER transforms:

- If $f \in \mathcal{C}^1(\mathbb{R}) \cap L^1(\mathbb{R})$ and $f' \in L^1(\mathbb{R})$, then:

$$\widehat{f'} = 2i\pi\widehat{f}.$$

- If $f, xf, ..., x^n f \in L^1(\mathbb{R})$, then $\widehat{f} \in \mathcal{C}^n(\mathbb{R})$ and:

$$\text{For } k = 0, ..., n-1, \ \left(\widehat{f}\right)^{(k)}(\xi) = (-2i\pi)^k \, \widehat{x^k f}\ (\xi).$$

(b) As already recalled, $L^2(\mathbb{R})$ is (a HILBERT space) equipped with the usual scalar product $\langle f, g\rangle = \int_{\mathbb{R}} f(x)g(x)\ \mathrm{d}x$ and the derived norm $\|f\| = \sqrt{\langle f, f\rangle}$.

- Firstly, show that h_n and h_m, $m \neq n$, are orthogonal in $L^2(\mathbb{R})$.

- Secondly, prove that the only function in $L^2(\mathbb{R})$ which is orthogonal to all the HERMITE functions h_n is the zero function.

Answers. 1°) (b). We have:

$$H_6(x) = x^6 - 15x^4 + 45x^2 - 15.$$

2°) (a). We have:

$$h_2(x) = 4\pi(4\pi x^2 - 1)e^{-\pi x^2}.$$

Comments. - This "Hilbertian" way of defining the FOURIER transform on $L^2(\mathbb{R})$ is due to N. WIENER (1933). The h_n's are eigenfunctions for the four possible eigenvalues of the FOURIER transform, namely $1, -i, -1, i$. On each eigenspace $\mathcal{H}_k$, $k = 0, 1, 2, 3$, the FOURIER transform acts like a rotation by an angle $0, -\pi/2, \pi$ and $\pi/2$, respectively.

- A consequence of the results in 3°) (b) is the possibility of developing $f \in L^2(\mathbb{R})$ in the orthonormal basis $\left(e_n = \frac{h_n}{\|h_n\|}\right)_n$:

$$f = \sum_{n=0}^{+\infty} \frac{\langle f, h_n \rangle}{\|h_n\|^2} h_n.$$

- For a little more on the subject of this Tapa 208, we refer to the note:

J.-B. HIRIART-URRUTY and M. PRADEL,
La transformation de FOURIER *de fonctions: au-delà de* L^1 *et* L^2. Revue de la filière Mathématiques (ex-Revue de Mathématiques Spéciales), n°2 (2002-2003), 25–31.

---

209. ★★★ *On the power of* PLANCHEREL's *theorem on the* FOURIER *transform*

For functions in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, the FOURIER transform $\mathcal{F}$ acts nicely:

$$\|f\| = \|\mathcal{F}f\|\,,\ \langle f, g\rangle = \langle \mathcal{F}f, \mathcal{F}g\rangle\,,\ \mathcal{F}(f * g) = \mathcal{F}f \times \mathcal{F}g, \qquad (1)$$

where $\langle \cdot, \cdot \rangle$ denotes the usual scalar product in $L^2(\mathbb{R})$ (and $\|\cdot\|$ the associated norm), and $*$ is the convolution operation $((\varphi * \psi)(x) = \int_{\mathbb{R}} \varphi(x-y)\psi(y)\ \mathrm{d}y)$.

1°) Let us consider the following two classes of functions:

$$f_a(x) \quad = \quad 1 \text{ if } \ x \in \left[-\frac{a}{2}, \frac{a}{2}\right], \text{ 0 if not (with } a > 0); \qquad (2)$$

$$g_b(x) \quad = \quad 1 - \frac{2}{b}\,|x| \ \text{ if } x \in \left[-\frac{b}{2}, \frac{b}{2}\right], \text{ 0 if not (with } b > 0). \qquad (3)$$

Calculate the FOURIER transforms of f_a and g_b.

2°) Using PLANCHEREL's theorem, determine the following integrals:

$$\int_{\mathbb{R}} \left(\frac{\sin\xi}{\xi}\right)^2 \mathrm{d}\xi, \ \int_{\mathbb{R}} \left(\frac{\sin\xi}{\xi}\right)^4 \mathrm{d}\xi, \ \int_{\mathbb{R}} \left(\frac{\sin\xi}{\xi}\right)^3 \mathrm{d}\xi.$$

Hints. 1°) After the calculation of $\mathcal{F}f_a$, in order to calculate $\mathcal{F}g_b$, one can either use the definition itself $(\mathcal{F}g_b(\xi) = \int_{\mathbb{R}} e^{-2i\pi x\xi} g_b(x)\ \mathrm{d}x)$, or observe that $f_a * f_a$ is related to g_{2a} and apply the third formula in (1).

2°) Use the above calculated FOURIER transforms and make use of (1).

Answers. 1°) We have:

$$\mathcal{F}f_a(\xi) = \frac{\sin(\pi a \xi)}{\pi \xi}. \tag{4}$$

We implicitly mean that $\mathcal{F}f_a(0) = 1$.

For g_b, we obtain:

$$\mathcal{F}g_b(\xi) = 2\ \frac{\sin^2(\frac{\pi b\ \xi}{2})}{\pi^2 b\ \xi^2}. \tag{5}$$

2°) By applying the first two formulas in (1) to functions chosen in the previous examples, we obtain:

$$\int_{\mathbb{R}} \left(\frac{\sin \xi}{\xi}\right)^2 \ \mathrm{d}\xi \quad = \quad \pi, \tag{6}$$

$$\int_{\mathbb{R}} \left(\frac{\sin \xi}{\xi}\right)^4 \ \mathrm{d}\xi \quad = \quad \frac{2\pi}{3}, \tag{7}$$

$$\int_{\mathbb{R}} \left(\frac{\sin \xi}{\xi}\right)^3 \ \mathrm{d}\xi \quad = \quad \frac{3\pi}{4}. \tag{8}$$

Comments. - The function $\mathcal{F}f_1 : x \mapsto \mathcal{F}f_1(x) = \frac{\sin(\pi\xi)}{\pi\xi}$ is the so-called (normalized) *cardinal sine function*, frequently appearing in Signal theory. Although $\mathcal{F}f_1 \notin L^1(\mathbb{R})$, it is possible to define an improper integral $\int_{-\infty}^{+\infty} \frac{\sin \xi}{\xi}\ \mathrm{d}\xi$ and, surprisingly enough,

$$\int_{-\infty}^{+\infty} \frac{\sin \xi}{\xi}\ \mathrm{d}\xi = \pi, \text{ like in (6).} \tag{9}$$

- A sort of "discrete version" of the results (6) and (9) is:

$$\sum_{n=1}^{+\infty} \frac{\sin(n)}{n} = \sum_{n=1}^{+\infty} \frac{\sin^2(n)}{n^2} = \frac{\pi - 1}{2}. \tag{10}$$

As expected, this is proved with the help of FOURIER series (development of the 2π-periodic function f defined as: $f(x) = 1$ if $x \in (0, 2]$; $f(x) = 0$ if $x \in (2, 2\pi]$).

210. ★★★ *The diffusion semigroup as a solution to the heat equation*

For all $t > 0$, define the function $p_t : \mathbb{R} \to \mathbb{R}$ by $p_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$.

1°) Verify that p_t is a $\mathcal{C}^\infty$ function all of whose derivatives are rapidly decreasing, *i.e.*,

$$\lim_{|x| \to +\infty} x^m \left(p_t\right)^{(n)}(x) = 0 \text{ for all nonnegative integers } m, n.$$

2°) Semi-group property. Prove that:

$$p_{t+s} = p_t * p_s \text{ for all } t > 0 \text{ and } s > 0. \tag{1}$$

Denote by $\mathcal{P}_t$ the mapping from $L^1(\mathbb{R})$ into itself defined by:

$$\mathcal{P}_t(f) = p_t * f.$$

3°) Check that:

$$\mathcal{P}_t \circ \mathcal{P}_s = \mathcal{P}_{t+s} \text{ for all } t > 0 \text{ and } s > 0. \tag{2}$$

4°) (*a*) Let $f = 1_{[0,1]}$. Show that $[\mathcal{P}_t(f)](0)$ does not tend to $f(0)$ as $t \to 0$.

(*b*) Let $f \in L^1(\mathbb{R})$ be continuous and bounded on $\mathbb{R}$. Prove that $\mathcal{P}_t(f)$ converges pointwise to f on $\mathbb{R}$ as $t \to 0$.

5°) Let $f \in L^1(\mathbb{R})$. Prove that the function $\varphi : (0, +\infty) \times \mathbb{R} \to \mathbb{R}$ defined by $\varphi(t, x) = [\mathcal{P}_t(f)](x)$ is a $\mathcal{C}^\infty$ function which solves the following partial differential equation

$$\frac{\partial \varphi}{\partial t} = \frac{1}{2} \frac{\partial^2 \varphi}{\partial x^2} \quad \text{(one-dimensional heat equation)}.$$

Hints. 2°) The result (1) is obtained by direct calculation of integrals or via the FOURIER transform.

The convolution operation $*$ has a "regularizing effect": with this p_t, when $f \in L^1(\mathbb{R})$, the regularized function $p_t * f$ is continuous and belongs to $L^1(\mathbb{R})$.

3°) The convolution operation $*$ is commutative and associative.

Answers. 4°) (a). For the chosen f, we have:

$$\begin{aligned}\left[\mathcal{P}_t(f)\right](0) &= \int_0^1 \frac{e^{-y^2/2t}}{\sqrt{2\pi t}}\,\mathrm{d}y \\ &= \int_0^{1/\sqrt{2t}} \frac{e^{-u^2/2}}{\sqrt{2\pi}}\,\mathrm{d}u \quad \left[\begin{array}{c}\text{through the change}\\ \text{of variables } u=y/\sqrt{t}\end{array}\right].\end{aligned}$$

Hence, when $t \to 0$, $\left[\mathcal{P}_t(f)\right](0) \to \int_0^{+\infty} \frac{e^{-u^2/2}}{\sqrt{2\pi}}\,\mathrm{d}u = \frac{1}{2} \neq f(0) = 1$.

(b) We have:

$$\begin{aligned}\left[\mathcal{P}_t(f)\right](x) &= \int_{\mathbb{R}} p_t(y) f(x-y)\,\mathrm{d}y \quad \text{[by definition]} \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-u^2/2} f(x-u\sqrt{t})\,\mathrm{d}u \\ &\qquad \left[\begin{array}{c}\text{through the change}\\ \text{of variables } u=y/\sqrt{t}\end{array}\right].\end{aligned}$$

We define: $F : (t,u) \in [0,+\infty) \times \mathbb{R} \mapsto F(t,u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} f(x-u\sqrt{t})$. Then:

$$\left\{\begin{array}{c}\text{For all } u \in \mathbb{R},\ F(\cdot,u) \text{ is continuous on } [0,+\infty)\,; \\ \lim_{t\to 0} F(t,u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} f(x); \\ |F(t,u)| \leqslant \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \times \|f\|_\infty \in L^1(\mathbb{R})\,.\end{array}\right.$$

Thus, according to the continuity theorem for a parameterized integral,

$$\left[\mathcal{P}_t(f)\right](x) \to \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-u^2/2} f(x)\,\mathrm{d}u = f(x) \text{ when } t \to 0.$$

5°) To show that $\varphi : (t,x) \mapsto \varphi(t,x) = \int_{\mathbb{R}} p_t(x-y) f(y)\,\mathrm{d}y$ possesses partial derivatives of any order, one uses the theorem allowing differentiation of a parameterized integral. Indeed, for nonnegative integers k, l such that $k+l \leqslant n$, there exists a polynomial function $P_{k,l}$ of degree bounded above by $2n$ such that

$$\frac{\partial^k}{\partial t^k} \frac{\partial^l}{\partial x^l} p_t(x-y) = \frac{1}{t^{1/2+2k+l}} P_{k,l}(t, x-y).$$

This can easily be checked by induction on n.

One deduces that:

$$\left| \frac{\partial^k}{\partial t^k} \frac{\partial^l}{\partial x^l} p_t(x-y) \times f(y) \right| \leqslant C\,|f(y)| \text{ whenever } t \in [a,A]\,.$$

Consequently,

$$\frac{\partial\varphi}{\partial t}(t,x) - \frac{1}{2}\frac{\partial^2\varphi}{\partial x^2}(t,x) = \int_{\mathbb{R}} \left(\frac{\partial p_t}{\partial t} - \frac{1}{2}\frac{\partial^2 p_t}{\partial x^2} \right)(x-y) \times f(y)\ \mathrm{d}y.$$

But

$$\frac{\partial p_t}{\partial t}(v) = \frac{1}{2}\frac{\partial^2 p_t}{\partial x^2}(v) = \frac{1}{2\sqrt{\pi}t^{5/2}}(v^2 - t)e^{-v^2/2t}.$$

We thus have proved that

$$\frac{\partial\varphi}{\partial t}(t,x) = \frac{1}{2}\frac{\partial^2\varphi}{\partial x^2}(t,x) \text{ for all } (t,x) \in (0,+\infty) \times \mathbb{R}.$$

---

211. ★★★ *When is $L^r(\mathbb{R})$ contained in $L^p(\mathbb{R}) + L^q(\mathbb{R})$?*

For $1 \leqslant p < +\infty$, let $L^p(\mathbb{R})$ denote the usual LEBESGUE space, equipped with the norm $\|\cdot\|_p$.

1°) On the vector space $L^p(\mathbb{R}) + L^q(\mathbb{R})$, with $1 \leqslant p, q < +\infty$, one defines the following function $\|\cdot\|_{p,q}$:

$$\|f\|_{p,q} = \inf \left\{ \|g\|_p + \|h\|_q : f = g + h \text{ with } g \in L^p(\mathbb{R}),\ h \in L^q(\mathbb{R}) \right\}. \quad (1)$$

(*a*) Show that $\|\cdot\|_{p,q}$ defines a norm on $L^p(\mathbb{R}) + L^q(\mathbb{R})$.

(*b*) Prove that $\left(L^p(\mathbb{R}) + L^q(\mathbb{R}), \|\cdot\|_{p,q} \right)$ is a BANACH space.

2°) Let $1 \leqslant p, q < +\infty$. Prove that:

(*a*) $L^r(\mathbb{R})$ is contained in $L^p(\mathbb{R}) + L^q(\mathbb{R})$ whenever $r \in [p, q]$.

(*b*) If r does not lie in the interval $[p, q]$, then $L^r(\mathbb{R})$ is not contained in $L^p(\mathbb{R}) + L^q(\mathbb{R})$.

3°) Let $1 \leqslant p, q < +\infty$ and $r \in [p, q]$. Check that the injection of $L^r(\mathbb{R})$ into $L^p(\mathbb{R}) + L^q(\mathbb{R})$ is continuous.

Hints. 1°) (*a*). The process proposed in Tapa 30 is used here.

