1

Why convex?

The first modern formalization of the concept of convex function appears in J. L. W. V. Jensen, "Om konvexe funktioner og uligheder mellem midelvaerdier." Nyt Tidsskr. Math. B ***16*** *(1905), pp. 49–69. Since then, at first referring to "Jensen's convex functions," then more openly, without needing any explicit reference, the definition of convex function becomes a standard element in calculus handbooks.* (A. Guerraggio and E. Molho)[1]

Convexity theory . . . reaches out in all directions with useful vigor. Why is this so? Surely any answer must take account of the tremendous impetus the subject has received from outside of mathematics, from such diverse fields as economics, agriculture, military planning, and flows in networks. With the invention of high-speed computers, large-scale problems from these fields became at least potentially solvable. Whole new areas of mathematics (game theory, linear and nonlinear programming, control theory) aimed at solving these problems appeared almost overnight. And in each of them, convexity theory turned out to be at the core. The result has been a tremendous spurt in interest in convexity theory and a host of new results. (A. Wayne Roberts and Dale E. Varberg)[2]

1.1 Why 'convex'?

This introductory polemic makes the case for a study focusing on convex functions and their structural properties. We highlight the centrality of convexity and give a selection of salient examples and applications; many will be revisited in more detail later in the text – and many other examples are salted among later chapters. Two excellent companion pieces are respectively by Asplund [15] and by Fenchel [212]. A more recent survey article by Berger has considerable discussion of convex geometry [53].

It has been said that most of number theory devolves to the Cauchy–Schwarz inequality and the only problem is deciding 'what to Cauchy with'. In like fashion, much mathematics is tamed once one has found the right convex 'Green's function'. Why convex? Well, because . . .

- For convex sets topological, algebraic, and geometric notions often coincide; one sees this in the study of the simplex method and of continuity of convex functions. This allows one to draw upon and exploit many different sources of insight.

[1] A. Guerraggio and E. Molho, "The origins of quasi-concavity: a development between mathematics and economics," *Historia Mathematica*, **31**, 62–75, (2004).
[2] Quoted by Victor Klee in his review of [366], *SIAM Review*, **18**, 133–134, (1976).

- In a computational setting, since the *interior-point revolution* [331] in linear optimization it is now more or less agreed that 'convex' = 'easy' and 'nonconvex' = 'hard' – both theoretically and computationally. A striking illustration in combinatorial optimization is discussed in Exercise 3.3.9. In part this easiness is for the prosaic reason that local and global minima coincide.
- 'Differentiability' is understood and has been exploited throughout the sciences for centuries; 'convexity' less so, as the opening quotations attest. It is not emphasized in many instances in undergraduate courses – convex principles appear in topics such as the second derivative test for a local extremum, in linear programming (extreme points, duality, and so on) or via Jensen's inequality, etc. but often they are not presented as part of any general corpus.
- Three-dimensional convex pictures are surprisingly often realistic, while two-dimensional ones are frequently not as their geometry is too special. (Actually in a convex setting even two-dimensional pictures are much more helpful compared to those for nonconvex functions, still three-dimensional pictures are better. A good illustration is Figure 2.16. For example, working two-dimensionally, one may check convexity along lines, while seeing equal right-hand and left-hand derivatives in all directions implies differentiability.)

1.2 Basic principles

First we define some of the fundamental concepts. This is done more methodically in Chapter 2. Throughout this book, we will typically use E to denote the finite-dimensional real vector space $\mathbb{R}^n$ for some $n \in \mathbb{N}$ endowed with its usual norm, and typically X will denote a real infinite-dimensional Banach space – and sometimes merely a normed space. In this introduction we will tend to state results and introduce terminology in the setting of the Euclidean space E because this more familiar and concrete setting already illustrates their power and utility.

A set $C \subset E$ is said to be *convex* if it contains all line segments between its members: $\lambda x + (1 - \lambda)y \in C$ whenever $x, y \in C$ and $0 \leq \lambda \leq 1$. Even in two dimensions this deserves thought: every set S with $\{(x, y) : x^2 + y^2 < 1\} \subset S \subset \{(x, y) : x^2 + y^2 \leq 1\}$ is convex.

