Grundlehren Text Editions

JEAN-BAPTISTE HIRIART-URRUTY
CLAUDE LEMARÉCHAL

FUNDAMENTALS OF CONVEX ANALYSIS

凸分析基础

Springer
世界图书出版公司

Jean-Baptiste Hiriart-Urruty
Claude Lemaréchal

Fundamentals of Convex Analysis

With 66 Figures

Springer
世界图书出版公司

书　　名：Fundamentals of Convex Analysis
作　　者：Jean-Baptiste Hiriart-Urruty Claude Lemaréchal
中 译 名：凸分析基础
出 版 者：世界图书出版公司北京公司
印 刷 者：北京世图印刷厂
发　　行：世界图书出版公司北京公司（北京朝内大街 137 号　100010）
联系电话：010-64015659, 64038347
电子信箱：kjsk@vip.sina.com
开　　本：24 开　　　印　张：11.5
出版年代：2004 年 4 月
书　　号：7-5062-6617-2/O·470
版权登记：图字: 01-2004-1164
定　　价：38.00 元

Preface

This book is an abridged version of our two-volume opus *Convex Analysis and Minimization Algorithms* [18], about which we have received very positive feedback from users, readers, lecturers ever since it was published – by Springer-Verlag in 1993. Its pedagogical qualities were particularly appreciated, in the combination with a rather advanced technical material.

Now [18] has a dual but clearly defined nature:

– an introduction to the basic concepts in convex analysis,
– a study of convex minimization problems (with an emphasis on numerical algorithms),

and insists on their mutual interpenetration. It is our feeling that the above basic introduction is much needed in the scientific community. This is the motivation for the present edition, our intention being to create a tool useful to teach convex analysis. We have thus extracted from [18] its "backbone" devoted to convex analysis, namely Chaps III–VI and X. Apart from some local improvements, the present text is mostly a copy of the corresponding chapters. The main difference is that we have deleted material deemed too advanced for an introduction, or too closely attached to numerical algorithms.

Further, we have included exercises, whose degree of difficulty is suggested by 0, 1 or 2 stars *. Finally, the index has been considerably enriched.

Just as in [18], each chapter is presented as a "lesson", in the sense of our old masters, treating of a given subject in its entirety. After an introduction presenting or recalling elementary material, there are five such lessons:

– A Convex sets (corresponding to Chap. III in [18]),
– B Convex functions (Chap. IV in [18]),
– C Sublinearity and support functions (Chap. V),
– D Subdifferentials in the finite-valued case (VI),
– E Conjugacy (X).

Thus, we do not go beyond conjugacy. In particular, subdifferentiability of extended-valued functions is intentionally left aside. This allows a lighter book, easier to master and to go through. The same reason led us to skip duality which, besides, is more related to optimization. Readers interested by these topics can always read the relevant chapters in [18] (namely Chaps XI and XII).

During the French Revolution, the writer of a bill on public instruction complained: "Le défaut ou la disette de bons ouvrages élémentaires a été, jusqu'à présent, un des plus grands obstacles qui s'opposaient au perfectionnement de l'instruction. La raison de cette disette, c'est que jusqu'à présent les savants d'un mérite éminent ont, presque toujours, *préféré la gloire d'élever l'édifice de la science à la peine d'en éclairer l'entrée.*[1]" Our main motivation here is precisely to "light the entrance" of the monument Convex Analysis. This is therefore not a reference book, to be kept on the shelf by experts who already know the building and can find their way through it; it is far more a book for the purpose of learning and teaching. We call above all on the intuition of the reader, and our approach is very gradual. Nevertheless, we keep constantly in mind the suggestion of A. Einstein: "Everything should be made as simple as possible, but not simpler". Indeed, the content is by no means elementary, and will be hard for a reader not possessing a firm mastery of basic mathematical skill.

We could not completely avoid cross-references between the various chapters; but for many of them, the motivation is to suggest an intellectual link between apparently independent concepts, rather than a technical need for previous results. More than a tree, our approach evokes a spiral, made up of loosely interrelated elements.

Many sections are set in smaller characters. They are by no means reserved to advanced material; rather, they are there to help the reader with illustrative examples and side remarks, that help to understand a delicate point, or prepare some material to come in a subsequent chapter. Roughly speaking, sections in smaller characters can be compared to footnotes, used to avoid interrupting the flow of the development; it can be helpful to skip them during a deeper reading, with pencil and paper. They can often be considered as additional informal exercises, useful to keep the reader alert.