(*b*) Here we have an opportunity to use a characterization of completeness of normed vector spaces which is not well-known (see Tapa 29). Indeed:

$$((X, \|\cdot\|) \text{ is complete})$$

$$\Updownarrow$$

$$\left(\begin{array}{c} \text{Every series in } X \text{ with general term } a_k, \text{ for} \\ \text{which } \sum_{k=0}^{+\infty} \|a_k\| < +\infty, \text{ converges in } X \\ \text{(towards a sum denoted as } \sum_{k=0}^{+\infty} a_k) \end{array} \right).$$

2°) (a). For $f \in L^r(\mathbb{R})$, consider $X = \{x \in \mathbb{R} : |f(x)| > 1\}$ and its complementary set X^c. Decompose f as follows:

$$f = f \cdot \mathbf{1}_X + f \cdot \mathbf{1}_{X^c},$$

where $\mathbf{1}_S$ denotes the indicator function of S.

(b) Distinguish two cases: $r < p$ and $r > q$.

Comment. For more details on this subject, we refer to the note:

J.-B. Hiriart-Urruty and P. Lassère, *When is $L^r(\mathbb{R})$ contained in $L^p(\mathbb{R}) + L^q(\mathbb{R})$?* American Math. Monthly 120, n°1 (2013), 55–61.

---

212. ★★★ *Least squares problems with inequalities*

Consider a system of linear inequalities in $\mathbb{R}^n$

$$a_i^T x \leqslant b_i \text{ for } i = 1, 2, ..., m \qquad (\mathcal{S})$$

where $a_i \in \mathbb{R}^n, b_i \in \mathbb{R}$. Suppose that $(\mathcal{S})$ is infeasible, that is to say there is no x satisfying all the inequalities in $(\mathcal{S})$. However, we are interested in "solving" $(\mathcal{S})$ in the best way. For that, we propose to tackle the following optimization problem:

$$\text{Minimize } f(x) = \sum_{i=1}^{m} \left(a_i^T x - b_i\right)_+^2 \text{ over } \mathbb{R}^n, \qquad (\mathcal{P})$$

where r_+ stands for the nonnegative part of r, *i.e.*, $r_+ = \max(r, 0)$. Mimicking the classical case dealing with systems of linear equalities, we call the solutions of $(\mathcal{P})$ the *least squares solutions* to the system of linear inequalities $(\mathcal{S})$.

We denote by A the $m \times n$ matrix whose columns are $a_1, a_2, ..., a_m$, and by b the vector in $\mathbb{R}^m$ whose components are $b_1, b_2, ..., b_m$. Accordingly, $(Ax-b)_+$ is the vector in $\mathbb{R}^m$ whose components are $(a_1^T x - b_1)_+, (a_2^T x - b_2)_+, ..., (a_m^T x - b_m)_+$.

1°) Warm-up.

(a) Let $n = 2, m = 3$,

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad b = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}.$$

Solve $(\mathcal{P})$ in this case.

(b) Let $n = 2, m = 4$,

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & -1 \\ 0 & 1 \end{bmatrix}, \quad b = \begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}.$$

Solve $(\mathcal{P})$ in this case.

2°) (a) Check that f is a differentiable convex function and determine its gradient $\nabla f(x)$ at any $x \in \mathbb{R}^n$.

(b) Deduce from the above that the solutions in $(\mathcal{P})$ are characterized by the equation:

$$\sum_{i=1}^{m} \left(a_i^T x - b_i\right)_+ a_i = A^T (Ax - b)_+ = 0. \tag{1}$$

3°) Prove that the least squares problem $(\mathcal{P})$ does have solutions, *i.e.*, there are points minimizing f on $\mathbb{R}^n$.

Hints. 2°) Easy to answer if one views $f(x)$ as $\sum_{i=1}^{m} r_i(h_i(x))$, where $h_i : \mathbb{R}^n \to \mathbb{R}$ is the affine function defined by $h_i(x) = a_i^T x - b_i$, and $r : \mathbb{R} \to \mathbb{R}$ is the increasing differentiable convex function defined by $r(t) = 1/2(t_+)^2$.

3°) There is no theorem in Analysis whose application would lead directly to the existence of minimizers of the function f. We therefore suggest for that purpose two possible alternative paths.

First approach. Using elementary techniques from Analysis (like extracting converging subsequences from bounded sequences) and Linear algebra (like $V \cap V^\perp = \{0\}$ for a vector space V). Since f is bounded from below (by 0 for example), consider a minimizing sequence, that is, a sequence (x_k) such that $f(x_k)$ converges to $\inf_{x \in \mathbb{R}^n} f(x)$ when $k \to +\infty$. Distinguish two cases: when (x_k) is bounded, and when (x_k) is unbounded.

Second approach. Transform the original problem $(\mathcal{P})$ into the following "equivalent" one (in a sense to be made precise):

$$\text{Minimize } \|v - b\|^2 \text{ over the set } \operatorname{Im} A + K, \tag{$\mathcal{P}'$}$$

where K denotes the closed convex cone in $\mathbb{R}^m$ consisting of vectors whose components are all nonpositive (in other words, $K = -\mathbb{R}^n_+$).

Answers. 1°) (a). The solution set is the half-line $\{0\} \times (-\infty, 1]$ in $\mathbb{R}^2$. We see in this example that the kernel of A reduces to $\{0\}$ while the solution set is unbounded... This phenomenon does not occur when solving least squares problems with equalities (the classical case).

(b) Here there is only one solution to the least squares problem for $(\mathcal{S})$, it is the origin $(0, 0) \in \mathbb{R}^2$.

2°) The function f is convex, $\mathcal{C}^1$ (but not $\mathcal{C}^2$), with

$$\nabla f(x) = 2A^T(Ax - b)_+ = \sum_{i=1}^{m} \left(a_i^T x - b_i\right)_+ a_i.$$

3°) Problem $(\mathcal{P}')$ (see Hints) is just a reformulation of the original problem $(\mathcal{P})$, taking into account the fact that $\sum_{i=1}^m \left(a_i^T x - b_i\right)_+^2$ is the square of the distance from $Ax - b$ to the closed convex cone $K = -\mathbb{R}^n_+$.

The optimal values in $(\mathcal{P})$ and $(\mathcal{P}')$ are equal. If $\overline{b}$ solves $(\mathcal{P}')$, then $\overline{b}$, expressed as $A\overline{x} + \overline{u}$ for some $\overline{x}$ in $\mathbb{R}^n$ and some $\overline{u}$ in K, allows us to bring out a solution $\overline{x}$ in $(\mathcal{P})$.

Since $K = -\mathbb{R}^n_+$ is a polyhedral closed convex cone, the cone $\operatorname{Im} A + K$ is closed convex as well. Solving Problem $(\mathcal{P}')$ amounts to projecting b on $\operatorname{Im} A + K$; this projection problem does have a solution $\overline{b}$ (which is actually unique).

In short: The solution set in $(\mathcal{P})$ is a nonempty closed convex polyhedron in $\mathbb{R}^n$, which can be characterized by the nonlinear equation $A^T(Ax - b)_+ = 0$.

Comments. For various simple approaches to the existence problem for least squares solutions of inequality systems, see:

J.-B. Hiriart-Urruty, L. Contesse and J.-P. Penot, *Least squares solutions of linear inequality systems: a pedestrian approach.* RAIRO - Operations Research. Published online: June 2016.

---

213. ★★★ *Following the path indicated by the opposite of the gradient*

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function such that $f(x) \to +\infty$ when $\|x\| \to +\infty$. We consider the following Cauchy problem:

$$\left\{ \begin{array}{l} \frac{\mathrm{d}x}{\mathrm{d}t} = -\nabla f\ (x(t)), \\ x(0) = x_0 \in \mathbb{R}^n. \end{array} \right. \qquad (\mathcal{C})$$

1°) Establish that the (unique) maximal solution of the CAUCHY problem $(\mathcal{C})$ is defined on an (open) interval I which is unbounded from above (hence, I contains $[0, +\infty)$).

2°) Prove that $\|\nabla f(x(t))\|$ tends to 0 when $t \to +\infty$.

Hints. 1°) Since the mapping $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable, the CAUCHY–LIPSCHITZ theorem applies to $(\mathcal{C})$: there exists a maximal solution $x(\cdot)$ of $(\mathcal{C})$ defined on an open interval I containing 0. It remains to prove that I is unbounded from above. A key-ingredient is to consider the function

$$\ell : t \in I \mapsto \ell(t) = f(x(t)) \text{ (the level of } f \text{ at } x(t)), \tag{1}$$

whose derivative is

$$\ell'(t) = \left\langle \nabla f(x(t)), \frac{\mathrm{d}x}{\mathrm{d}t} \right\rangle = - \left\| \nabla f\ (x(t)) \right\|^2. \tag{2}$$

The function ℓ is therefore decreasing on I.

2°) The key-tool to be considered is the function $m(t) = -\ \ell'(t) = \|\nabla f(x(t))\|^2$. The objective here is to prove that $m(t)$ tends to 0 when $t \to +\infty$. For that purpose, prove that the nonnegative function $m(\cdot)$ is integrable and LIPSCHITZ on $[0, +\infty)$.

Answers. 1°) For $r \in \mathbb{R}$, let K_r denote the sub-level set of f at level r, that is

$$K_r = \{u \in \mathbb{R}^n : f(u) \leqslant r\}.$$

Due to the assumptions made on f, all these sublevel-sets are compact.

Suppose that the right-endpoint β of the open interval I (on which is defined the maximal solution $x(\cdot)$ to $(\mathcal{C})$) is finite; we show that this assumption leads to a contradiction.

Because the function $\ell : t \in I \mapsto \ell(t) = f(x(t))$ is decreasing, we have $x([0, \beta)) \subset K_{x_0}$. Due to this bound and to the continuity of ∇f, $\mathrm{d}x/\mathrm{d}t = -\nabla f(x(\cdot))$ is bounded on $[0, \beta)$. This implies that $x(t) = x_0 + \int_0^t x'(s)\, \mathrm{d}s$ has a (left-) limit in β. But, in that case, it could be possible to extend the solution $x(\cdot)$ to $(\mathcal{C})$ beyond the point β, which contradicts the "maximality" of the interval I. Hence, the right endpoint β of I cannot be finite, $\beta = +\infty$.

2°) The function l is decreasing on $[0,+\infty)$ and bounded from below (because f is bounded from below); it therefore has a limit ℓ^* at $+\infty$. Considering the function $m(t) = -\,\ell'(t) = \|\nabla f(x(t))\|^2$, we note that

$$\lim_{t\to+\infty} \int_0^t m(s)\, ds = \lim_{t\to+\infty} [f(x_0) - f(x(t))] = f(x_0) - \ell^*. \qquad (3)$$

The function $m(\cdot)$ is therefore integrable on $[0,+\infty)$. The function $m'(\cdot)$ is continuous and bounded on $[0,+\infty)$; hence the function $m(\cdot)$ is LIPSCHITZ on $[0,+\infty)$. Combined with the integrability of $m(\cdot)$ on $[0,+\infty)$, this implies that $m(t)$ necessarily tends to 0 as $t \to +\infty$.

Comments. Since $f(x) \to +\infty$ when $\|x\| \to +\infty$ (a property sometimes called 1-coercivity of the function f), we are sure that there are minimizers of f on $\mathbb{R}^n$, hence critical points (*i.e.,* points $\overline{x}$ at which $\nabla f(\overline{x}) = 0$). However, the structure of the set of critical points can be fairly complicated, and the result proved in this Tapa does not say much about the behavior of the solution $t \in [0,+\infty) \mapsto x(t) \in \mathbb{R}^n$ (not $\nabla f(x(t))$ when $t \to +\infty$. We nevertheless mention the following partial results:

- If f has only finitely many critical points, then $x(t)$ does indeed have a limit when $t \to +\infty$ (actually one of these critical points).

- If the set of critical points of f has connected components which do not reduce to single points, and if $n \geqslant 2$, then $x(t)$ may not have any limit when $t \to +\infty$.

The situation is a bit simpler when f is convex (then critical points are also minimizers of f); an example of such a situation has been considered in Tapa 130, where $f : x \mapsto f(x) = \frac{1}{2}x^T Bx$ was a convex quadratic function associated with the $n \times n$ positive definite matrix B.

---

214. ★★★ *Directional derivative of a max function*

Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a continuous function, let Y be a nonempty closed subset in $\mathbb{R}^m$, and let g be defined on $\mathbb{R}^n$ by

$$g(x) = \sup_{y\in Y} f(x,y). \qquad (1)$$

The question we address here is that of possibly expressing the directional derivative of g in terms of the (partial) gradients of f (assumed to exist).

Having to minimize a function defined as in (1) frequently appears in applications. There is no hope of obtaining a differentiable g even if

the function f is (very) smooth as a function of x; even with a discrete set of parameters Y, the max operation used in defining g generally destroys differentiability of g (see Tapa 108). We therefore consider conditions ensuring that the directional derivative of g exists. Recall that the directional derivative $g'(x_0, d)$ of g at x_0 in the direction $d \in \mathbb{R}^n$ is defined as

$$g'(x_0, d) = \lim_{t>0\to 0} \frac{g(x_0 + td) - f(x)}{t}. \tag{2}$$

1°) General case. We make the following assumptions, around the considered point $x_0 \in \mathbb{R}^n$:

$$\begin{array}{c} Y(x) = \{y \in Y : g(x) = f(x, y)\} \text{ is nonempty;} \\ \cup_{x\in N} Y(x) \text{ is bounded for some neighborhood } N \text{ of } x_0; \\ \text{The partial gradients } \nabla_x f \text{ of } f \text{ with respect to } x \text{ exist} \\ \text{and } (x, y) \mapsto \nabla_x f(x, y) \text{ is continuous.} \end{array} \tag{$\mathcal{H}$}$$

Note that these assumptions are satisfied whenever Y is compact and $f : (x, y) \mapsto f(x, y)$ is a continuously differentiable function.