The *lower level sets* of a function $f : E \to [-\infty, +\infty]$ are the sets $\{x \in E : f(x) \leq \alpha\}$ where $\alpha \in \mathbb{R}$. The *epigraph* of a function $f : E \to [-\infty, +\infty]$ is defined by

$$\operatorname{epi} f := \{(x, t) \in E \times \mathbb{R} : f(x) \leq t\}.$$

We will see a function as *convex* if its epigraph is a convex set; and we will use ∞ and $+\infty$ interchangeably, but we prefer to use $+\infty$ when $-\infty$ is nearby.

Consider a function $f : E \to [-\infty, +\infty]$; we will say f is *closed* if its epigraph is closed; whereas f is *lower-semicontinuous* (lsc) if $\liminf_{x \to x_0} f(x) \geq f(x_0)$ for all $x_0 \in E$. These two concepts are intimately related for convex functions. Our primary focus will be on *proper functions*, those functions $f : E \to [-\infty, +\infty]$ that do not take the value $-\infty$ and whose *domain* of f, denoted by $\operatorname{dom} f$, is defined by $\operatorname{dom} f := \{x \in E : f(x) < \infty\}$. The *indicator function* of a nonempty set D

is the function δ_D defined by $\delta_D(x) := 0$ if $x \in D$ and $\delta_D(x) := +\infty$ otherwise. These notions allow one to study convex functions and convex sets interchangeably, however, our primary focus will center on convex functions.

A sketch of a real-valued differentiable convex function very strongly suggests that the derivative of such a function is monotone increasing, in fact this is true more generally – but in a nonobvious way. If we denote the derivative (or *gradient*) of a real function g by ∇g, then using the inner product the monotone increasing property of ∇g can be written as

$$\langle \nabla g(y) - \nabla g(x), y - x \rangle \geq 0 \text{ for all } x \text{ and } y.$$

The preceding inequality leads to the definition of the *monotonicity* of the gradient mapping on general spaces. Before stating our first basic result, let us recall that a set $K \subset E$ is a *cone* if $tK \subset K$ for every $t \geq 0$; and an *affine mapping* is a translate of a linear mapping.

We begin with a recapitulation of the useful preservation and characterization properties convex functions possess:

Lemma 1.2.1 (Basic properties). *The convex functions form a convex cone closed under pointwise suprema: if f_γ is convex for each $\gamma \in \Gamma$ then so is $x \mapsto \sup_{\gamma \in \Gamma} f_\gamma(x)$.*

(a) A function g is convex if and only if epi g *is convex if and only if $\delta_{\operatorname{epi} g}$ is convex.*
(b) A differentiable function g is convex on an open convex set D if and only if ∇g is a monotone operator on D, while a twice differentiable function g is convex if and only if the Hessian $\nabla^2 g$ is a positive semidefinite matrix for each value in D.
(c) $g \circ \alpha$ and $m \circ g$ are convex when g is convex, α is affine and m is monotone increasing and convex.
(d) For $t > 0$, the function $(x, t) \mapsto tg(x/t)$ is convex if and only if the function g is convex.

Proof. See Lemma 2.1.8 for (a), (c) and (d). Part (b) is developed in Theorem 2.2.6 and Theorem 2.2.8, where we are more precise about the form of differentiability used. In (d) one may be precise also about the lsc hulls, see [95] and Exercise 2.3.9. □

Before introducing the next result which summarizes many of the important continuity and differentiability properties of convex functions, we first introduce some crucial definitions. For a proper function $f : E \to (-\infty, +\infty]$, the *subdifferential* of f at $\bar{x} \in E$ where $f(\bar{x})$ is finite is defined by

$$\partial f(\bar{x}) := \{\phi \in E : \langle \phi, y - \bar{x} \rangle \leq f(y) - f(\bar{x}), \text{ for all } y \in E\}.$$

If $f(\bar{x}) = +\infty$, then $\partial f(\bar{x})$ is defined to be empty. Moreover, if $\phi \in \partial f(\bar{x})$, then ϕ is said to be a *subgradient* of f at $\bar{x}$. Note that, trivially but importantly, $0 \in \partial f(x)$ – and we call x a *critical point* – if and only if x is a minimizer of f.