The numbering of sections restarts at 1 in each chapter, and chapter numbers are dropped in a reference to an equation or result from within the same chapter.

Toulouse and Grenoble,
March 2001

J.-B. Hiriart-Urruty, C. Lemaréchal

[1] "The lack or scarcity of good, elementary books has been, until now, one of the greatest obstacles in the way of better instruction. The reason for this scarcity is that, until now, scholars of great merit have almost always preferred the glory of constructing the monument of science over the effort of lighting its entrance." D. Guedj: *La Révolution des Savants*, Découvertes, Gallimard Sciences (1988) 130 – 131.

Contents

Preface V

0. Introduction: Notation, Elementary Results 1
1 Some Facts About Lower and Upper Bounds 1
2 The Set of Extended Real Numbers 5
3 Linear and Bilinear Algebra 6
4 Differentiation in a Euclidean Space 9
5 Set-Valued Analysis 12
6 Recalls on Convex Functions of the Real Variable 14
Exercises 16

A. Convex Sets 19
1 Generalities 19
1.1 Definition and First Examples 19
1.2 Convexity-Preserving Operations on Sets 22
1.3 Convex Combinations and Convex Hulls 26
1.4 Closed Convex Sets and Hulls 31
2 Convex Sets Attached to a Convex Set 33
2.1 The Relative Interior 33
2.2 The Asymptotic Cone 39
2.3 Extreme Points 41
2.4 Exposed Faces 43
3 Projection onto Closed Convex Sets 46
3.1 The Projection Operator 46
3.2 Projection onto a Closed Convex Cone 49
4 Separation and Applications 51
4.1 Separation Between Convex Sets 51
4.2 First Consequences of the Separation Properties 54
– Existence of Supporting Hyperplanes 54
– Outer Description of Closed Convex Sets 55
– Proof of Minkowski's Theorem 57
– Bipolar of a Convex Cone 57
4.3 The Lemma of Minkowski-Farkas 58
5 Conical Approximations of Convex Sets 62

5.1 Convenient Definitions of Tangent Cones 62
5.2 The Tangent and Normal Cones to a Convex Set 65
5.3 Some Properties of Tangent and Normal Cones 68
Exercises 70

B. Convex Functions 73
1 Basic Definitions and Examples 73
1.1 The Definitions of a Convex Function 73
1.2 Special Convex Functions: Affinity and Closedness 76
– Linear and Affine Functions 77
– Closed Convex Functions 78
– Outer Construction of Closed Convex Functions 80
1.3 First Examples 82
2 Functional Operations Preserving Convexity 87
2.1 Operations Preserving Closedness 87
2.2 Dilations and Perspectives of a Function 89
2.3 Infimal Convolution 92
2.4 Image of a Function Under a Linear Mapping 96
2.5 Convex Hull and Closed Convex Hull of a Function 98
3 Local and Global Behaviour of a Convex Function 102
3.1 Continuity Properties 102
3.2 Behaviour at Infinity 106
4 First- and Second-Order Differentiation 110
4.1 Differentiable Convex Functions 110
4.2 Nondifferentiable Convex Functions 114
4.3 Second-Order Differentiation 115
Exercises 117

C. Sublinearity and Support Functions 121
1 Sublinear Functions 123
1.1 Definitions and First Properties 123
1.2 Some Examples 127
1.3 The Convex Cone of All Closed Sublinear Functions 131
2 The Support Function of a Nonempty Set 134
2.1 Definitions, Interpretations 134
2.2 Basic Properties 136
2.3 Examples 140
3 Correspondence Between Convex Sets and Sublinear Functions 143
3.1 The Fundamental Correspondence 143
3.2 Example: Norms and Their Duals, Polarity 146
3.3 Calculus with Support Functions 151
3.4 Example: Support Functions of Closed Convex Polyhedra .. 158
Exercises 161

D. Subdifferentials of Finite Convex Functions 163
1 The Subdifferential: Definitions and Interpretations 164
1.1 First Definition: Directional Derivatives 164
1.2 Second Definition: Minorization by Affine Functions 167
1.3 Geometric Constructions and Interpretations 169
2 Local Properties of the Subdifferential 173
2.1 First-Order Developments 173
2.2 Minimality Conditions 177
2.3 Mean-Value Theorems 178
3 First Examples 180
4 Calculus Rules with Subdifferentials 183
4.1 Positive Combinations of Functions 183
4.2 Pre-Composition with an Affine Mapping 184
4.3 Post-Composition with an Increasing Convex Function of Several Variables 185
4.4 Supremum of Convex Functions 188
4.5 Image of a Function Under a Linear Mapping 191
5 Further Examples 194
5.1 Largest Eigenvalue of a Symmetric Matrix 194
5.2 Nested Optimization 196
5.3 Best Approximation of a Continuous Function on a Compact Interval 198
6 The Subdifferential as a Multifunction 199
6.1 Monotonicity Properties of the Subdifferential 199
6.2 Continuity Properties of the Subdifferential 201
6.3 Subdifferentials and Limits of Subgradients 204
Exercises 205