Prove that the directional derivative $g'(x_0, d)$ of g at x_0 in the direction $d \in \mathbb{R}^n$ exists and equals

$$g'(x_0, d) = \sup_{y_0 \in Y(x_0)} [\nabla_x f(x_0, y_0)]^T d. \tag{3}$$

2°) An application. Let S be a nonempty closed set in $\mathbb{R}^n$, and let d_S be the distance function to S (defined via the usual Euclidean distance on $\mathbb{R}^n$). Check that the directional derivative of d_S^2 exists and that it can be expressed with the help of

$$P_S(x) = \{s \in S : d_S(x) = \|x - s\|\}.$$

Hints. 1°) Consider an arbitrary sequence $(t_k)_k$ of positive real numbers converging to 0. Firstly, obtain a lower bound of

$$\liminf_{k\to+\infty} \frac{g(x_0 + t_k d) - g(x_0)}{t_k}$$

by just using the definitions of $g(x_0)$ and $Y(x_0)$. Secondly, obtain an upper bound of

$$\limsup_{k\to+\infty} \frac{g(x_0 + t_k d) - g(x_0)}{t_k}$$

with the help of mean value theorems applied to the partial functions $f(\cdot, y_k)$ where $y_k \in Y(x_0 + t_k d)$. At this stage, one should check the following "closedness" property for the set-valued mapping $Y(\cdot)$:

$$\left.\begin{array}{c} y_k \in Y(x_k) \text{ for all } k, \\ x_k \to x \text{ when } k \to +\infty, \\ y_k \to y \text{ when } k \to +\infty. \end{array}\right\} \Rightarrow y \in Y(x).$$

2°) Clearly,

$$-\, d_S^2(x) = \sup_{s \in S} \left(- \left\| x - s \right\|^2 \right).$$

Answers. 1°) Consider any sequence $(t_k)_k$ of positive real numbers converging to 0. On the one hand, according to the definitions of $g(x_0)$ and $Y(x_0)$, we have, for all $y_0 \in Y(x_0)$:

$$g(x_0 + t_k d) - g(x_0) \geqslant f(x_0 + t_k d, y_0) - f(x_0, y_0),$$

so that

$$\begin{array}{rcl} \liminf\limits_{k \to +\infty} \dfrac{g(x_0 + t_k d) - g(x_0)}{t_k} & \geqslant & \lim\limits_{k \to +\infty} \dfrac{f(x_0 + t_k d, y_0) - f(x_0, y_0)}{t_k} \\ & = & \left[\nabla_x f(x_0, y_0) \right]^T d. \end{array}$$

On the other hand,

$$g(x_0 + t_k d) - g(x_0) \leqslant f(x_0 + t_k d, y_k) - f(x_0, y_k),$$

where $y_k \in Y(x_0 + t_k d)$. Now, due to the mean value theorem applied to the partial function $f(\cdot, y_k)$, there exists a $\widetilde{x}_k$ lying in the line-segment joining x_0 to $x_0 + t_k d$ such that

$$\frac{f(x_0 + t_k d, y_k) - f(x_0, y_k)}{t_k} = \left[\nabla_x f(\widetilde{x}_k, y_k) \right]^T d.$$

Whence we have:

$$\frac{g(x_0 + t_k d) - g(x_0)}{t_k} \leqslant \left[\nabla_x f(\widetilde{x}_k, y_k) \right]^T d. \tag{4}$$

The key-phase is now to pass to the limit on k. The first ingredient is the following "closedness" property for the set-valued mapping $Y(\cdot)$:

$$\left.\begin{array}{c} y_k \in Y(x_k) \text{ for all } k, \\ x_k \to x \text{ when } k \to +\infty, \\ y_k \to y \text{ when } k \to +\infty. \end{array}\right\} \Rightarrow y \in Y(x).$$

This follows readily from the continuity of f and the definition of $Y(\cdot)$. The second ingredient is that the sequences involved in (4), namely $(\widetilde{x}_k), (y_k)$, could be supposed convergent, subsequencing if necessary; this follows from the assumed boundedness property of $Y(\cdot)$ around x_0 (see the assumptions in $(\mathcal{H})$). As a result, one can suppose that y_k converges to some y_0 in $Y(x_0)$. The assumed continuity of $(x, y) \mapsto \nabla_x f(x, y)$ now means that we can infer from (4):

$$\limsup_{k \to +\infty} \frac{g(x_0 + t_k d) - g(x_0)}{t_k} \leqslant \sup_{y_0 \in Y(x_0)} \left[\nabla_x f(x_0, y_0)\right]^T d.$$

We therefore have proved that

$$\lim_{k \to +\infty} \frac{g(x_0 + t_k d) - g(x_0)}{t_k} = \sup_{y_0 \in Y(x_0)} \left[\nabla_x f(x_0, y_0)\right]^T d.$$

2°) We have that

$$-d_S^2(x) = \sup_{s \in S} \left(- \|x - s\|^2 \right).$$

The function $f : (x, s) \mapsto f(x, s) = - \|x - s\|^2$ is continuously differentiable, with $\nabla_x f(x, s) = -2(x - s)$. Here $Y = S$ and

$$Y(x) = P_S(x) = \{s \in S : d_S(x) = \|x - s\|\}.$$

All the assumptions in $(\mathcal{H})$ are satisfied. Consequently, for all points x_0 and all directions d, the directional derivative of the function d_S^2 at x_0 in the d direction exists and is equal to

$$\left(d_S^2\right)'(x_0, d) = \inf_{s_0 \in P_S(x_0)} 2\left[(x_0 - s_0)\right]^T d. \tag{5}$$

Comments. - When $Y(x_0)$ is single-valued, say $Y(x_0) = \{y_0\}$, then formula (3) boils down to

$$g'(x_0, d) = \left[\nabla_x f(x_0, y_0)\right]^T d \text{ for all } d \in \mathbb{R}^n.$$

In other words, g is Gateaux-differentiable at x_0 with

$$\nabla g(x_0) = \nabla_x f(x_0, y_0). \tag{6}$$

In the context of 2°), when S is convex, $P_S(x)$ contains only a single point, the so-called orthogonal projection $p_S(x)$ of x onto S; then

$$\nabla(d_S^2)(x) = 2(x - p_S(x)).$$

- Question 2°). We recover in this illustration results on the directional differentiability of d_S^2 that had been observed in another context; for that, peruse Tapas 75 and 198.

- The first roots of a formula like (3) go back to J.M. Danskin (1967), at least when Y is compact. When the functions $f(\cdot, y)$ are convex, the max function $g(\cdot) = \sup_{y \in Y} f(\cdot, y)$ is also convex, and things are therefore easier to handle (the existence of directional derivatives, for example).

---

215. ★★★ *An everywhere continuous but nowhere differentiable vector-valued function*

In 1872, Weierstrass proposed an example of a pathological real-valued function of a real variable: it was continuous everywhere but differentiable nowhere. The definition of the function was not simple, it was given as a sum of a series. The situation is easier, in a certain sense, if infinite-dimensional spaces are allowed; there is more room for pathologies. We propose in this Tapa everywhere continuous but nowhere differentiable mappings taking values in the Lebesgue spaces $L^p([0,1])$.

Let $f : [0,1] \to L^p([0,1])$ be defined as follows:

$$t \in [0,1] \mapsto f(t) = \mathbf{1}_{[0,t]} \text{ (the indicator function of } [0,t]\,). \qquad (1)$$

We denote by $\|\cdot\|_{L^p}$ the usual norm on $L^p([0,1])$.

1°) General case: $p > 1$.

Prove that f is continuous everywhere but differentiable nowhere.

2°) The particular case $p = 1$.

Show that f is continuous, even an isometry (or distance preserving), but differentiable nowhere.

3°) The particular case $p = 2$.

Here, $L^2([0,1])$ is a Hilbert space when endowed with the usual scalar product $\langle u, v \rangle = \int_{[0,1]} u(x)v(x)\, \mathrm{d}\lambda(x)$.

Let $s < s' \leqslant t < t'$. Check the following strange behavior of successive "variations" $f(t') - f(t)$ and $f(s') - f(s)$ or "slopes" $\frac{f(t')-f(t)}{t'-t}$ and $\frac{f(s')-f(s)}{s'-s}$: they are always orthogonal elements in $L^2([0,1])$.

Hints. 1°) Calculate $\|f(s) - f(t)\|_{L^p}$ and $\left\| \frac{f(t_0+h)-f(t)}{h} \right\|_{L^p}$ explicitly.

2°) Here, in contrast to the case where $p > 1$, the differential quotient $q_h = \frac{f(t_0+h)-f(t_0)}{h}$ is always of norm 1. The objective is to show that q_h cannot converge to a function in $L^1([0,1])$.

Answers. 1°) If t and s lie in $[0,1]$, $t \neq s$, the "variation" $f(s) - f(t)$ is the indicator function of $(t,s]$ or of $(t,s]$. Thus,

$$\|f(s) - f(t)\|_{L^p} = |s-t|^{1/p}. \tag{2}$$

Consequently, f is a continuous function. But it is just HÖLDER continuous, not LIPSCHITZ continuous.

Suppose that f is differentiable at $t_0 \in [0,1]$: $\frac{f(t_0+h)-f(t)}{h}$ has a limit $f'(t_0)$ in $L^p([0,1])$ when $h \to 0$. That would imply

$$\left\| \frac{f(t_0+h)-f(t)}{h} \right\|_{L^p} \to \|f'(t_0)\|_{L^p} \text{ when } h \to 0.$$

But, $\left\| \frac{f(t_0+h)-f(t)}{h} \right\|_{L^p} = \frac{|h|^{1/p}}{|h|} = \frac{1}{|h|^{(p-1)/p}} \to +\infty$ when $h \to 0$. Thus, we obtain a contradiction. It is impossible for f to be differentiable at t_0.

2°) For $p = 1$, we see with (2) that

$$\|f(s) - f(t)\|_{L^1} = |s-t| \text{ for all } s,t.$$

Indeed, f is an isometry (or distance preserving).

Suppose that f is differentiable at $t_0 \in [0,1]$: $\frac{f(t_0+h)-f(t)}{h}$ has a limit $f'(t_0)$ in $L^1([0,1])$ when $h \to 0$. With $h > 0$ for example, we have:

$$q_h = \frac{f(t_0+h)-f(t_0)}{h} = \frac{1}{h}\mathbf{1}_{(t_0,t_0+h]}.$$

It is impossible for q_h to converge, as $h \to 0$, to some function in $L^1([0,1])$. One simple reason is, for example, that the sequence $\left(q_{1/n}\right)_{n\geqslant 1}$ is not a CAUCHY sequence in $L^1([0,1])$ (indeed, for all n, $\|q_{1/2n} - q_{1/n}\|_{L^1} \geqslant \frac{1}{2}$). An alternate reason (of a more advanced mathematical level) is that q_h converges weakly to a DIRAC delta function supported at the point t_0.

3°) Since the intervals (s,s') and (t,t') are disjoint, we have:

$$\langle f(s') - f(s), f(t') - f(t)\rangle = \int_{[0,1]} \mathbf{1}_{(s,s')}(x).\mathbf{1}_{(t,t')}(x)\, \mathrm{d}\lambda(x) = 0.$$

Comments. - There is a beautiful differentiability result on locally LIPSCHITZ mappings due to RADEMACHER (1919): a locally LIPSCHITZ mapping f from $\mathbb{R}^n$ into $\mathbb{R}^m$ is differentiable almost everywhere, that is, the points in $\mathbb{R}^n$ at which f is not differentiable form a set of LEBESGUE measure zero.

- When f is a mapping from $\mathbb{R}$ into a BANACH space E, the situation is more complicated, as expected. However, if E enjoys the so-called RADON–NIKODYM (RN) property, a LIPSCHITZ mapping from $\mathbb{R}$ into E is also differentiable almost everywhere. Reflexive BANACH spaces (like HILBERT spaces) do enjoy the RN property, $L^1([0,1])$ does not... The examples displayed in this Tapa show that if f just satisfies a HÖLDER property or if the image space E does not enjoy the RN property, the above mentioned differentiability result breaks down.

---

216. ★★★ *Bounding, averaging, relaxing... in a quadratic minimization problem*

An important problem in Combinatorial optimization can be formulated as follows:

$$(\mathcal{P}) \quad \left\{ \begin{array}{c} \text{Maximize } q(x) = \langle Ax, x\rangle \text{ (other possible notation } x^T Ax) \\ \text{subject to } x_i^2 = 1 \text{ for all } i = 1, 2, ..., n \end{array} \right. ,$$

where $A = [a_{ij}] \in \mathcal{S}_n(\mathbb{R})$ is given and $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ is the optimization variable vector.

There is little to say concerning the existence and uniqueness of solutions of $(\mathcal{P})$: the constraint set $\{(\pm 1, \pm 1, \dots, \pm 1\}$ is finite; if $\overline{x}$ is a solution, it is nonzero and $-\overline{x}$ is also a solution.

We denote by q^* the optimal value in $(\mathcal{P})$. Our objective in this Tapa is to illustrate what can be done to bound q^*, to "relax" the difficult problem $(\mathcal{P})$.

1°) *Averaging.* We suppose that the $x_i(\xi)$'s, $i = 1, ..., n$, are independent random variables, obeying the same probability law:

$$P(x_i(\xi) = 1) = P(x_i(\xi) = -1) = 1/2.$$

The $x_i(\xi)$'s are so-called BERNOULLI random variables.

(*a*) Observing that $\langle Ax, x\rangle = \sum_{i=1}^n a_{ii} x_i^2 + 2\sum_{i<j} a_{ij} x_i x_j$, calculate the expectation of the random variable $\langle Ax(\xi), x(\xi)\rangle$.

(*b*) Deduce that:

$$\mathrm{tr}(A) \leqslant q^*. \tag{1}$$

2°) *Bounding.* A problem akin to $(\mathcal{P})$ is the following one:

$$(Q)\ \left\{\begin{array}{c}\text{Maximize } q(x)=\langle Ax,x\rangle\\ \text{subject to } \|x\|^2=\sum_{i=1}^n x_i^2=1\end{array}\right.,$$

whose optimal value turns out to be $\lambda_{\max}(A)$, that is, the largest eigenvalue of A.

By comparing $(\mathcal{P})$ and (Q), rescaling the variables x if necessary, show that:

$$q^* \leqslant n\cdot\lambda_{\max}(A). \tag{2}$$

3°) *Relaxation.* Consider the new optimization problem:

$$(\mathcal{P})_{sdp}\ \left\{\begin{array}{c}\text{Maximize } \mathrm{tr}(AX)\\ \text{subject to } X\succcurlyeq 0 \text{ and } X_{ii}=1 \text{ for all } i=1,...,n.\end{array}\right.$$

Here, $X=[X_{ij}]\in\mathcal{S}_n(\mathbb{R})$ is the optimization variable matrix. Denote by q^*_{sdp} the optimal value in $(\mathcal{P})_{sdp}$.

(a) After having checked that $\mathrm{tr}(AX)=\langle Ax,x\rangle$ whenever $X=xx^T$, explain why $(\mathcal{P})_{sdp}$ is a "relaxed" form of $(\mathcal{P})$.

(b) Deduce from the above that:

$$q^*\leqslant q^*_{sdp}. \tag{3}$$

(c) Comment on the advantages and disadvantages of solving $(\mathcal{P})_{sdp}$ instead of $(\mathcal{P})$.

Hints. 1°) - Beware: The objective function in $(\mathcal{P})$ is quadratic, nothing more (*a priori* neither convex nor concave).

- For independent random variables, the expectation of their product is the product of expectations.

3°) - The involved variables are positive semidefinite matrices X; which explains the acronym *sdp* attached to this problem.

- Here, we have enlarged the underlying optimization space (from $\mathbb{R}^n$ into $\mathcal{S}_n(\mathbb{R})$); which explains the term "relaxation" used for this new version of the problem.

Answers. 1°) (a). We have:

$$E\left\langle Ax(\xi),x(\xi)\right\rangle=\sum_{i=1}^n a_{ii}E\left(x_i^2(\xi)\right)+2\sum_{i<j}a_{ij}E\left(x_i(\xi)x_j(\xi)\right).$$

But $x_i^2(\xi)$ is a random variable which takes the value 1 with probability 1, so that $E\left(x_i^2(\xi)\right) = 1$. The random variables $x_i(\xi)$ and $x_j(\xi)$, $i < j$, are independent; since $E\left(x_k(\xi)\right) = 0$, we therefore obtain $E\left(x_i(\xi)x_j(\xi)\right) = E\left(x_i(\xi)\right) \cdot E(x_j(\xi)) = 0$. Consequently,

$$E\left\langle Ax(\xi), x(\xi)\right\rangle = \sum_{i=1}^{n} a_{ii} = \mathrm{tr}(A).$$

(b) Any stochastic realization of the random vector $x(\xi) = (x_1(\xi), ..., x_n(\xi))$ satisfies the constraints in $(\mathcal{P})$ (we say that it is *feasible* in $(\mathcal{P})$), because $x_i^2(\xi) = 1$ almost surely for all i. As a result, $\langle Ax(\xi), x(\xi)\rangle \leqslant q^*$ almost surely, and:

$$\left(\langle Ax(\xi), x(\xi)\rangle \leqslant q^* \text{ almost surely}\right) \Longrightarrow \left(E\left\langle Ax(\xi), x(\xi)\right\rangle \leqslant q^*\right).$$

2°) If $x_i^2 = 1$ for all $i = 1, ..., n$, then $\|x\|^2 = \sum_{i=1}^n x_i^2 = n$. Thus,

$$\begin{aligned} q^* &\leqslant \max_{\|x\|^2=n} \langle Ax, x\rangle \\ &\quad \text{(because we have enlarged the constraint set in } (\mathcal{P})) \\ &= n \cdot \max_{\|u\|^2=1} \langle Au, u\rangle \ \text{ (by just rescaling, } u = \frac{x}{\sqrt{n}}). \end{aligned}$$

But $\max_{\|u\|^2=1} \langle Au, u\rangle$ is the variational formulation of $\lambda_{\max}(A)$. Hence the announced inequality (2) is proved.