While it is possible for the subdifferential to be empty, we will see below that very often it is not. An important consideration for this is whether $\bar{x}$ is in the boundary of the domain of f or in its interior, and in fact, in finite dimensions, the *relative interior*

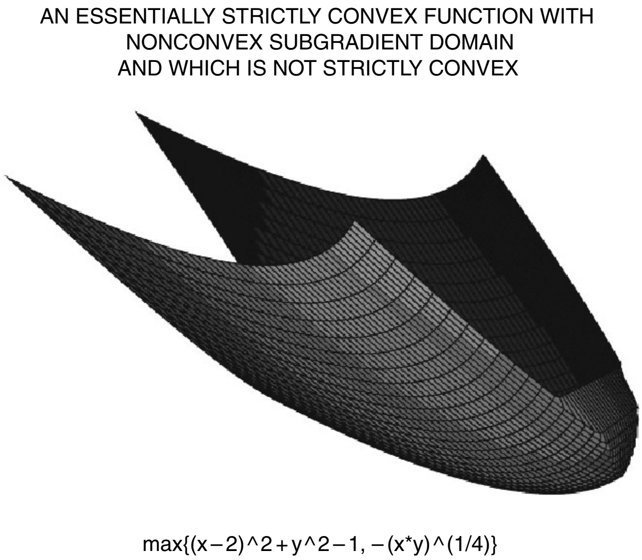

Figure 1.1 A subtle two-dimensional function from Chapter 6.

(i.e. the interior relative to the affine hull of the set) plays an important role. The function f is *Fréchet differentiable* at $\bar{x} \in \operatorname{dom} f$ with Fréchet derivative $f'(\bar{x})$ if

$$\lim_{t \to 0} \frac{f(\bar{x} + th) - f(\bar{x})}{t} = \langle f'(\bar{x}), h \rangle$$

exists uniformly for all h in the unit sphere. If the limit exists only pointwise, f is *Gâteaux differentiable* at $\bar{x}$. With these terms in mind we are now ready for the next theorem.

Theorem 1.2.2. *In Banach space, the following are central properties of convexity:*

(a) Global minima and local minima coincide for convex functions.
(b) Weak and strong closures coincide for convex functions and convex sets.
(c) A convex function is locally Lipschitz if and only if it is continuous if and only if it is locally bounded above. A finite lsc convex function is continuous; in finite dimensions lower-semicontinuity is not automatic.
(d) In finite dimensions, say n=dim E, the following hold.
  (i) The relative interior of a convex set always exists and is nonempty.
  (ii) A convex function is differentiable if and only if it has a unique subgradient.
  (iii) Fréchet and Gâteaux differentiability coincide.
  (iv) 'Finite' if and only if '$n + 1$' or 'n' (e.g. the theorems of Radon, Helly, Carathéodory, and Shapley–Folkman stated below in Theorems 1.2.3, 1.2.4, 1.2.5, and 1.2.6). These all say that a property holds for all finite sets as soon as it holds for all sets of cardinality of order the dimension of the space.

Proof. For (a) see Proposition 2.1.14; for (c) see Theorem 2.1.10 and Proposition 4.1.4. For the purely finite-dimensional results in (d), see Theorem 2.4.6 for (i); Theorem 2.2.1 for (ii) and (iii); and Exercises 2.4.13, 2.4.12, 2.4.11, and 2.4.15, for Helly's, Radon's, Carathéodory's and Shapley–Folkman theorems respectively. $\square$

Theorem 1.2.3 (Radon's theorem). *Let* $\{x_1, x_2, \ldots, x_{n+2}\} \subset \mathbb{R}^n$. *Then there is a partition* $I_1 \cup I_2 = \{1, 2, \ldots, n+2\}$ *such that* $C_1 \cap C_2 \neq \emptyset$ *where* $C_1 = \operatorname{conv}\{x_i : i \in I_1\}$ *and* $C_2 = \operatorname{conv}\{x_i : i \in I_2\}$.