E. Conjugacy in Convex Analysis 209
1 The Convex Conjugate of a Function 211
1.1 Definition and First Examples 211
1.2 Interpretations 214
1.3 First Properties 216
– Elementary Calculus Rules 216
– The Biconjugate of a Function 218
– Conjugacy and Coercivity 219
1.4 Subdifferentials of Extended-Valued Functions 220
2 Calculus Rules on the Conjugacy Operation 222
2.1 Image of a Function Under a Linear Mapping 222
2.2 Pre-Composition with an Affine Mapping 224
2.3 Sum of Two Functions 227
2.4 Infima and Suprema 229
2.5 Post-Composition with an Increasing Convex Function 231
3 Various Examples 233
3.1 The Cramer Transformation 234

3.2 The Conjugate of Convex Partially Quadratic Functions 234
3.3 Polyhedral Functions 235
4 Differentiability of a Conjugate Function 237
4.1 First-Order Differentiability 238
4.2 Lipschitz Continuity of the Gradient Mapping 240
Exercises 241

Bibliographical Comments 245

The Founding Fathers of the Discipline 249

References 251

Index 253

0. Introduction: Notation, Elementary Results

We start this chapter by listing some basic concepts, which are or should be well-known – but it is good sometimes to return to basics. This gives us the opportunity of making precise the system of notation used in this book. For example, some readers may have forgotten that "i.e." means *id est*, the literal translation of "that is (to say)". If we get closer to mathematics, $S\backslash\{x\}$ denotes the set obtained by depriving a set S of a point $x \in S$. We also mention that, if f is a function, $\overset{-1}{f}(y)$ is the *inverse image* of y, i.e. the set of all points x such that $f(x) = y$. When f is invertible, this set is the singleton $\{f^{-1}(y)\}$.

After these basic recalls, we prove some results on convex functions of one real variable. They are just as basic, but are easily established and will be of some use in this book.

1 Some Facts About Lower and Upper Bounds

1.1 In the *totally ordered* set $\mathbb{R}$, $\inf E$ and $\sup E$ are respectively the greatest lower bound – the *infimum* – and least upper bound – the *supremum* – of a nonempty subset E, when they exist (as real numbers). Then, they may or may not belong to E; when they do, a more accurate notation is $\min E$ and $\max E$. Whenever the relevant infima exist, the following relations are clear enough:

$$\left.\begin{array}{c} \inf(E \cup F) = \min\{\inf E, \inf F\}, \\ F \subset E \implies \inf F \geqslant \inf E, \\ \inf(E \cap F) \geqslant \max\{\inf E, \inf F\}. \end{array}\right\} \tag{1.1}$$

If E is characterized by a certain property P, we use the notation

$$E = \{r \in \mathbb{R} \,:\, r \text{ satisfies } P\}.$$

Defining (in $\mathbb{R}$ considered as a real *vector space*) the standard operations on nonempty sets

$$\begin{array}{c} E + F := \{r = e + f \,:\, e \in E, f \in F\}, \\ tE := \{tr \,:\, r \in E\} \quad \text{for } t \in \mathbb{R} \end{array}$$

(the sign ":=" means "equals by definition"), it is also clear that

$$\left.\begin{array}{c}\inf(E+F)=\inf E+\inf F\,,\\ \inf tE=t\inf E \quad \text{if } t>0\,,\\ \inf(-E)=-\sup E\,,\end{array}\right\} \tag{1.2}$$

whenever the relevant extrema exist.