3°) Starting from a vector $x \in \mathbb{R}^n$, we build a matrix $X = [X_{ij}] \in \mathcal{S}_n(\mathbb{R})$ as follows: $X = xx^T$. What do we get? Clearly:

- X is symmetric, even of rank 1 if $x \neq 0$;
- X is positive semidefinite since $\langle Xu, u\rangle = \left(x^T u\right)^2 \geqslant 0$ for all $u \in \mathbb{R}^n$;
- if x is feasible in $(\mathcal{P})$, then $X_{ii} = x_i^2 = 1$ for all $i = 1, ..., n$;
- using properties of the trace function, $\mathrm{tr}(AX) = \mathrm{tr}(Axx^T) = \mathrm{tr}\left\langle x^T A^T x\right\rangle = \langle Ax, x\rangle$.

Hence, passing from $(\mathcal{P})$ to $(\mathcal{P})_{sdp}$, we have enlarged the domain of possibilities; $(\mathcal{P})_{sdp}$ is indeed "a relaxed form" of $(\mathcal{P})$. As a consequence, we have : $q^* \leqslant q^*_{sdp}$.

Advantages of solving $(\mathcal{P})_{sdp}$ instead of $(\mathcal{P})$:

- the objective function $X \mapsto \mathrm{tr}(AX)$ is linear;
- the constraints are either linear ($X_{ii} = 1$) or conical convex ($X \succcurlyeq 0$).

Disadvantages, or what has been lost in passing from $(\mathcal{P})$ to $(\mathcal{P})_{sdp}$:
- the conical convex constraint $(X \succcurlyeq 0)$ is not easy to handle numerically;
- the gap $q^*_{sdp} - q^*$ $(\geqslant 0)$ should be controlled, i.e., bounded from above.

Comments. - One could add $\alpha \sum_{i=1}^n x_i^2$, with $\alpha > 0$, to the objective function $\langle Ax, x\rangle$ in $(\mathcal{P})$ without altering $(\mathcal{P})$ (since $x_i^2 = 1$ for any feasible vector $x = (x_1, \dots, x_n)$ in $(\mathcal{P})$). Therefore, there is no loss of generality in assuming that A is positive definite, hence $x \mapsto \langle Ax, x\rangle$ is strictly convex.

The points $\{(\pm 1, \pm 1, \dots, \pm 1\}$ are vertices of the box $[-1, +1]^n$ in $\mathbb{R}^n$. Thus, solving $(\mathcal{P})$ could be seen as equivalent to the next optimization problem:

$$(\mathcal{P})' \quad \begin{cases} \text{Maximize } q(x) = \langle Ax, x\rangle \\ \text{subject to } x \in [-1, +1]^n, \end{cases}$$

with A positive definite.

We therefore have to *maximize* a strictly *convex* quadratic function over a convex polyhedron, an optimization problem as difficult as the original one...

- To be fully equivalent to $(\mathcal{P})$, one should add another constraint in problem $(\mathcal{P})_{sdp}$: X has to be of rank one, *i.e.* of the form xx^T for some x in $\mathbb{R}^n$. Now, the constraint "X is a rank one matrix" can be translated into mathematical terms in various ways, like the equality constraint: "$\mathrm{rank}(X) = 1$" (but the rank function behaves very wildly, see Tapa 165), or this alternative: "$\|X\| = n$" (due to the other constraints imposed on X). In both cases, this equality constraint is difficult to handle, that is why it does not appear in the relaxed form $(\mathcal{P})_{sdp}$.

- Problem $(\mathcal{P})$ is the matricial formulation of an important combinatorial optimization problem, called *Max-cut*, whose roots are in Statistical physics or Circuit layout design.

---

217. ★★★ *An unusual integral inequality*

Let $f : \mathbb{R} \to \mathbb{R}$ be a function satisfying the following three integrability properties:

$$\begin{cases} \text{- The function } f \text{ itself is in the LEBESGUE space } L^1(\mathbb{R}); \\ \text{- The function } x \mapsto [x\ f(x)]^2 \text{ belongs to } L^1(\mathbb{R}); \\ \text{- The function } x \mapsto [f(x)]^2 \text{ belongs to } L^1(\mathbb{R}). \end{cases}$$

Prove that

$$\int_{\mathbb{R}} |f(x)|\ \mathrm{d}x \leqslant \sqrt{8} \left(\int_{\mathbb{R}} [x\ f(x)]^2\ \mathrm{d}x \right)^{\frac{1}{4}} \left(\int_{\mathbb{R}} [\ f(x)]^2\ \mathrm{d}x \right)^{\frac{1}{4}}. \qquad (1)$$

Hints. - The power 1/4 in the integrals is understandable for "homogeneity" reasons: if αf, with $\alpha > 0$, is substituted for f in (1), then the coefficient α can be factorized out in the inequality (1).

- Let $p > 0$. For a term like $xf(x)$ to appear in the left-hand side integral, it is advisable to "isolate" the origin 0:

$$\int_{\mathbb{R}} |f(x)| \ \mathrm{d}x = \int_{[-p,p]} |f(x)| \ \mathrm{d}x + \int_{\mathbb{R}\setminus[-p,p]} \frac{1}{|x|} \, |xf(x)| \ \mathrm{d}x.$$

Then, apply the Cauchy–Schwarz inequality to both right-hand integrals, and minimize with respect to the parameter.

Answer. Let $p > 0$. We divide $\int_{\mathbb{R}} |f(x)| \, \mathrm{d}x$ into two pieces, with $|xf(x)|$ appearing in the second one:

$$\int_{\mathbb{R}} |f(x)| \ \mathrm{d}x = \int_{[-p,p]} |f(x)| \ \mathrm{d}x + \int_{\mathbb{R}\setminus[-p,p]} \frac{1}{|x|} \, |xf(x)| \ \mathrm{d}x. \qquad (2)$$

By applying the Cauchy–Schwarz inequality to both integrals in (2), we obtain:

$$\int_{\mathbb{R}} |f(x)| \, \mathrm{d}x \leqslant \sqrt{2p} \, \sqrt{\int_{[-p,p]} [\ f(x)]^2 \, \mathrm{d}x} + \sqrt{\frac{2}{p}} \, \sqrt{\int_{\mathbb{R}\setminus[-p,p]} [x \ f(x)]^2 \, \mathrm{d}x}.$$

Consequently,

$$\int_{\mathbb{R}} |f(x)| \ \mathrm{d}x \leqslant \sqrt{2pI} + \sqrt{\frac{2}{p} J}, \qquad (3)$$

where $I = \int_{\mathbb{R}} [f(x)]^2 \, \mathrm{d}x$ and $J = \int_{\mathbb{R}} [x \ f(x)]^2 \, \mathrm{d}x$. This inequality (3) is valid for all $p > 0$. We therefore minimize the function

$$p > 0 \mapsto \sqrt{2pI} + \sqrt{\frac{2}{p} J}$$

and easily obtain its minimal value $2\sqrt{2} \ I^{\frac{1}{4}} J^{\frac{1}{4}}$. Hence, the proposed inequality (1) is proved.

Comments. There are dozens of functional inequalities..., this one is unusual however; it is loosely related to the so-called "uncertainty principle".

We do not know of the (non-trivial) functions f for which equality holds true in the inequality (1). Neither do we know of any "variational" proof of this inequality (*i.e.*, as the result of a minimization problem in some appropriate space of functions).

The CAUCHY–SCHWARZ inequality played an instrumental role in the proof of inequality (1). Actually, this inequality was taken from the following nice booklet (pages 105–106), the kind of mathematical document I would have liked to have written.

J.M. STEELE, *The* CAUCHY–SCHWARZ *Master Class.* Cambridge University Press (2004).

---

218. ★★★ *Prime numbers which are sums of two squares*

2017, the year when this collection of Tapas was published, is a prime number... But it is more than that, it is the sum of two squares of integers: $44^2 + 9^2$. There are not so many numbers like this... Indeed, between 2000 and 2050, only two numbers are prime and sums of two squares: 2017, as we said above, and $2029 = 2^2 + 45^2$. How to characterize such integers? This is the aim of the present Tapa. We shall prove the following theorem of FERMAT and EULER: *a prime number (greater than 3) is the sum of two squares of integers if and only if it is of the form $4k + 1$, with k a positive integer.* This is the case for 5, 13, 17,... , 2017, 2029. The path that we follow to prove it has a very "combinatorial" flavor.

A. The easy part.

Check that a prime number which is the sum of two squares of integers is necessarily of the form $4k + 1$.

B. The converse part.

1°) Warm-up. Let S be a finite set and let $f : S \to S$.

- If f is an involution which has only one fixed point, then the number of elements in S is odd.

- If the number of elements in S is odd, any involution f on S possesses at least one fixed point.

2°) Let p be a prime of the form $4k + 1$, with k a positive integer. Let

$$\mathcal{E}(p) = \left\{(a, b, c) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N} : 4ab + c^2 = p \ (= 4k + 1)\right\}. \qquad (1)$$

(a) Check that $\mathcal{E}(p)$ is a nonempty finite set consisting of triples of positive integers.

(b) Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be defined as follows:

$$f(a, b, c) = \begin{cases} (b, a - c - b, c + 2b) & \text{if } c < a - b; \\ (a, c - a + b, 2a - c) & \text{if } a - b < c < 2a; \\ (c - a + b, a, c - 2a) & \text{if } c > 2a. \end{cases} \qquad (2)$$

- Prove that f is an involution on $\mathcal{E}(p)$.

- Check that $(1, k, 1)$ is the only fixed point of f.

(c) Let $g : \mathbb{R}^3 \to \mathbb{R}^3$ be the involution defined by: $g(a, b, c) = (b, a, c)$.

Deduce from the results above that f has at least one fixed point $(a^*, a^*, c^*) \in \mathcal{E}(p)$, and that such a fixed point satisfies

$$4a^*a^* + (c^*)^2 = (2a^*)^2 + (c^*)^2 = p.$$

C. An illustrative example.

Consider $p = 41 = 4 \times 10 + 1$. Describe in detail the process developed in 2°) above.

Hints. B. 1°) An involution f on S is a bijection from S into S such that $f \circ f$ is the identity map on S, or equivalently such that $f = f^{-1}$.

An element s in S is said to be a fixed point of f when $f(s) = s$.

2°) (b). Clearly, f is a continuous piecewise-linear function on $\mathbb{R}^3$; but only the action of f on the finite set $\mathcal{E}(p)$ is considered here.

Answers. A. The square of every even integer $2n$ is $4n^2$, while the square of every odd integer $2m + 1$ is $4m^2 + 4m + 1$. Consequently, the sum of squares of two integers is of the form $4k, 4k + 1$ or $4k + 2$. Thus, if moreover this sum is assumed to be prime (greater than 3), it is necessarily of the form $4k + 1$.

B. 1°). When $f : S \to S$ is an involution, we have: for all $s \in S$,

$$f(s) = s' \ (\in S) \text{ and } f(s') = s.$$

So, there are only two possibilities: fixed points s (*i.e.,* $f(s) = s$) or exchange of two elements $s \neq s'$ (*i.e.,* $f(s) = s'$ and $f(s') = s$).

- On one hand. If we assume that f has only one fixed point, the actions of f on the other elements in S consist in exchanges of two different elements; they therefore go two by two. As a result, the number of elements in S is odd.

- On the other hand. If the number of elements in S is odd, the number of fixed points for f is $1, 3, 5, ...$ (an odd number); so f possesses at least one fixed point.

2°) (a). Clearly, $(1, k, 1) \in \mathcal{E}(p)$. Also, because p is a prime number, it cannot be factorized (as $p = c^2$ or as $p = 4ab$ (an even number)); thus: $a > 0, b > 0$ and $c > 0$ whenever $(a, b, c) \in \mathcal{E}(p)$.

As crude upper bounds for a, b, c (when $(a, b, c) \in \mathcal{E}(p)$), one gets:

$$a \leqslant \frac{p}{4},\ b \leqslant \frac{p}{4},\ c \leqslant \sqrt{p}.$$

Obviously, once a and b have been chosen, there is only one possibility for c.

(b) As it is easy to check, $f(a, b, c) = (a', b', c') \in \mathcal{E}(p)$ whenever $(a, b, c) \in \mathcal{E}(p)$ (just because $4a'b' + (c')^2 = 4ab + c^2$).

f is indeed an involution on $\mathcal{E}(p)$; it suffices to check that

$$(f \circ f)(a, b, c) = f(a', b', c') = (a, b, c)$$

in all three possible cases exhibited in the definition (2) of f. To just take an example, suppose that $c < a - b$ so that $f(a, b, c) = (a', b', c') = (b, a - c - b, c + 2b)$. Then $c' > 2a'$ and

$$f(a', b', c') = (c' - a' + b', a', c' - 2a') = (a, b, c).$$

Because $1 - k < 1 < 2$, the image of $(1, k, 1)$ under f is $(1, k, 1)$. It remains to check that it is the only fixed point of f in $\mathcal{E}(p)$. Indeed, to have $f(a, b, c) = (a', b', c') = (a, b, c)$ implies that we are in the second case in the definition "by parts" of f (see (2)); thus $c = a$. But, this implies that $a(4b + 1) = p$; because $p = 4k + 1$ is prime, this leads to $a = 1$ and $b = k$.

(c) g is obviously an involution on $\mathcal{E}(p)$. Now, because the number of elements in S is odd, the involution g possesses at least one fixed point, say (a^*, a^*, c^*). As a result, there indeed exist positive integers a^* and c^* satisfying

$$4a^*a^* + (c^*)^2 = (2a^*)^2 + (c^*)^2 = p.$$

C. Here, $\mathcal{E}(41)$ consists of 11 elements (a, b, c), that we describe below together with their images (a', b', c') under the involution f defined in (2):

$$\left\{ \begin{array}{cccc} \begin{pmatrix} a, b, c \\ a', b', c' \end{pmatrix} & \begin{pmatrix} 2, 5, 1 \\ 2, 4, 3 \end{pmatrix} & \begin{pmatrix} 5, 2, 1 \\ 2, 2, 5 \end{pmatrix} & \begin{pmatrix} 1, 8, 3 \\ 10, 1, 1 \end{pmatrix} \\ \begin{pmatrix} 8, 1, 3 \\ 1, 4, 5 \end{pmatrix} & \begin{pmatrix} 2, 4, 3 \\ 2, 5, 1 \end{pmatrix} & \begin{pmatrix} 10, 1, 1 \\ 1, 8, 3 \end{pmatrix} & \begin{pmatrix} 1, 10, 1 \\ 1, 10, 1 \end{pmatrix} \\ \begin{pmatrix} 4, 2, 3 \\ 4, 1, 5 \end{pmatrix} & \begin{pmatrix} 1, 4, 5 \\ 8, 1, 3 \end{pmatrix} & \begin{pmatrix} 4, 1, 5 \\ 4, 2, 3 \end{pmatrix} & \begin{pmatrix} 2, 2, 5 \\ 5, 2, 1 \end{pmatrix}. \end{array} \right.$$

The only fixed point for f is $(1, k, 1) = (1, 10, 1)$.