Theorem 1.2.4 (Helly's theorem). *Suppose* $\{C_i\}_{i \in I}$ *is a collection of nonempty closed bounded convex sets in* $\mathbb{R}^n$, *where* I *is an arbitrary index set. If every subcollection consisting of* $n+1$ *or fewer sets has a nonempty intersection, then the entire collection has a nonempty intersection.*

In the next two results we observe that when positive as opposed to convex combinations are involved, '$n + 1$' is replaced by 'n'.

Theorem 1.2.5 (Carathéodory's theorem). *Suppose* $\{a_i : i \in I\}$ *is a finite set of points in* E. *For any subset* J *of* I, *define the cone*

$$C_J = \left\{ \sum_{i \in J} \mu_i a_i \ : \ \mu_i \in [0, +\infty), \ i \in J \right\}.$$

(a) The cone C_I *is the union of those cones* C_J *for which the set* $\{a_j : j \in J\}$ *is linearly independent. Furthermore, any such cone* C_J *is closed. Consequently, any finitely generated cone is closed.*
(b) If the point x *lies in* $\operatorname{conv}\{a_i : i \in I\}$ *then there is a subset* $J \subset I$ *of size at most* $1 + \dim E$ *such that* $x \in \operatorname{conv}\{a_i : i \in J\}$. *It follows that if a subset of* E *is compact, then so is its convex hull.*

Theorem 1.2.6 (Shapley–Folkman theorem). *Suppose* $\{S_i\}_{i \in I}$ *is a finite collection of nonempty sets in* $\mathbb{R}^n$, *and let* $S := \sum_{i \in I} S_i$. *Then every element* $x \in \operatorname{conv} S$ *can be written as* $x = \sum_{i \in I} x_i$ *where* $x_i \in \operatorname{conv} S_i$ *for each* $i \in I$ *and moreover* $x_i \in S_i$ *for all except at most* n *indices.*

Given a nonempty set $F \subset E$, the *core* of F is defined by $x \in \operatorname{core} F$ if for each $h \in E$ with $\|h\| = 1$, there exists $\delta > 0$ so that $x + th \in F$ for all $0 \leq t \leq \delta$. It is clear from the definition that the *interior* of a set F is contained in its core, that is, $\operatorname{int} F \subset \operatorname{core} F$. Let $f : E \to (-\infty, +\infty]$. We denote the set of points of continuity of f by $\operatorname{cont} f$. The *directional derivative* of f at $\bar{x} \in \operatorname{dom} f$ in the direction h is defined by

$$f'(\bar{x}; h) := \lim_{t \to 0^+} \frac{f(\bar{x} + th) - f(\bar{x})}{t}$$

if the limit exists – and it always does for a convex function. In consequence one has the following simple but crucial result.

Theorem 1.2.7 (First-order conditions). *Suppose $f : E \to (-\infty, +\infty]$ is convex. Then for any $x \in \operatorname{dom} f$ and $d \in E$,*

$$f'(x; d) \le f(x + d) - f(x). \tag{1.2.1}$$

In consequence, f is minimized (locally or globally) at x_0 if and only if $f'(x_0; d) \ge 0$ for all $d \in E$ if and only if $0 \in \partial f(x_0)$.

The following fundamental result is also a natural starting point for the so-called *Fenchel duality/Hahn–Banach theorem circle*. Let us note, also, that it directly relates differentiability to the uniqueness of subgradients.