The word *positive* means "> 0", and *nonpositive* therefore means "$\leqslant 0$"; same conventions with negative and nonnegative. The set of nonnegative numbers is denoted by $\mathbb{R}^+$ and, generally speaking, a substar deprives a set of the point 0. Thus, for example,

$$\mathbb{N}_* = \{1,2,\ldots\} \quad \text{and} \quad \mathbb{R}^+_* = \{t \in \mathbb{R} : t > 0\}\,.$$

Squared brackets are used to denote the intervals of $\mathbb{R}$: for example,

$$\mathbb{R} \supset]a,b] = \{t \in \mathbb{R} : a < t \leqslant b\}\,.$$

The symbol "$\downarrow$" means convergence from the right, *the limit being excluded*; thus, $t \downarrow 0$ means $t \to 0$ in $\mathbb{R}^+_*$. The words "increasing" and " decreasing" are taken in a broad sense: a sequence (t_k) is increasing when $k > k' \Rightarrow t_k \geqslant t_{k'}$. We use the notation (t_k), or $(t_k)_k$, or $(t_k)_{k\in\mathbb{N}_*}$ for a sequence of elements $t_1, t_2, \ldots$

1.2 Now, to denote a real-valued function f defined on a nonempty set X, we write

$$X \ni x \mapsto f(x) \in \mathbb{R}.$$

The *sublevel-set* of f at level $r \in \mathbb{R}$ is defined by

$$\mathrm{S}_r(f) := \{x \in X : f(x) \leqslant r\}\,.$$

If two functions f and g from X to $\mathbb{R}$ satisfy

$$f(x) \leqslant g(x) \quad \text{for all } x \in X\,,$$

we say that f *minorizes* g (on X), or that g majorizes f.

Computing the number

$$\inf\{f(x) : x \in X\} =: \bar{f} \tag{1.3}$$

represents a minimization problem posed in X: namely that of finding a so-called *minimizing sequence*, i.e. $(x_k) \subset X$ such that $f(x_k) \to \bar{f}$ when $k \to +\infty$ (note that no structure is assumed on X). In other words, $\bar{f}$ is the largest lower bound $\inf f(X)$ of the subset $f(X) \subset \mathbb{R}$, and will often be called the infimal value, or more simply the *infimum* of f on X. Another notation for (1.3) is $\inf_{x\in X} f(x)$, or also $\inf_X f$. The function f is usually called the objective function, or also infimand. We can also meet supremands, minimands, etc.

From the relations (1.1), (1.2), we deduce (hereafter, $\bar{f}_i$ denotes the infimum of f over X_i for $i = 1, 2$):

$$\begin{gathered}
\inf\{f(x) : x \in X_1 \cup X_2\} = \min\{\bar{f}_1, \bar{f}_2\}, \\
X_1 \subset X_2 \quad \Longrightarrow \quad \bar{f}_1 \geqslant \bar{f}_2, \\
\inf\{f(x) : x \in X_1 \cap X_2\} \geqslant \max\{\bar{f}_1, \bar{f}_2\}, \\
\inf\{f(x_1) + f(x_2) : x_1 \in X_1 \text{ and } x_2 \in X_2\} = \bar{f}_1 + \bar{f}_2, \\
\inf\{tf(x) : x \in X\} = t\bar{f}, \quad \text{for } t \geqslant 0, \\
\inf\{-f(x) : x \in X\} = -\sup\{f(x) : x \in X\},
\end{gathered} \tag{1.4}$$

whenever the relevant extrema exist. The last relation is used very often.

The attention of the reader is drawn to (1.4), perhaps the only non-totally trivial among the above relations. Calling $E_1 := f(X_1)$ and $E_2 := f(X_2)$ the *images* of X_1 and X_2 under f, (1.4) represents the sum of the infima $\inf E_1$ and $\inf E_2$. There could just as well be two different infimands, i.e. (1.4) could be written more suggestively

$$\inf\{f(x_1) + g(x_2) : x_1 \in X_1 \text{ and } x_2 \in X_2\} = \bar{f}_1 + \bar{g}_2$$

(g being another real-valued function). This last relation must not be confused with

$$\inf\{f(x) + g(x) : x \in X\} \geqslant \bar{f} + \bar{g};$$

here, in the language of (1.4), $X_1 = X_2 = X$, but only the image by f of the diagonal of $X \times X$ is considered.

Another relation requiring some attention is the *decoupling*, or transitivity, of infima: if g sends the Cartesian product $X \times Y$ to $\mathbb{R}$, then

$$\begin{aligned}
&\inf\{g(x,y) : x \in X \text{ and } y \in Y\} = \\
&= \inf_{x\in X}[\inf_{y \in Y} g(x,y)] = \inf_{y\in Y}[\inf_{x\in X} g(x,y)].
\end{aligned} \tag{1.5}$$

1.3 An *optimal solution* of (1.3) is an $\bar{x} \in X$ such that

$$f(\bar{x}) = \bar{f} \leqslant f(x) \quad \text{for all } x \in X;$$

such an $\bar{x}$ is often called a minimizer, a minimum point, or more simply a *minimum* of f on X. We will also speak of *global minimum*. To say that there exists a minimum is to say that the inf in (1.3) is a min; the infimum $\bar{f} = f(\bar{x})$ can then be called the minimal value. The notation

$$\min\{f(x) : x \in X\}$$

is the same as (1.3), and says that there does exist a solution; we stress the fact that this notation – as well as (1.3) – represents at the same time a *number* and a *problem to solve*. It is sometimes convenient to denote by

$$\operatorname{Argmin}\{f(x) : x \in X\}$$

the set of optimal solutions of (1.3), and to use "argmin" if the solution is unique.