Concerning the involution $g : (a, b, c) \mapsto (b, a, c)$, a fixed point of it in $\mathcal{E}(41)$ is $(a^*, b^* = a^*, c^*) = (2, 2, 5)$; this provides the decomposition $41 = 4^2 + 5^2$.

Comments. - In the decomposition of p as $m^2 + n^2$, necessarily one integer, say m, is even, and the other one, n, is odd. It turns out that this pair (m, n) is unique.

- The process followed here to prove the existence of a decomposition of p as a sum of squares is not constructive, except for rather small p. There are indeed algorithms to determine such decompositions.

- This theorem proving that a prime number p of the form $4k + 1$ can be expressed as the sum of squares of integers was stated by FERMAT; the first true proof, however, is due to EULER. There are other results of this kind in the literature on Number theory, for example: prime numbers of the form $8k + 3$ can be decomposed as $m^2 + 2n^2$; an integer is the sum of 3 squares if and only if it is of the form $4^k(8l + 7)$ with integers k and l (a result due to LEGENDRE); every integer can be decomposed as the sum of 4 squares (a result due to LAGRANGE). Only big names!

- The idea of the proof of the result presented in this Tapa is due to D. ZAGIER,

D. ZAGIER, *A one-sentence proof that every prime $p \equiv 1$ (mod 4) is a sum of two squares.* Amer. Math. Monthly 97, 144 (1990).

However, the presentation there is very condensed; in particular, one does not see why and how the "diabolic" involution f was designed; indeed one can explain, if necessary, how one obtains this piecewise-description of f (in short: the necessity of having only $(1, k, 1)$ as a fixed point of f on $\mathcal{E}(p)$).

---

219. ★★★ *On the areas of convex polygons*

We consider a convex quadrilateral $ABCD$ in the plane, whose area is $\mathcal{A}$; each of the sides is divided into 5 equal parts. This gives rise to 25 small quadrilaterals of various areas inside the main quadrilateral $ABCD$. Our objective is to prove that the area of the central quadrilateral $EFGH$ is exactly $\mathcal{A}/25$. See Figure 1 below.

This could be done by cleverly using results in plane geometry like THALES' theorem on triangles, formulas for areas of quadrilaterals, etc.; this is the objective of Question 1°). But the posed question is a pretext to do more advanced mathematics and to prove a more general result (Question 2°)).

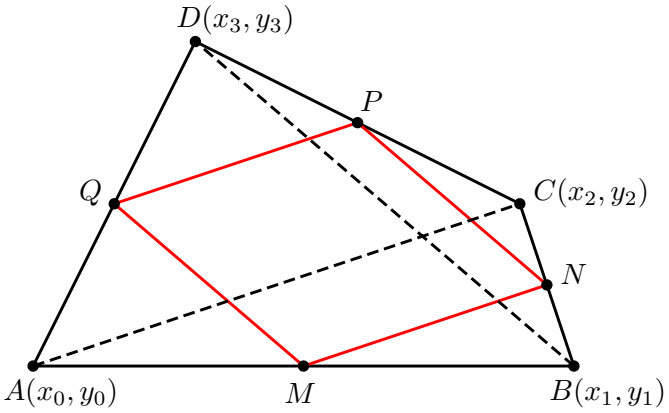

Figure 1

1°) *Areas of convex quadrilaterals in the plane*

(*a*) Let $ABCD$ be a convex quadrilateral in the plane; the Cartesian coordinates of the vertices A, B, C, D are denoted by (x_0, y_0), (x_1, y_1), (x_2, y_2), (x_3, y_3), respectively. See Figure 2 below.

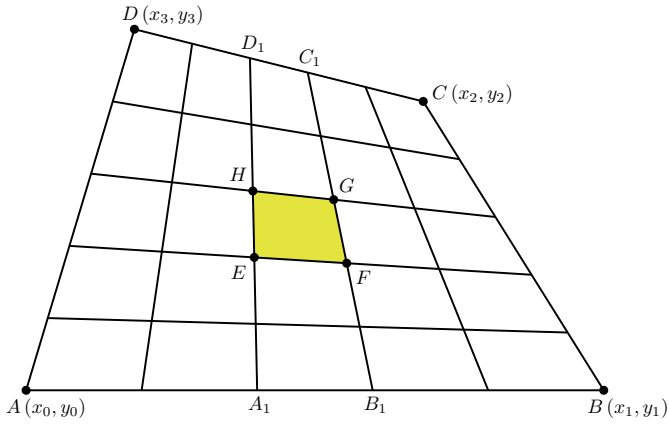

Figure 2

Prove that the area of the quadrilateral $ABCD$ is

$$\begin{aligned}\mathcal{A}(ABCD) &= \frac{1}{2}\,|(x_2-x_0)(y_3-y_1)-(x_3-x_1)(y_2-y_0)| \qquad (1)\\ &= \frac{1}{2}|\,(x_0y_1-y_0x_1)+(x_1y_2-y_1x_2)+(x_2y_3-y_2x_3)\\ &\quad +(x_3y_0-y_3x_0)\,|. \qquad (2)\end{aligned}$$

(b) Application to the posed problem. Determine the coordinates of E, F, G, H and, using formula (2) above, calculate the area of the quadrilateral $EFGH$.

2°) *A general result on the area of a convex polygon.*

Let Π be a compact convex polygon in the plane, whose $n+1$ vertices $A_0, A_1, \ldots, A_n$ are labelled by proceeding counter-clockwise. The Cartesian coordinates of the vertex A_k are (x_k, y_k); the edge (or side) joining A_k to A_{k+1} is denoted by Γ_k. With the help of Green's theorem, prove that the area of Π can be expressed in terms of the coordinates (x_k, y_k) of the vertices as follows:

$$\mathcal{A}(\Pi) = \frac{1}{2}\sum_{k=0}^{n}(x_ky_{k+1}-x_{k+1}y_k); \qquad (3)$$

$$\mathcal{A}(\Pi) = \frac{1}{2}\sum_{k=0}^{n}(x_k+x_{k+1})(y_{k+1}-y_k); \qquad (4)$$

$$\mathcal{A}(\Pi) = \frac{1}{2}\sum_{k=0}^{n}(y_k+y_{k+1})(x_k-x_{k+1}). \qquad (5)$$

In these formulas, we set by convenience $x_{n+1} = x_0$ and $y_{n+1} = y_0$ (that is, the starting vertex A_0 is found both at the start and end of the list of vertices). See Figure 3 below.

Hints. 1°) (b). The points $A_1, B_1, \ldots, E, F, \ldots$ are convex combinations of the initial points A, B, C, D.

2°) Green's theorem (also called the Green–Riemann or even Green–Ostrogradsky theorem) states that, for the convex polygon Π (whose boundary Γ is piecewise-smooth, actually a succession of line-segments (or edges) $\Gamma_0, \Gamma_1, \ldots, \Gamma_n$), and for $\mathcal{C}^1$ functions $(x,y) \mapsto P(x,y)$ and $(x,y) \mapsto Q(x,y)$, we have:

$$\oint_{\Gamma} P(x,y)\,\mathrm{d}x + Q(x,y)\,\mathrm{d}y = \iint_{\Pi}\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)\,\mathrm{d}x\mathrm{d}y.$$

Since the area of Π is $\iint_{\Pi} 1\;\mathrm{d}x\mathrm{d}y$, to express it as a line integral it suffices to choose P and Q such that $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 1$.

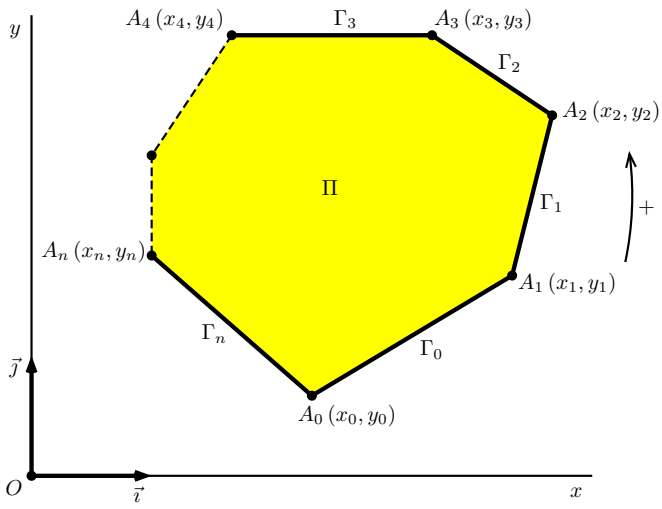

Figure 3

Answers. 1°) (*a*). Consider the quadrilateral formed by the midpoints M, N, P, Q of consecutive sides of the original quadrilateral $ABCD$; see Figure 2. It is easy to check that $MNPQ$ is a parallelogram (usually called VARIGNON's quadrilateral associated with $ABCD$), whose area is exactly half of the area of $ABCD$. Now,

$$\mathcal{A}(MNPQ) = \left\| \overrightarrow{MN} \times \overrightarrow{MQ} \right\|,$$

where $\overrightarrow{MN} \times \overrightarrow{MQ}$ denotes the vector or cross-product of the two vectors $\overrightarrow{MN}$ and $\overrightarrow{MQ}$ (considered as vectors in $\mathbb{R}^3$). But $\overrightarrow{MN}$ is half of the diagonal $\overrightarrow{AC}$ and $\overrightarrow{MQ}$ is half of the diagonal $\overrightarrow{BD}$; consequently

$$\mathcal{A}(ABCD) = \frac{1}{2} \left\| \overrightarrow{AC} \times \overrightarrow{BD} \right\|. \tag{6}$$

Then, it remains to note that the components of the vector $\overrightarrow{AC} \times \overrightarrow{BD}$ $\in \mathbb{R}^3$ are $0, 0, (x_2 - x_0)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_0)$. Hence, formula (1) is proved.

Formula (2) is another way of expressing the same result as in (1); in doing so, the determinants of the four 2×2 matrices enter into the picture

$$\begin{bmatrix} x_0 & y_0 \\ x_1 & y_1 \end{bmatrix}, \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}, \begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \end{bmatrix} \begin{bmatrix} x_3 & y_3 \\ x_0 & y_0 \end{bmatrix}.$$

1°) (b). We have: $A_1 = \left(\frac{3x_0+2x_1}{5}, \cdot\right), ...,$
$E = \frac{1}{5}\left(3\frac{2x_1+3x_0}{5} + 2\frac{2x_2+3x_3}{5}, \cdot\right) = (\frac{9x_0}{25} + \frac{6x_1}{25} + \frac{4x_2}{25} + \frac{6x_3}{25}, \cdot)$, etc.
Consequently

$$\overrightarrow{EG} = \frac{1}{5}(x_2 - x_0, y_2 - y_0); \; \overrightarrow{EG} = \frac{1}{5}(x_3 - x_1, y_3 - y_1)$$

and

$$\begin{aligned} \mathcal{A}(EFGH) &= \frac{1}{2}\left\|\overrightarrow{EG} \times \overrightarrow{FH}\right\| \\ &= \frac{1}{50}\left|(x_2 - x_0)(y_3 - y_1) - (x_3 - x_1)(y_2 - y_0)\right| \\ &= \frac{1}{25}\mathcal{A}(ABCD). \end{aligned}$$

2°) In Green's theorem

$$\oint_\Gamma P(x,y)\ \mathrm{d}x + Q(x,y)\ \mathrm{d}y = \iint_\Pi \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\ \mathrm{d}x\mathrm{d}y,$$

choose for example $P(x,y) = 0$ and $Q(x,y) = x$. Then, the line integral that we have to evaluate is

$$\oint_\Gamma x\ \mathrm{d}y = \oint_{\Gamma_0} x\ \mathrm{d}y + \oint_{\Gamma_1} x\ \mathrm{d}y + ... + \oint_{\Gamma_n} x\ \mathrm{d}y.$$

To compute the k-th line integral above, parametrize the edge Γ_k from A_k to A_{k+1}:

$$t \in [0,1] \mapsto \overrightarrow{\gamma(t)} = (x_k + t(x_{k+1} - x_k), (y_k + t(y_{k+1} - y_k)).$$

Substituting this parametrization into the line integral, we get :

$$\begin{aligned} \oint_{\Gamma_k} x\ \mathrm{d}y &= \int_0^1 (x_k + t(x_{k+1} - x_k))(y_{k+1} - y_k)\ \mathrm{d}t \\ &= \frac{1}{2}(x_k + x_{k+1})(y_{k+1} - y_k). \end{aligned}$$

Consequently,

$$\mathcal{A}(\Pi) = \frac{1}{2}\sum_{k=0}^{n}(x_k + x_{k+1})(y_{k+1} - y_k).$$

Another choice for P and Q could be $P(x,y) = -y$ and $Q(x,y) = 0$. Then

$$\begin{aligned} \oint_{\Gamma_k} (-y)\ \mathrm{d}x &= -\int_0^1 (y_k + t(y_{k+1} - y_k))(x_{k+1} - x_k)\ \mathrm{d}t \\ &= -\frac{1}{2}(x_{k+1} - x_k)(y_{k+1} + y_k), \end{aligned}$$

and

$$\mathcal{A}(\Pi) = -\frac{1}{2}\sum_{k=0}^{n}(x_{k+1} - x_k)(y_{k+1} + y_k).$$

Finally, a "mixed" choice for P and Q could be $P(x, y) = -y/2$ and $Q(x, y) = x/2$. Then,

$$\mathcal{A}(\Pi) = \frac{1}{2}\oint_{\Gamma} x \ \mathrm{d}y - y \ \mathrm{d}x,$$

a familiar formula indeed when expressing an area in terms of a line integral. In the present case,

$$\frac{1}{2}\oint_{\Gamma_k} x \ \mathrm{d}y - y \ \mathrm{d}x = \frac{1}{2}(x_k y_{k+1} - x_{k+1} y_k),$$

so that

$$\mathcal{A}(\Pi) = \frac{1}{2}\sum_{k=0}^{n}(x_k y_{k+1} - x_{k+1} y_k).$$

Comments. 1°) (*b*). - Note the proportionality of diagonals of quadrilaterals $EFGH$ and $ABCD$: $\overrightarrow{EG} = \frac{1}{5}\overrightarrow{AC}$, $\overrightarrow{FH} = \frac{1}{5}\overrightarrow{BD}$. This could have been "intuited", but needed to be proved.

- Along the same lines, one can show that

$$\mathcal{A}(A_1B_1C_1D_1) = \frac{1}{5}\mathcal{A}(ABCD),$$

and that the area of $EFGH$ is the average of the areas of the horizontal (or vertical) neighboring parallelograms.