It is worth mentioning that the decoupling property (1.5) has a translation in terms of Argmin's. More precisely, the following properties are easy to see:

– If $(\bar{x}, \bar{y})$ minimizes g over $X \times Y$, then $\bar{y}$ minimizes $g(\bar{x}, \cdot)$ over Y and $\bar{x}$ minimizes over X the function

$$\varphi(x) := \inf \{g(x, y) \, : \, y \in Y\} .$$

– Conversely, if $\bar{x}$ minimizes φ over X and if $\bar{y}$ minimizes $g(\bar{x}, \cdot)$ over Y, then $(\bar{x}, \bar{y})$ minimizes g over $X \times Y$.

Needless to say, symmetric properties are established, interchanging the roles of x and y.

1.4 In our context, X is equipped with a topology; actually X is a subset of some finite-dimensional real vector space, call it $\mathbb{R}^n$; the topology is then that induced by a norm. The *interior* and *closure* of X are denoted by $\operatorname{int} X$ and $\operatorname{cl} X$ respectively; its boundary is $\operatorname{bd} X$.

The concept of limit is assumed familiar. We recall that the *limes inferior* (in the ordered set $\mathbb{R}$) is the smallest cluster point.

Remark 1.1 The standard terminology is *lower limit* ("abbreviated" as lim inf!) This terminology is unfortunate, however: a limit must be a well-defined unique element; otherwise, expressions such as "$f(x)$ has a limit" are ambiguous. □

Thus, to say that $\ell = \liminf_{x \to x^*} f(x)$, with $x^* \in \operatorname{cl} X$, means: for all $\varepsilon > 0$,

$$\text{there is a neighborhood } N(x^*) \text{ such that } f(x) \geqslant \ell - \varepsilon \text{ for all } x \in N(x^*) ,$$

and

$$\text{in any neighborhood } N(x^*), \text{ there is } x \in N(x^*) \text{ such that } f(x) \leqslant \ell + \varepsilon ;$$

in particular, if $x^* \in X$, we certainly have $\ell \leqslant f(x^*)$.

Let $x^* \in X$. If $f(x^*) \leqslant \liminf_{x \to x^*} f(x)$, then f is said to be *lower semi-continuous* (l.s.c) at x^*; *upper semi-continuity*, which means $f(x^*) \geqslant \limsup f(x)$, is not much used in our context. It is well-known that, if X is a compact set on which f is continuous, then the lower bound $\bar{f}$ exists and (1.3) has a solution. Actually, lower semi-continuity (of f on the whole compact X) suffices: if (x_k) is a minimizing sequence, with some cluster point $x^* \in X$, we have

$$f(x^*) \leqslant \liminf_{k \to \infty} f(x_k) = \lim_{k \to \infty} f(x_k) = \bar{f} .$$

Another observation is: let E be such that $\operatorname{cl} E \subset X$; if f is continuous on $\operatorname{cl} E$, then

$$\inf \{f(x) \, : \, x \in E\} = \inf \{f(x) \, : \, x \in \operatorname{cl} E\} .$$

This relation is wrong if f is only l.s.c, though: then, only (1.1) gives useful relations.

Related with (1.3), another problem is whether a given minimizing sequence (x_k) converges to an optimal solution when $k \to +\infty$. This problem is really distinct from (1.3): for example, with $X := \mathbb{R}$, $f(0) := 0$, $f(x) := 1/|x|$ for $x \neq 0$, the sequence defined by $x_k = k$ is minimizing but does not converge to the minimum 0 when $k \to +\infty$.

2 The Set of Extended Real Numbers

In convex analysis, there are serious reasons for wanting to give a meaning to (1.3), for arbitrary f and X. For this, two additional elements are appended to $\mathbb{R}$: $+\infty$ and $-\infty$.

If $E \subset \mathbb{R}$ is nonempty but unbounded from above, we set $\sup E = +\infty$; similarly, $\inf E = -\infty$ if E is unbounded from below. Then consider the case of an empty set: to maintain a relation such as (1.1)

$$\inf (E \cup \emptyset)[= \inf E] = \min \{\inf E, \inf \emptyset\} \quad \text{for all } \emptyset \neq E \subset \mathbb{R}\,,$$

we have no choice and we set $\inf \emptyset = +\infty$. Naturally, $\sup \emptyset = -\infty$, and this maintains the relation $\inf(-E) = -\sup E$ in (1.2).

It should be noted that the world of convex analysis is not symmetric, it is *unilateral*. In particular, $+\infty$ and $-\infty$ do not play the same role, and it suffices for our purpose to consider the set $\mathbb{R} \cup \{+\infty\}$. Extending the notation of the intervals of $\mathbb{R}$, this set will also be denoted by $]-\infty, +\infty]$.

To extend the structure of $\mathbb{R}$ to this new set, the natural rules are adopted:

$$\begin{array}{rc} \text{order:} & x \leqslant +\infty \text{ for all } x \in \mathbb{R} \cup \{+\infty\}\,; \\ \text{addition:} & (+\infty) + x = x + (+\infty) = +\infty \text{ for all } x \in \mathbb{R} \cup \{+\infty\}\,; \\ \text{multiplication:} & t \cdot (+\infty) = +\infty \text{ for all } 0 < t \in \mathbb{R} \cup \{+\infty\}\,. \end{array}$$

Thus, we see that

- the structured set $(\mathbb{R} \cup \{+\infty\}, +)$ is not a group, just because $+\infty$ has no opposite;
- it is a fortiori not a field, a second reason being that we avoid writing $t \times (+\infty)$ for $t \leqslant 0$.

On the other hand, we leave it to the reader to check that the other axioms are preserved (for the order, the addition and the multiplication); so some calculus can at least be done in $\mathbb{R} \cup \{+\infty\}$.

Actually, $\mathbb{R} \cup \{+\infty\}$ is nothing more than an *ordered convex cone*, analogous to the set $\mathbb{R}_*^+$ of positive numbers. In particular, observe the following continuity properties:

$$\begin{array}{lcl} (x_k, y_k) \to (x, y) \text{ in } [\mathbb{R} \cup \{+\infty\}]^2 & \Longrightarrow & x_k + y_k \to x + y \text{ in } \mathbb{R} \cup \{+\infty\}\,; \\ (t_k, x_k) \to (t, x) \text{ in } \mathbb{R}_*^+ \times (\mathbb{R} \cup \{+\infty\}) & \Longrightarrow & t_k x_k \to tx \text{ in } \mathbb{R} \cup \{+\infty\}\,. \end{array}$$

In this book, starting from Chap. B, the minimization problems of §1 – and in particular (1.3) – will be understood as posed in $\mathbb{R} \cup \{+\infty\}$. The advantage of this is to give a systematic meaning to all the relations of §1. On the other hand, the reader should not feel too encumbered by this new set, which takes the place of the familiar set of real numbers where algebra is "easy". First of all, $\mathbb{R} \cup \{+\infty\}$ is relevant only as far as images of functions are concerned: any algebraic manipulations involving no term $f(x)$ is "safe" and requires no special attention. When some $f(x)$ is involved, the following pragmatic attitude can be adopted:

- comparison and addition: no problems in $\mathbb{R} \cup \{+\infty\}$, just as in $\mathbb{R}$;
- subtraction: before subtracting $f(x)$, make sure that $f(x) < +\infty$;
- multiplication: think of a term like $tf(x)$ as the multiplication of the vector $f(x)$ by the scalar t; if $t \leqslant 0$, make sure that $f(x) < +\infty$ (note: in convex analysis, the product of functions $f(x)g(x)$ is rarely used, and multiplication by -1 puts (1.3) in a different world);
- division: same problems as in $\mathbb{R}$, namely avoid division by 0;
- convergence: same problems as in $\mathbb{R}$, namely pay attention to $\infty - \infty$ and $0 \cdot (+\infty)$;
- in general, do not overuse expressions like $tf(x)$ with $t \leqslant 0$, or $r - f(x)$, etc.: they do not fit well with the conical structure of $\mathbb{R} \cup \{+\infty\}$.