- We mention another way of proving the announced result by using vector calculus. Suppose that the (oriented) plane is marked with an orthonormal basis $\left(O; \overrightarrow{i}, \overrightarrow{j}\right)$. A parametrization of the surface $ABCD$ is:

$$\begin{aligned} (u, v) &\in [0, 1] \times [0, 1] \mapsto \overrightarrow{\sigma}(u, v) \in \mathbb{R}^2 \\ \overrightarrow{\sigma}(u, v) &= u\left[t\overrightarrow{OA} + (1-t)\overrightarrow{OB}\right] + (1-v)\left[t\overrightarrow{OD} + (1-t)\overrightarrow{OC}\right] \quad (7) \end{aligned}$$

Then, easy vector calculus leads to

$$\begin{aligned} \left\| \frac{\overrightarrow{\partial\sigma}}{\partial u} \times \frac{\overrightarrow{\partial\sigma}}{\partial v} \right\| &= 2uv\mathcal{A}(ADB) + 2(1-u)v\mathcal{A}(DCA) \\ &\quad + 2u(1-v)\mathcal{A}(BAC) + 2(1-u)(1-v)\mathcal{A}(DCB). \quad (8) \end{aligned}$$

Then

$$\begin{aligned}\mathcal{A}(EFGH) &= \iint_{[2/5,3/5]} \left\| \frac{\overrightarrow{\partial\sigma}}{\partial u} \times \frac{\overrightarrow{\partial\sigma}}{\partial v} \right\| \, \mathrm{d}u\mathrm{d}v \\ &= \frac{1}{100} \left[2\mathcal{A}(ADB) + 2\mathcal{A}(DCA) + 2\mathcal{A}(BAC) + 2\mathcal{A}(DCB) \right] \\ &= \frac{1}{50} \left[2\mathcal{A}(ABCD) \right].\end{aligned}$$

- As one might imagine, there is nothing special about the number 5 in the proved result; if each of the sides of the quadrilateral $ABCD$ is divided into $2n+1$ equal parts, then the area of the "central" quadrilateral is $\mathcal{A}/\left(2n+1\right)^2$.

2°) To keep in mind formula (3), observe that it contains a succession of determinants of 2×2 matrices $\begin{bmatrix} x_k & y_k \\ x_{k+1} & y_{k+1} \end{bmatrix}$, $k = 0, 1, ..., n$. So, one proceeds as in lacing up shoes (when they have laces!)

$$\begin{bmatrix} x_0 \searrow & \swarrow y_0 \\ x_1 \searrow & \swarrow y_1 \\ x_2 \searrow & \swarrow y_2 \\ \cdots & \cdots \\ x_n \searrow & \swarrow y_n \\ x_0 & y_0 \end{bmatrix}.$$

This is why formula (3) is sometimes called the *shoelace formula*.

---

220. ★★★ *Is $\sqrt{f}$ differentiable whenever the nonnegative f is differentiable?*

Let $f : \mathbb{R} \to [0, +\infty)$ be a nonnegative differentiable function. The questions we would like to answer in this Tapa are: is $\sqrt{f}$ differentiable? Is $\sqrt{f}$ of class $\mathcal{C}^1$ whenever f is of class $\mathcal{C}^2$? Of course, these questions are posed only in the case where there are points a where $f(a) = 0$ (such points are called *zeros* of f).

1°) Warm-up. With the help of simple polynomial functions f, show that $\sqrt{f}$ can be differentiable or not differentiable at zeros of f.

2°) Suppose that f is twice differentiable at a, a zero of f. Prove that $\sqrt{f}$ is differentiable at a if and only if $f''(a) = 0$.

3°) Suppose that f is of class $\mathcal{C}^2$ on $\mathbb{R}$. Prove that $\sqrt{f}$ is of class $\mathcal{C}^1$ on $\mathbb{R}$ if and only if $f''(a) = 0$ at all zeros a of f.

Hints. 2°) - Of course, $\sqrt{f}$ is differentiable at any point x where $f(x) > 0$; at such points x, we have $\left(\sqrt{f}\right)'(x) = \frac{f'(x)}{2\sqrt{f(x)}}$.

- If a is a zero of f, it is a minimizer of f; thus: $f'(a) = 0$ and $f''(a) \geqslant 0$ [these are first and second-order necessary conditions for optimality].

- Besides that, if $\sqrt{f}$ is differentiable at a zero a of f, then $\left(\sqrt{f}\right)'(a) = 0$ [again because a is a minimizer of $\sqrt{f}$].

3°) The main point is to prove the continuity of $\left(\sqrt{f}\right)'$ at zeros a of f. For that purpose, it is suggested to prove the following inequality: For all $\alpha > 0$ and all $x \in [a - \alpha, a + \alpha]$,

$$|f'(x)|^2 \leqslant 2f(x)M_2(\alpha), \tag{1}$$

where $M_2(\alpha) = \sup\{|f''(x)|, x \in [a - 2\alpha, a + 2\alpha]\}$.

Answers. 1°) Let $f(x) = x^2$; then $\sqrt{f}(x) = |x|$ is not differentiable at $a = 0$, the only zero of f.

Let $f(x) = x^4$; then $\sqrt{f}(x) = x^2$ is differentiable at $a = 0$.

Another example is $f(x) = \left(x^2 - 1\right)^2$. At the two zeros of f, namely -1 and 1, $\sqrt{f}(x) = \left|x^2 - 1\right|$ has only a left-derivative -2 and a right-derivative 2.

2°) Since a is a minimizer of f, we have $f'(a) = 0$. Because f has been assumed twice differentiable at a, we have the second-order Taylor–Young expansion below:

$$f(a + h) = \frac{f''(a)}{2}h^2 + h^2\varepsilon(h),$$

with $\lim_{h\to 0} \varepsilon(h) = 0$. As a consequence on the difference quotient involving the $\sqrt{f}$ function,

$$\frac{\sqrt{f}(a + h) - \sqrt{f}(a)}{h} = \frac{|h|}{h}\sqrt{\frac{f''(a)}{2} + \varepsilon(h)}. \tag{2}$$

Here, beware of $|h|/h$, which gives the sign of $h \neq 0$. From (2) above, it is now clear that the considered difference quotient has a limit when $h \to 0$ if and only if $f''(a) = 0$. In that case, as expected, $\left(\sqrt{f}\right)'(a) = 0$.

Here is the proved result in a global statement: *Assuming that the non-negative f is twice differentiable on $\mathbb{R}$, we have that $\sqrt{f}$ is differentiable on $\mathbb{R}$ if and only if $f''(a) = 0$ at all zeros a of f.*

3°) The trivial implication. If $\sqrt{f}$ is of class $\mathcal{C}^1$ on $\mathbb{R}$, then it is differentiable on $\mathbb{R}$ and, according to the previous result (Question 2°)), $f''(a) = 0$ at any point a where $f(a) = 0$.

The converse implication, a bit more difficult to prove. Assume that f is of class $\mathcal{C}^2$ on $\mathbb{R}$ and that $f''(a) = 0$ at any a where $f(a) = 0$. We intend to prove that $\left(\sqrt{f}\right)'$ is a continuous function on $\mathbb{R}$. Clearly, $\left(\sqrt{f}\right)' = \frac{f'}{2\sqrt{f}}$ is continuous at any x where $f(x) > 0$.

Therefore, let a be a zero f. We begin by proving the following inequality: For all $\alpha > 0$ and all $x \in [a - \alpha, a + \alpha]$,

$$|f'(x)|^2 \leqslant 2f(x)M_2(\alpha), \tag{1'}$$

where $M_2(\alpha) = \max\{|f''(x)|, x \in [a - 2\alpha, a + 2\alpha]\}$. Note that $M_2(\alpha)$ is well-defined and that $M_2(\alpha) \to f''(a) = 0$ when $\alpha \to 0$ (since f'' is a continuous function).

When $x \in [a - \alpha, a + \alpha]$ and $h \in [-\alpha, \alpha]$, a second-order Taylor–Lagrange expansion of f around x gives rise to:

$$f(x + h) = f(x) + f'(x)h + \frac{f''(c)}{2}h^2, \tag{3}$$

where c lies in the line-segment joining x to $x + h$.
Hence, $c \in [a - 2\alpha, a + 2\alpha]$, so that it readily follows from (3) that

$$0 \leqslant f(x + h) \leqslant f(x) + f'(x)h + \frac{M_2(\alpha)}{2}h^2 = P(h). \tag{4}$$

Beware, however, that $h \in [-\alpha, \alpha]$.

The derivative of the convex trinomial function $P(h)$ vanishes at $\overline{h} = -f'(x)/M_2(\alpha)$. But $f'(x) = |x - a| \cdot |f''(c_1)|$, where c_1 is a point between a and x. Thus

$$|\overline{h}| = \frac{|f'(x)|}{M_2(\alpha)} = \frac{|x - a| \cdot |f''(c_1)|}{M_2(\alpha)} \leqslant \alpha,$$

and, according to (4),

$$f(x) - \frac{|f'(x)|^2}{2M_2(\alpha)} = P(\overline{h}) \geqslant f(x + \overline{h}) \geqslant 0.$$

Another way of obtaining this is to note the following: The polynomial $P(h)$ is nonnegative *for all* h in $\mathbb{R}$ (not only for $h \in [-\alpha, \alpha]$); hence its discriminant is $\leqslant 0$.

The inequality (1') has therefore been proved.

Conclusive step: We have to check that $\left(\sqrt{f}\right)'(x) \to \left(\sqrt{f}\right)'(a) = 0$ when $x \to a$. Two cases occur:

- If $f(x) = 0$, then $\left(\sqrt{f}\right)'(x) = 0$;

- If $f(x) > 0$, then

$$\left|\left(\sqrt{f}\right)'(x)\right| = \frac{|f'(x)|}{2\sqrt{f(x)}} \leqslant \sqrt{\frac{M_2(\alpha)}{2}}.$$

When $x \to a$, we can choose $\alpha \to 0$. We conclude that $\left(\sqrt{f}\right)'(x) \to 0$ when $x \to a$.

Comments. - The result of Question 3°) is G. Glaeser's theorem (1963) for functions of a real variable.

- 2°) We have actually proved that, at a zero a of f, the function $\sqrt{f}$ has a right-derivative $\sqrt{\frac{f''(a)}{2}}$ and a left-derivative $-\sqrt{\frac{f''(a)}{2}}$. An illustrative example is $f(x) = \left(x^2-1\right)^2$ and $a = \pm 1$.

- 3°) Contrary to what one might imagine, even with f of class $\mathcal{C}^2$ on $\mathbb{R}$, one just gets that $\sqrt{f}$ is $\mathcal{C}^1$ (not necessarily $\mathcal{C}^2$) on $\mathbb{R}$. Indeed, there are counterexamples with f of class $\mathcal{C}^\infty$ on $\mathbb{R}$ and $\sqrt{f}$ just $\mathcal{C}^1$ (not $\mathcal{C}^2$) on $\mathbb{R}$.

- In the same vein, let us mention the following nice theorem of J. Boman (1967): The function $f : \mathbb{R}^n \to \mathbb{R}$ is of class $\mathcal{C}^\infty$ if and only if $f \circ c$ is of class $\mathcal{C}^\infty$ for every function $c : \mathbb{R} \to \mathbb{R}^n$ of class $\mathcal{C}^\infty$.

- One might think of extending the result of this Tapa to f^α with $0 < \alpha < 1/2$. Unfortunately, there exist nonnegative functions f, of class $\mathcal{C}^\infty$ on $\mathbb{R}$, whose derivatives all vanish at zeros of f, and still f^α is not $\mathcal{C}^1$ on $\mathbb{R}$.

- Another facet of differentiability questions on powers of functions is Joris' theorem; see Tapa 294 in

J.-B. Hiriart-Urruty, *Mathematical tapas*, Volume 1 (for Undergraduates). Springer (2016).

---

221. ★★★ *An amazingly simple descent method for minimizing a function*

We are going to consider an algorithmic procedure to solve the following minimization problem:

$$(\mathcal{P}) \quad \left\{ \begin{array}{c} \text{Minimize } f(x) \\ x \in \mathbb{R}^n. \end{array} \right.$$

We suppose, throughout, that:

- f is differentiable on $\mathbb{R}^n$ with a gradient mapping satisfying a LIPSCHITZ condition with constant $L > 0$;
- f is bounded from below on $\mathbb{R}^n$ ($\overline{f}$ denotes $\inf_{x \in \mathbb{R}^n} f(x)$).

1°) Recall quickly why the following "squeezing property" on f holds true: For all x and d in $\mathbb{R}^n$,

$$f(x) + [\nabla f(x)]^T d - \frac{L}{2} \|d\|^2 \leqslant f(x+d) \tag{1-1}$$

$$\leqslant f(x) + [\nabla f(x)]^T d + \frac{L}{2} \|d\|^2 . \tag{1-2}$$

2°) Consider the following algorithmic procedure:

$$\left\{ \begin{array}{l} \text{Initial point } x_0; \\ \text{For all } k,\ x_{k+1} = x_k - \frac{1}{L} \nabla f(x_k). \end{array} \right. \tag{2}$$

(a) Show that, for all integers k,

$$f(x_{k+1}) \leqslant f(x_k) - \frac{1}{2L} \|\nabla f(x_k)\|^2 , \tag{3-1}$$

$$\|\nabla f(x_k)\|^2 \leqslant 2L \left[f(x_k) - f(x_{k+1}) \right] , \tag{3-2}$$

and then, for all integers $N \geqslant 1$,

$$\sum_{k=0}^{N-1} \|\nabla f(x_k)\|^2 \leqslant 2L \left[f(x_0) - f(x_N) \right] . \tag{4}$$

(b) Deduce from (4): $\nabla f(x_k) \to 0$ when $k \to +\infty$.

(c) Derive from (4):

$$\min_{k=0,1,..,N-1} \|\nabla f(x_k)\| \leqslant \sqrt{\frac{2L \left[f(x_0) - \overline{f} \right]}{N}}, \tag{5}$$

i.e., " the smallest gradient encountered in the first N iterations shrinks like $1/\sqrt{N}$ ".

3°) We suppose, moreover, that f is convex and that problem $(\mathcal{P})$ has a solution $\overline{x}$ (whence $f(\overline{x}) = \overline{f}$).

Prove that, for all integers $N \geqslant 1$,

$$0 \leqslant f(x_N) - \overline{f} \leqslant \frac{L}{2N} \|x_0 - \overline{x}\|^2 . \tag{6}$$

Conclusion concerning the sequence $(f(x_k))_k$?

Hints. 1°) This squeezing property on f has been proved in Tapa 157.

3°) Use the following inequality, deduced from the convexity of f (the graph of f is above any of its tangent lines),

$$f(\overline{x}) \geqslant f(x_k) + [\nabla f(x_k)]^T (\overline{x} - x_k)$$

and simple calculus rules like

$$\|x_k + u - \overline{x}\|^2 = \|x_k - \overline{x}\|^2 + \|u\|^2 + 2(x_k - \overline{x})^T u.$$

Answers. 2°) (*a*). Using the right-part of inequality (1) applied to $x = x_k$ and $d = -\frac{1}{L}\nabla f(x_k)$, we get

$$f(x_{k+1}) \leqslant f(x_k) - \frac{1}{2L} \|\nabla f(x_k)\|^2, \tag{3-1}$$

which is equivalent to

$$\|\nabla f(x_k)\|^2 \leqslant 2L\,[f(x_k) - f(x_{k+1})]. \tag{3-2}$$

So, this amazingly simple algorithm is a *descent method*, in the sense that

$$f(x_{k+1}) < f(x_k) \text{ if } x_k \text{ is not a critical point of } f \text{ (if } \nabla f(x_k) \neq 0).$$

The decreasing sequence $(f(x_k))_k$ is bounded from below by $\overline{f}$, hence it converges to some value $\widetilde{f}$.