3 Linear and Bilinear Algebra

3.0 Let us start with the *model-situation* of $\mathbb{R}^n$, the real n-dimensional vector space of n-uples $x = (\xi^1, \ldots, \xi^n)$. In this space, the vectors $e_1, \ldots, e_n$, where each e_i has coordinates $(0, \ldots, 0, 1, 0, \ldots, 0)$ (the "1" in i^{th} position) form a basis, called the *canonical* basis. The linear mappings from $\mathbb{R}^m$ to $\mathbb{R}^n$ are identified with the $n \times m$ *matrices* which represent them in the canonical bases; *vectors* of $\mathbb{R}^n$ are thus naturally identified with $n \times 1$ matrices.

The space $\mathbb{R}^n$ is equipped with the canonical, or standard, Euclidean structure with the help of the scalar product

$$x = (\xi^1, \ldots, \xi^n),\ y = (\eta^1, \ldots, \eta^n) \quad \mapsto \quad x^\top y := \sum_{i=1}^n \xi^i \eta^i$$

(also denoted by $x \cdot y$). Then we can speak of the Euclidean space $(\mathbb{R}^n, {}^\top)$.

3.1 More generally, a *Euclidean space* is a real vector space, say X, of *finite dimension*, say n, equipped with a *scalar product* denoted by $\langle \cdot, \cdot \rangle$. Recall that a scalar (or inner) product is a bilinear symmetric mapping $\langle \cdot, \cdot \rangle$ from $X \times X$ to $\mathbb{R}$, satisfying $\langle x, x \rangle > 0$ for $x \neq 0$.

(a) If a basis $\{b_1, \ldots, b_n\}$ has been chosen in X, along which two vectors x and y have the coordinates $(\xi^1, \ldots, \xi^n)$ and $(\eta^1, \ldots, \eta^n)$, we have

$$\langle x, y \rangle = \sum_{i,j=1}^n \xi^i \eta^j \langle b_i, b_j \rangle .$$

This can be written $\langle x, y \rangle = x^\top Q y$, where Q is a symmetric positive definite $n \times n$ matrix ($\mathrm{S}_n(\mathbb{R})$ will denote the set of symmetric matrices). In this situation, to equip X with a scalar product is actually to take a symmetric positive definite matrix.

The simplest matrix Q is the identity matrix I, or I_n, which corresponds to the scalar product

$$\langle x, y\rangle = x^\top y = \sum_{i=1}^{n} \xi^i \eta^i ,$$

called the *dot-product*. For this particular product, one has $\langle b_i, b_j\rangle = \delta_{ij}$ (δ_{ij} is the symbol of Kronecker: $\delta_{ij} = 0$ if $i \neq j$, $\delta_{ii} = 1$). The basis $\{b_1, \ldots, b_n\}$ is said to be *orthonormal* for this scalar product; and this scalar product is of course the only one for which the given basis is orthonormal.

Thus, whenever we have a basis in X, we know all the possible ways of equipping X with a Euclidean structure.

(b). Reasoning in the other direction, let us start from a Euclidean space $(X, \langle\cdot,\cdot\rangle)$ of dimension n. It is possible to find a basis $\{b_1, \ldots, b_n\}$ of X, which is orthonormal for the given scalar product (i.e. which satisfies $\langle b_i, b_j\rangle = \delta_{ij}$ for $i, j = 1, \ldots, n$). If two vectors x and y are expressed in terms of this basis, $\langle x, y\rangle$ can be written $x^\top y$.

Use the space $\mathbb{R}^n$ of §3.0 and denote by $\varphi : \mathbb{R}^n \to X$ the unique *linear operator* (isomorphism of vector spaces) satisfying $\varphi(e_i) = b_i$ for $i = 1, \ldots, n$. Then

$$x^\top y = \langle \varphi(x), \varphi(y)\rangle \quad \text{for all } x \text{ and } y \text{ in } \mathbb{R}^n ,$$

so the Euclidean structure is also carried over by φ, which is therefore an isomorphism of Euclidean spaces as well. Thus, *any* Euclidean space $(X, \langle\cdot,\cdot\rangle)$ of dimension n is *isomorphic* to $(\mathbb{R}^n, ^\top)$, which explains the importance of this last space. However, given a Euclidean space, an orthonormal basis need not be easy to construct; said otherwise, one must sometimes content oneself with a scalar product imposed by the problem considered.

Example 3.1 Vector spaces of matrices form a rich field of applications for the techniques and results of convex analysis. The set of $p \times q$ matrices forms a vector space of dimension pq, in which a natural scalar product of two matrices M and N is ($\operatorname{tr} A := \sum_{i=1}^n A_{ii}$ is the *trace* of the $n \times n$ matrix A)

$$\langle M, N\rangle := \operatorname{tr} M^\top N = \sum_{i=1}^{p} \sum_{j=1}^{q} M_{ij} N_{ij} . \qquad \square$$

(c). A subspace V of $(X, \langle\cdot,\cdot\rangle)$ can be equipped with the Euclidean structure defined by

$$V \times V \ni (x, y) \mapsto \langle x, y\rangle .$$

Unless otherwise specified, we will generally use this *induced* structure, with the same notation for the scalar product in V and in X.

More importantly, let $(X_1, \langle\cdot,\cdot\rangle_1)$ and $(X_2, \langle\cdot,\cdot\rangle_2)$ be two Euclidean spaces. Their Cartesian product $X = X_1 \times X_2$ can be made Euclidean via the scalar product

$$((x_1, x_2), (y_1, y_2)) = (x, y) \mapsto \langle\!\langle x, y\rangle\!\rangle := \langle x_1, y_1\rangle_1 + \langle x_2, y_2\rangle_2 .$$

This is not compulsory: cases may occur in which the product-space X has its own Euclidean structure, not possessing this "decomposability" property.

3.2 Let $(X, \langle\cdot,\cdot\rangle)$ and $(Y, \langle\!\langle\cdot,\cdot\rangle\!\rangle)$ be two Euclidean spaces, knowing that we could write just as well $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$.

(a). If A is a linear operator from X to Y, the *adjoint* of A is the unique operator A^* from Y to X, defined by

$$\langle A^*y, x\rangle = \langle\!\langle y, Ax\rangle\!\rangle \quad \text{for all } (x,y) \in X \times Y .$$

There holds $(A^*)^* = A$. When both X and Y have orthonormal bases (as is the case with canonical bases for the dot-product in the respective spaces), the matrix representing A^* in these bases is the *transpose* of the matrix representing A.

Consider the case $(Y, \langle\!\langle\cdot,\cdot\rangle\!\rangle) = (X, \langle\cdot,\cdot\rangle)$. When A is invertible, so is A^*, and then $(A^*)^{-1} = (A^{-1})^*$. When $A^* = A$, we say that A is self-adjoint, or *symmetric*. If, in addition,

$$\langle Ax, x\rangle > 0 \quad [\text{resp.} \geqslant 0] \quad \text{for all } 0 \neq x \in X ,$$

then A is *positive definite* [resp. positive semi-definite] and we write $A \succcurlyeq 0$ [resp. $A \succ 0$]. When $X = Y$ is equipped with an orthonormal basis, symmetric operators can be characterized in terms of matrices: A is symmetric [resp. symmetric positive (semi)-definite] if and only if the matrix representing A (in the orthonormal basis) is symmetric [resp. symmetric positive (semi)-definite].

(b). When the image-space Y is $\mathbb{R}$, an operator is rather called a *form*. If ℓ is a linear form on $(X, \langle\cdot,\cdot\rangle)$, there exists a unique $s \in X$ such that $\ell(x) = \langle s, x\rangle$ for all $x \in X$. If q is a quadratic form on $(X, \langle\cdot,\cdot\rangle)$, there exists a unique symmetric operator Q such that

$$q(x) := \tfrac{1}{2}\langle Qx, x\rangle \quad \text{for all } x \in X$$

(the coefficient 1/2 is useful to simplify most algebraic manipulations).

Remark 3.2 The correspondence $\ell \leftrightarrows s$ is a triviality in $(\mathbb{R}^n, {}^\top)$ (just transpose the $1 \times n$ matrices to vectors) but this is deceiving. Indeed, it is the correspondence $X \leftrightarrows X^*$ between a space and its *dual* that is being considered. For two vectors s and x of X, it is good practice to think of the scalar product $\langle s, x\rangle$ as the action of the first argument s (a slope, representing an element in the dual) on the second argument x; this helps one to understand what one is doing. Likewise, the operator Q associated with a quadratic form sends X to X^*; and an adjoint A^* is from Y^* to X^*. □

3.3 Two subspaces U and V of $(X, \langle\cdot,\cdot\rangle)$ are mutually *orthogonal* if $\langle u, v\rangle = 0$ for all $u \in U$ and $v \in V$, a relation denoted by $U \perp V$. On the other hand, U and V are *generators* of X if $U + V = X$. For given U, we denote by $U^\perp$ the *orthogonal supplement* of U, i.e. the unique subspace orthogonal to U such that U and $U^\perp$ form a generator of X.

Let $A : X \to Y$ be an arbitrary linear operator, X and Y having arbitrary scalar products. As can easily be seen,