By writing (3-2) for $k = 0$ to $k = N - 1$ and summing up (the "telescoping effect"), we obtain

$$\sum_{k=0}^{N-1} \|\nabla f(x_k)\|^2 \leqslant 2L\,[f(x_0) - f(x_N)]. \tag{4}$$

(*b*) We deduce from (4): For all positive integers N,

$$\sum_{k=0}^{N} \|\nabla f(x_k)\|^2 \leqslant 2L\,\left[f(x_0) - \overline{f}\right].$$

Hence, the numerical series with general term $\|\nabla f(x_k)\|^2$ is convergent; thus, $\nabla f(x_k) \to 0$ as $k \to +\infty$.

(c) Still with (4),

$$N \times \min_{k=0,1,..,N-1} \|\nabla f(x_k)\|^2 \leqslant 2L\left[f(x_0) - \overline{f})\right],$$

and so

$$\min_{k=0,1,..,N-1} \|\nabla f(x_k)\| \leqslant \sqrt{\frac{2L\left[f(x_0) - \overline{f})\right]}{N}}. \tag{5}$$

3°) Due to the convexity of f,

$$\begin{array}{rcl} f(\overline{x}) & \geqslant & f(x_k) + [\nabla f(x_k)]^T (\overline{x} - x_k), \\ f(x_k) & \leqslant & \overline{f} - [\nabla f(x_k)]^T (\overline{x} - x_k). \end{array}$$

With (3-1), we obtain

$$f(x_{k+1}) \leqslant \overline{f} - [\nabla f(x_k)]^T (\overline{x} - x_k) - \frac{1}{2L} \|\nabla f(x_k)\|^2. \tag{7}$$

But $x_{k+1} - \overline{x} = x_k - \overline{x} - \frac{1}{L}\nabla f(x_k)$, so that

$$-\frac{L}{2}\left[\|x_{k+1} - \overline{x}\|^2 - \|x_k - \overline{x}\|^2\right] = -[\nabla f(x_k)]^T (\overline{x} - x_k) - \frac{1}{2L} \|\nabla f(x_k)\|^2.$$

We therefore derive from (7):

$$f(x_{k+1}) - \overline{f} \leqslant \frac{L}{2}\left[\|x_k - \overline{x}\|^2 - \|x_{k+1} - \overline{x}\|^2\right]. \tag{8}$$

By summing up (8) written for $k = 0, 1, .., N-1$, we obtain

$$\begin{array}{rcl} \displaystyle\sum_{k=0}^{N-1} \left[f(x_{k+1}) - \overline{f}\right] & \leqslant & \dfrac{L}{2}\left[\|x_0 - \overline{x}\|^2 - \|x_N - \overline{x}\|^2\right] \\ & \leqslant & \dfrac{L}{2}\|x_0 - \overline{x}\|^2. \end{array} \tag{9}$$

Since the sequence $\left(f(x_k) - \overline{f}\right)_k$ is decreasing, we infer from (9):

$$f(x_N) - \overline{f} \leqslant \frac{1}{N} \sum_{k=0}^{N-1} \left[f(x_{k+1}) - \overline{f}\right] \leqslant \frac{L}{2N} \|x_0 - \overline{x}\|^2.$$

As an immediate consequence, $f(x_k)$ converges to $\overline{f}$ as $k \to +\infty$.

Comments. - Even if, when $k \to +\infty$, $f(x_k)$ converges to some value $\widetilde{f}$ (actually the minimal value $\overline{f}$ when f is convex) and $\nabla f(x_k) \to 0$, we have not exhibited any results concerning the convergence of the sequence (x_k) itself. However, with additional assumptions on f, like its so-called strong convexity (*i.e.,* $f - \alpha \|\cdot\|^2$ is still convex for some $\alpha > 0$), one can prove the following:
$f(x_{k+1}) - \overline{f} \leqslant M \left[f(x_k) - \overline{f} \right]$ for some $M > 0$ depending on L and α; the sequence (x_k) converges towards the unique solution $\overline{x}$ in $(\mathcal{P})$.
- This Tapa 219 is in a sense the discrete version of what has been proposed in Tapa 213.

---

222. ★★★ *Two simple approaches to the "uniform boundedness principle"*

Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be two normed vector spaces, let $A : E \to F$ be a continuous linear mapping. Recall that the "operator norm" $N(A)$ (encountered several times in this collection of Tapas) is defined as

$$N(A) = \sup_{\|x\|_E \leqslant 1} \|A(x)\|_F \,, \tag{1}$$

or, equivalently, as the greatest lower bound of those $L \geqslant 0$ satisfying:

$$\|A(x)\|_F \leqslant L \|x\|_E \text{ for all } x \in E. \tag{2}$$

Clearly, if $(A_\alpha)_{\alpha \in A}$ is a family of continuous linear mappings from E into F, and if, for some constant $C > 0$,

$$N(A_\alpha) \leqslant C \text{ for all } \alpha \in A, \tag{3}$$

then, for any $x \in E$,

$$\|A_\alpha(x)\|_F \leqslant C \|x\|_E \text{ for all } \alpha \in A. \tag{4}$$

Inequality (3) expresses that $(A_\alpha)_{\alpha \in A}$ is *norm-bounded*, while (4) expresses that $(A_\alpha)_{\alpha \in A}$ is just *pointwise-bounded*. The so-called "uniform boundedness principle" is devoted to proving that (4) implies (3) under a suitable assumption on E. Its precise statement is as follows:

Assume that $(E, \|\cdot\|_E)$ is a Banach *space. Let $(A_\alpha)_{\alpha \in A}$ be a family of continuous linear mappings from E into F. We suppose that for all $x \in E$, there exists a $C_x > 0$ such that*

$$\|A_\alpha(x)\|_F \leqslant C_x \|x\|_E \text{ for all } \alpha \in A. \tag{5}$$

There then exists a $C > 0$ such that

$$N(A_\alpha) \leqslant C \text{ for all } \alpha \in A, \tag{6}$$

or, in other words,

$$\|A_\alpha(x)\|_F \leqslant C \|x\|_E \text{ for all } \alpha \in A \text{ and all } x \in E.$$

1°) First simple approach.

(a) A preparatory lemma. Let $(A_\alpha)_{\alpha \in A}$ be a family of continuous linear mappings from E into F. Suppose that there exists a closed ball $\overline{B(x_0, r)}$ and a constant $K > 0$ such that

$$\|A_\alpha(x)\|_F \leqslant K \text{ for all } x \in \overline{B(x_0, r)} \text{ and all } \alpha \in A. \tag{7}$$

Then check that

$$N(A_\alpha) \leqslant \frac{2K}{r} \text{ for all } \alpha \in A. \tag{8}$$

(b) Proof of the "uniform boundedness principle".

The idea of the proof is by contradiction. Suppose that $N(A_\alpha)$ may be as large as desired. Then, according to the preliminary lemma (Question (a)), for any ball $\overline{B(x_0, r)}$ and any $K > 0$, there exists $x \in \overline{B(x_0, r)}$ and $\alpha \in A$ such that $\|A_\alpha(x)\|_F > K$.

- Construct inductively a sequence of nested closed balls $\overline{B(x_k, r)} = S_k$ and mappings A_{α_k} such that

$$\|A_{\alpha_k}(x)\|_F \geqslant k \text{ for all } x \in S_k.$$

- Conclude, with the help of a point x^* belonging to the intersection of the S_k's, that this leads to a contradiction.

2°) Second simple approach.

(a) A preparatory lemma. Let A be a continuous linear mapping from E into F. Then

$$\sup_{\|x' - x\|_E \leqslant r} \|A(x')\|_F \geqslant rN(A). \tag{9}$$

(b) Proof of the "uniform boundedness principle".

The idea of the proof is again by contradiction. Suppose that $N(A_\alpha)$ may be as large as desired. Choose therefore a sequence $(A_{\alpha_n})_{n \geqslant 1}$ such that $N(A_{\alpha_n}) \geqslant 4^n$ for all n.

- Start with $x_0 = 0$, and for $n \geqslant 1$, choose inductively $x_n \in E$ such that

$$\begin{aligned} \|x_n - x_{n-1}\|_E &\leqslant \frac{1}{3^n}, \\ \|A_{\alpha_n}(x_n)\|_F &\geqslant \frac{2}{3}\frac{1}{3^n}\, N(A_{\alpha_n}). \end{aligned}$$

- Conclude with a contradiction concerning the limit point x^* of the sequence (x_n).

Hints. The completeness property of E is essential here. It is used in two ways:

- In the first proof. Use the property: For any sequence of closed balls $S_1 \supset S_2 ... \supset S_k \supset ...$, their intersection is nonempty. When the diameter of these balls tends to 0, the intersection is even reduced to a single point; see Tapa 118.

- In the second proof. Use the property: Any CAUCHY sequence is convergent.

Answers. 1) (a). Let $x \neq 0$ be an arbitrary element in E. Then, x_0 and $x_0 + \frac{r}{\|x\|}x$ belong to $\overline{B(x_0, r)}$. Consequently, one derives from (7):

$$\begin{aligned} K &\geqslant \left\| A_\alpha \left(x_0 + \frac{r}{\|x\|_E} x \right) \right\|_F = \left\| A_\alpha(x_0) + \frac{r}{\|x\|_E} A_\alpha(x) \right\|_F \\ &\geqslant \frac{r}{\|x\|_E} \|A_\alpha(x)\|_F - K \quad \text{(due to the triangle inequality).} \end{aligned}$$

Hence,

$$\|A_\alpha(x)\|_F \leqslant \frac{2K}{r} \|x\|_E .$$

Since this inequality holds true for all $x \neq 0$, the inequality (8) follows.

(b) Consider a first ball $\overline{B(x_0, r_0)}$. Since $\{\|A_\alpha(x)\|_F : \alpha \in A\}$ is unbounded on $\overline{B(x_0, r)}$, there exists $x_1 \in \overline{B(x_0, r_0)} = S_0$ and $\alpha_1 \in A$ such that $\|A_{\alpha_1}(x_1)\|_F > 1$. Due to the continuity of A_{α_1}, there exists a ball $\overline{B(x_1, r_1)} = S_1$ contained in S_0 such that $\|A_{\alpha_1}(x)\|_F > 1$ for all $x \in S_1$. But $\{\|A_\alpha(x)\|_F : \alpha \in A\}$ is also unbounded on S_1. There therefore exists $x_2 \in S_1$ and $\alpha_2 \in A$ such that $\|A_{\alpha_2}(x_2)\|_F > 2$. Due to the continuity of A_{α_2}, there exists a ball $\overline{B(x_2, r_2)} = S_2$ contained in S_1 such that $\|A_{\alpha_2}(x)\|_F > 2$ for all $x \in S_2$. And so on. We finally

get a sequence of closed balls $S_1 \supset S_2 ... \supset S_n \supset ...$ and a sequence $(A_{\alpha_k}(x_k))_k$, such that

$$\|A_{\alpha_k}(x)\|_F > k \text{ for all } x \in S_k.$$

The intersection of the S_k's is nonempty. Take a point x^* belonging to this intersection. Then, according to the inequality above,

$$\|A_{\alpha_k}(x^*)\|_F > k \text{ for all } k,$$

which contradicts the assumption (5) on the boundedness of

$$\left\{\|A_{\alpha_k}(x^*)\|_F : \alpha \in A\right\}.$$

2) (a). Let $u \in E$ satisfying $\|u\|_E \leqslant 1$. Then, due to the linearity of A,

$$A(x) = \frac{1}{2}A(x+u) + \frac{1}{2}A(x-u).$$

Consequently,

$$\|A(x)\|_F \leqslant 2\frac{1}{2}\sup_{\|v\|\leqslant r}\|A(x+v)\|_F.$$

Hence, remembering the definition (1) or (2) of $N(A)$, the inequality (9) is proved.

(b) Choose a sequence $(A_{\alpha_n})_{n\geqslant 1}$ such that $N(A_{\alpha_n}) \geqslant 4^n$ for all n. We make use of the preparatory lemma written in the following form:

$$\sup_{\|x'-x_{n-1}\|_E \leqslant \frac{1}{3^n}} \|A_{\alpha_n}(x')\|_F \geqslant \frac{1}{3^n}\ N(A_{\alpha_n}).$$

Start with $x_0 = 0$, and for $n \geqslant 1$, use the inequality above to inductively choose $x_n \in E$ such that

$$\|x_n - x_{n-1}\|_E \leqslant \frac{1}{3^n}, \tag{10}$$

$$\|A_{\alpha_n}(x_n)\|_F \geqslant \frac{2}{3}\frac{1}{3^n}\ N(A_{\alpha_n}). \tag{11}$$

It clearly follows from (10) that $(x_n)_{n\geqslant 1}$ is a CAUCHY sequence, hence convergent to some $x^* \in E$. Still using (10), one gets an upper bound of $\|x_q - x_p\|$, $q > p$,

$$\|x_q - x_p\| \leqslant \sum_{k=p}^{q-1}\|x_{k+1} - x_k\| \leqslant \sum_{k=p}^{+\infty}\frac{1}{3^{p+1}}$$

and, passing to the limit (on q) in this inequality, one derives

$$\|x^* - x_p\| \leqslant \frac{1}{2}\frac{1}{3^p}.$$

Finally,

$$\begin{aligned}
\|A_{\alpha_n}(x^*)\|_F &= \|A_{\alpha_n}(x^* - x_n) + A_{\alpha_n}(x_n)\|_F \\
&\geqslant \|A_{\alpha_n}(x_n)\|_F - \|A_{\alpha_n}(x^* - x_n)\|_F \\
&\geqslant \|A_{\alpha_n}(x_n)\|_F - N(A_{\alpha_n})\,\|x^* - x_n\|_E \\
&\geqslant \left(\frac{2}{3} - \frac{1}{2}\right)\frac{1}{3^n}\; N(A_{\alpha_n}) \\
&\geqslant \frac{1}{6}\left(\frac{4}{3}\right)^n,
\end{aligned}$$

which contradicts the assumption (5) on the boundedness of

$$\left\{\|A_{\alpha_n}(x^*)\|_F : \alpha \in A\right\}.$$

Comments. - The "uniform boundedness principle" is one of the cornerstones of Functional analysis. It is usually proved as an application of a form of the Baire category theorem (that we alluded to in Tapa 190). As often in mathematics, to prove a theorem, it depends on what is known, what results and techniques are at disposal. Here, only simple properties in Banach spaces are used. The idea of the first proof (construction of a nested sequence of closed balls) is taken from

V. Trénoguine, *Analyse fonctionnelle [Functional analysis].* Editions Mir (1985),

while that in the second one (construction of a Cauchy sequence) comes from

A. Sokal, *A really simple elementary proof of the uniform boundedness theorem.* The American Math. Monthly 118, 450–452 (2011).

Part 4. Open problems

"Some problems open doors, some problems close doors, and some remain curiosities, but all sharpen our wits and act as a challenge and a test of our ingenuity and techniques"

M. Atiyah (1929–)

223. ♣♣♣ *Open question on the sequence of successive prime numbers*

We know that there are infinitely many prime numbers $p_1 = 2, p_2 = 3, p_3 = 5 ... , p_n < p_{n+1}, ...$ Many properties of this sequence of ordered prime numbers p_n are known, for example:
- The gap between p_n and its successor p_{n+1} may be as large as desired;
- After p_n, we don't have to wait for $2p_n$ to meet the next prime number p_{n+1} (in other words, $p_{n+1} < 2p_n$).

Open question: Do we have

$$\sqrt{p_{n+1}} - \sqrt{p_n} \leqslant 1 \text{ for all } n?$$

In 1985, D. Andrica conjectured that the answer is Yes. Indeed, the suggested inequality has been checked up to $n = 4 \times 10^{18}$; it seems highly likely that it is true for all n, but no mathematical proof has been proposed.

---

224. ♣♣♣ *Open question on the* Euler–Mascheroni *constant*

The so-called Euler–Mascheroni (or just Euler) constant is a mathematical constant recurring in Number theory and Analysis, usually denoted by γ (gamma). We have encountered it, for example, in Tapas 60 and 77 listed in

J.-B. Hiriart-Urruty, *Mathematical tapas*, Volume 1 (for Undergraduates). Springer (2016).

Its basic definition is:

$$\gamma = \lim_{n \to +\infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \ln n \right).$$

Open question: Is γ rational or irrational?

While it is fairly easy to prove that the mathematical constants e and π are irrational, the answer for the posed question on γ is unknown and seems out of reach... There is plenty of information on what γ could be or should not be; see the book referenced below. For example, if γ is a fraction p/q (with p and q positive integers), then the denominator q must be greater than 10^{242080}.

J. HAVIL, *Gamma. Exploring* EULER*'s constant*. Princeton University Press (2003).

---

225. ♣♣♣ *Open question on the* HADAMARD *matrices*

A HADAMARD (or SYLVESTER–HADAMARD) matrix is an $n \times n$ matrix whose entries are either $+1$ or -1 and whose distinct row vectors are orthogonal. In geometric terms, this means that each pair of row vectors of a HADAMARD matrix H represents two orthogonal vectors in $\mathbb{R}^n$, while the length of each row is $\sqrt{n}$; in matricial terms, we have: $H \cdot H^T = nI_n$. Clearly, HADAMARD matrices may be changed into other HADAMARD matrices by just permuting rows and columns, and by multiplying rows and columns by -1.

The following is easy to check (this is a necessary condition for existence): If a HADAMARD matrix of order $n > 3$ exists, then n should be of the form $4k$ for some positive integer k.

Open question: Is there a HADAMARD matrix of every order $n = 4k$? (In other words, is the above necessary condition sufficient?)

As yet, the answer has been proved to be Yes for all $n = 4k < 668$ and in some cases for larger n... but not for all of them.

For more on HADAMARD matrices and their applications (especially in Signal processing, Coding and Cryptography), we refer to the following book:

K.J. HORADAM, HADAMARD *matrices and their applications*. Princeton University Press (2007).

---

226. ♣♣♣ *Open question on the determinant of normal matrices*

Let A and B be two normal $n \times n$ matrices with prescribed eigenvalues $a_1, .., a_n$ and $b_1, .., b_n$, respectively.

Open question: Does

$$\det(A-B) \in \operatorname{co}\left\{\prod_{i=1}^{n}(a_i - b_{\sigma(i)}) : \sigma \text{ permutation of } \{1,2,...,n\}\right\}? \tag{1}$$

Recall that a matrix M is said to be normal if there exists a unitary matrix U such that $U^*MU = \operatorname{diag}(\lambda_1, ..., \lambda_n)$, where the λ_i's are eigenvalues of M. Hermitian, skew-Hermitian and unitary matrices are examples of normal matrices. Note, however, that $A-B$ is not necessarily normal even if both A and B are.

MARCUS (in 1973) and DE OLIVEIRA (in 1982) conjectured that the answer to the posed question (the OMC conjecture, for short) is Yes.

Indeed, the OMC conjecture is known to be true in several situations: when $n \leqslant 3$; when both A and B are Hermitian; when both A and B are unitary; when A is positive definite and B is skew-Hermitian; when A is positive definite and B is a non-real scalar multiple of a Hermitian matrix; when A is Hermitian and B is a non-real scalar multiple of a Hermitian matrix. For an easy way to grasp the meaning of (1), think of the two following specific cases:

- When both A and B are Hermitian (hence with *real* eigenvalues $a_1, .., a_n$ and $b_1, .., b_n$ respectively); then $\det(A - B)$ lies in the line-segment

$$\left[\min \prod_{i=1}^{n}(a_i - b_{\sigma(i)}), \max \prod_{i=1}^{n}(a_i - b_{\sigma(i)})\right],$$

where the min and the max are taken over all the permutations σ of $\{1, 2, ..., n\}$.

- When both A and B commute: in this case, there is a unitary matrix which simultaneously diagonalizes A and B, so that $\det(A-B)$ indeed equals $\prod_{i=1}^{n}(a_i - b_{\sigma(i)})$ for some permutation σ of $\{1, 2, ..., n\}$.

However the answer to the general question (1) is still unknown.

---

227. ♣♣♣ *An open global volume minimization problem*

This is a 3D-counterpart of (the Comments in) Tapa 123.

Consider a compact convex set C in the 3-dimensional space $\mathbb{R}^3$, with nonempty interior. The support function σ_C of C is defined as follows:

$$d \in \mathbb{R}^3 \mapsto \sigma_C(d) = \max_{c \in C} c^T d. \tag{1}$$

We call the width (or thickness) of C in the unitary direction $\overrightarrow{u} \in \mathbb{R}^3$ the quantity

$$e_C(\overrightarrow{u}) = \max_{c \in C} c^T \overrightarrow{u} - \min_{c \in C} c^T \overrightarrow{u} = \sigma_C(\overrightarrow{u}) + \sigma_C(-\overrightarrow{u}). \tag{2}$$

In geometrical terms, $e_C(\overrightarrow{u})$ represents the distance between two parallel planes orthogonal to $\overrightarrow{u}$ "squeezing" C (or tangent to C if the contact points are smooth).

We say that C is a convex set of constant width (or thickness) $\Delta > 0$ when $e_C(\overrightarrow{u}) = \Delta$ for all unitary vectors $\overrightarrow{u}$ in $\mathbb{R}^3$. If necessary, the condition can be parameterized by using spherical (or other) coordinates for unit vectors $\overrightarrow{u}$ in $\mathbb{R}^3$. Such compact convex sets of constant width in $\mathbb{R}^3$ are sometimes called *spheroforms* in the literature.

Open question:
What are the convex sets (in) $\mathbb{R}^3$ of constant width Δ and of minimal volume? (3)

In spite of the nice properties of the volume function (it is continuous, its 3-rd root is concave, etc.), we do not know all the solutions of the above minimization problem. What is known about it?

- It indeed has solutions: the continuity of the volume function and the usual topologies on families of compact convex sets are sufficient ingredients to ensure the existence of solutions. From one solution, one gets another one by rotating it.

- By taking the constant width Δ equal to 1, a theoretical lower bound for the volume is known to be ≈ 0.365.

- The candidates which are suspected to provide an answer to question (3) are the so-called MEISSNER spheroforms (sort of inflated tetrahedrons, whose curved faces contain pieces of spheres and toroidal parts).

---

228. ♣♣♣ *A shortest curve problem in three-dimensional space*

Warm-up in the plane. Consider a boat, lost at sea, whose captain knows that it is located 1 mile from the shore (in a straight line, as indicated by the measuring instruments), but the fog is so heavy that he is unable to assess the direction to take. The boat moves at a constant speed and the objective is to touch the seaside as soon as possible; so the question is: What is the path of minimal length that the boat should follow to be sure to touch or meet the seaside? In mathematical terms: *Given a circle centered at* $(0,0)$ *of radius 1, what is the curve of minimal length, drawn from the origin* $(0,0)$*, touching or meeting any tangent line to the circle?* By "a curve" we mean here "a continuous rectifiable curve". This question is completely solved, see the reference below and the works quoted therein.

The 3-dimensional version of the problem. The question posed above has a 3-dimensional version, but it is decidedly more difficult to handle. Open question:

Given a sphere centered at $(0,0,0)$ *of radius* 1*, what curve, emanating from the origin* $(0,0,0)$*, has minimal length and touches or meets each tangent plane to the sphere?*

Admittedly, we know nothing about the structure of the optimal path. Repeated considerations and collegian discussions have produced quite a few feasible solutions, some of a rather strange shape.

Another problem, similar to the previous one, posed by a colleague while reacting to and wrestling with the above challenging problem, is the next one:

What is the curve of minimal length in space which can be seen from any point on the earth?

This also remains an open question.

J.-B. Hiriart-Urruty, *A new series of conjectures and open questions in optimization and matrix analysis.* ESAIM: Control, Optimisation and Calculus of Variations 15, 454–470 (2009).

---

229. ♣♣♣ M. Crouzeix's *conjecture in Matrix theory*

Let $\mathbb{C}^n$ be equipped with the usual Hermitian inner product

$$(u = (u_1, ..., u_n) \in \mathbb{C}^n,\ v = (v_1, ..., v_n) \in \mathbb{C}^n) \mapsto \langle u, v\rangle = \sum_{i=1}^{n} \overline{u_i} v_i,$$

and the associated norm $u \mapsto \|u\| = \sqrt{\langle u, u\rangle}$.

For $A = [a_{ij}] \in \mathcal{M}_n(\mathbb{C})$, we set

$$\mathcal{H}(A) = \left\{ \langle u, Au\rangle = \sum_{i,j=1}^{n} a_{ij} u_i \overline{u_j} : \|u\| = 1 \right\}. \tag{1}$$

This set is called the *hausdorffian set* or *numerical range* or *field of values* of A.

Examples of determinations of $\mathcal{H}(A)$ as well as some of its properties have been displayed in Tapa 129.

Crouzeix's conjecture (2004) is one of the most intriguing questions in Matrix theory, it deals precisely with the numerical range of matrices. Crouzeix proved that, for any polynomial p and any matrix $A \in \mathcal{M}_n(\mathbb{C})$,

$$\nu_2 [p(A)] \leqslant 11.08 \times \max_{z \in \mathcal{H}(A)} |p(z)|. \tag{2}$$

Here $\nu_2 [M]$ stands for the 2-norm (or spectral norm) of the matrix M, that is

$$\nu_2 [M] = \sup_{\mathbb{R}^n \ni x \neq 0} \frac{\|Mx\|}{\|x\|} = \max \{\text{singular values of } M\}.$$

Open question:
Can the constant 11.08 *in inequality (2) be lowered to* 2?

Crouzeix conjectured that the answer is Yes. Indeed, it is true for any matrix $M \in \mathcal{M}_2(\mathbb{C})$ and for the so-called normal matrices $M \in \mathcal{M}_n(\mathbb{C})$ (normal matrices are those which can be diagonalized with the help of a unitary matrix). Actually, for normal matrices M, $\mathcal{H}(M)$ is the convex

polygon in $\mathbb{C}$ generated by the eigenvalues λ_i of M, that is to say, the convex hull of $\lambda_1, ..., \lambda_n$ (see Tapa 129 if necessary); in that case, it is easy to check that

$$\nu_2 \left[p(M) \right] \leqslant \max_{z \in \mathcal{H}(M)} |p(z)| . \tag{3}$$

So, the open question is posed for non-normal matrices.

Some final comments:

- If Crouzeix's conjecture holds true for polynomial functions p, it will also hold for all analytic functions p.

- Approaches by optimization and various numerical experiments strongly suggest that the answer to the open question is Yes. From the theoretical side, Crouzeix and Palencia proved recently (2017) that the inequality (2) could be greatly improved, replacing 11.08 with $1+\sqrt{2} \approx 2.414$.

- This is Michel Crouzeix, a mathematician who worked mainly in Numerical analysis... His brother Jean-Pierre Crouzeix is also a mathematician, whose works concern more Quasi-convex analysis and Optimization.

---

230. ♣♣♣ *A simple equivalent form of* Riemann's *open problem*

One of the most famous open problems (or conjectures) in mathematics, if not the most famous one today, is due to Riemann (1859) (one also speaks of the Riemann hypothesis). In its basis or classical form, it expresses that Riemann's ζ function (the analytic extension of the holomorphic function defined for all complex numbers z, except 1, of the function of the complex variable $z \mapsto \zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^z}$) has all of its non-trivial zeros located on the line with equation $\mathrm{Re}(z) = 1/2$. One of the fascinating aspects of this conjecture is that it can be related to various different domains in mathematics (see the paper quoted below). Also very astonishing are the *equivalent* forms of this conjecture; here is the one described by Lagarias (2002).

Let

$$\begin{aligned} H_n &= \sum_{i=1}^{n} \frac{1}{k} \text{ (the so-called harmonic numbers);} \\ \sigma(n) &= \sum_{d \text{ divides } n} d \text{ (the sum of all divisors of } n). \end{aligned}$$

For example, $\sigma(3) = 1 + 3 = 4$, $\sigma(16) = 1 + 2 + 4 + 8 + 16 = 31$.

Then, an equivalent form of Riemann's conjecture turns out to be:

$$\text{For all } n > 1, \sigma(n) \leqslant H_n + \exp(H_n)\ln(H_n). \tag{1}$$

Open question: *Does inequality (1) hold true or not?*

M. Balazard, *Un siècle et demi de recherches sur l'hypothèse de* Riemann *[One and a half centuries of research on the* Riemann *hypothesis].* La Gazette des mathématiques 126 (Bulletin of the French Mathematical Society), 7–24 (2010).

Epilogue

"You are never sure whether or not a problem is good unless you actually solve it"
M. Gromov (1943–)

References and sources

Journals

Section "Problems and solutions" of *The American Mathematical Monthly*
http://www.maa.org/publications/periodicals/american-mathematical-monthly

Rubrique "Questions et réponses" de la *Revue de la Filière Mathématique*
(ex-Revue de Mathématiques Spéciales, RMS in short)
http://www.rms-math.com/

Revue *Quadrature*
http://www.quadrature.info/

Books

G. Polya and J. Kilpatrick, *The Stanford mathematics problem book*, Dover editions (2009).

J.-B. Hiriart-Urruty, *Optimisation et Analyse convexe [Optimization and Convex analysis] (Summaries of results; exercises and problems with solutions).* Collection Enseignement SUP Mathématiques, Editions EDP Sciences, Paris (2009), 344 pages.
New printing of a book published in 1998 by another publishing house.

D. Azé, G. Constans and J.-B. Hiriart-Urruty, *Calcul différentiel et équations différentielles [Differential calculus and Differential equations] (Exercises and problems with solutions).* Collection Enseignement SUP Mathématiques, Editions EDP Sciences, Paris (2010), 224 pages.
New printing of a book published in 1998 by another publishing house.

D. Azé and J.-B. Hiriart-Urruty, *Analyse variationnelle et optimisation [Variational analysis and Optimization] (Summaries of lectures; exercises and problems with solutions).* Editions Cépaduès, Toulouse (2010), 332 pages.