Theorem 1.2.8 (Max formula). *Suppose $f : E \to (-\infty, +\infty]$ is convex (and lsc in the infinite-dimensional setting) and that $\bar{x} \in \operatorname{core}(\operatorname{dom} f)$. Then for any $d \in E$,*

$$f'(\bar{x}; d) = \max\{\langle \phi, d\rangle : \phi \in \partial f(\bar{x})\}. \tag{1.2.2}$$

In particular, the subdifferential $\partial f(\bar{x})$ is nonempty at all core points of $\operatorname{dom} f$.

Proof. See Theorem 2.1.19 for the finite-dimensional version and Theorem 4.1.10 for infinite-dimensional version. □

Building upon the Max formula, one can derive a quite satisfactory calculus for convex functions and linear operators. Let us note also, that for $f : E \to [-\infty, +\infty]$, the *Fenchel conjugate* of f is denoted by f^* and defined by $f^*(x^*) := \sup\{\langle x^*, x\rangle - f(x) : x \in E\}$. The conjugate is always convex (as a supremum of affine functions) while $f = f^{**}$ exactly if f is convex, proper and lsc. A very important case leads to the formula $\delta_C^*(x^*) = \sup_{x \in C}\langle x^*, x\rangle$, the *support function* of C which is clearly continuous when C is bounded, and usually denoted by σ_C. This simple conjugate formula will play a crucial role in many places, including Section 6.6 where some duality relationships between Asplund spaces and those with the Radon–Nikodým property are developed.

Theorem 1.2.9 (Fenchel duality and convex calculus). *Let E and Y be Euclidean spaces, and let $f : E \to (-\infty, +\infty]$ and $g : Y \to (-\infty, +\infty]$ and a linear map $A : E \to Y$, and let $p, d \in [-\infty, +\infty]$ be the primal and dual values defined respectively by the Fenchel problems*

$$p := \inf_{x \in E}\{f(x) + g(Ax)\} \tag{1.2.3}$$

$$d := \sup_{\phi \in Y}\{-f^*(A^*\phi) - g^*(-\phi)\}. \tag{1.2.4}$$

Cambridge University Press
978-0-521-85005-6 - Convex Functions: Constructions, Characterizations and Counterexamples
Jonathan M. Borwein and Jon D. Vanderwerff
Excerpt
More information

Then these values satisfy the weak duality inequality $p \geq d$. If, moreover, f and g are convex and satisfy the condition

$$0 \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} f) \tag{1.2.5}$$

or the stronger condition

$$A \operatorname{dom} f \cap \operatorname{cont} g \neq \emptyset \tag{1.2.6}$$

then $p = d$ and the supremum in the dual problem (1.2.4) is attained if finite.

At any point $x \in E$, the subdifferential sum rule,

$$\partial(f + g \circ A)(x) \supset \partial f(x) + A^* \partial g(Ax) \tag{1.2.7}$$

holds, with equality if f and g are convex and either condition (1.2.5) or (1.2.6) holds.

Proof. The proof for Euclidean spaces is given in Theorem 2.3.4; a version in Banach spaces is given in Theorem 4.4.18. □

A nice application of Fenchel duality is the ability to obtain primal solutions from dual ones; this is described in Exercise 2.4.19.

Corollary 1.2.10 (Sandwich theorem). *Let $f : E \to (-\infty, +\infty]$ and $g : Y \to (-\infty, +\infty]$ be convex, and let $A : E \to Y$ be linear. Suppose $f \geq -g \circ A$ and $0 \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} f)$ (or $A \operatorname{dom} f \cap \operatorname{cont} g \neq \emptyset$). Then there is an affine function $\alpha : E \to \mathbb{R}$ satisfying $f \geq \alpha \geq -g \circ A$.*

It is sometimes more desirable to symmetrize this result by using a *concave function* g, that is a function for which $-g$ is convex, and its *hypograph*, hyp g, as in Figure 1.2.

Using the sandwich theorem, one can easily deduce Hahn–Banach extension theorem (2.1.18) and the max formula to complete the so-called Fenchel duality/Hahn–Banach circle.

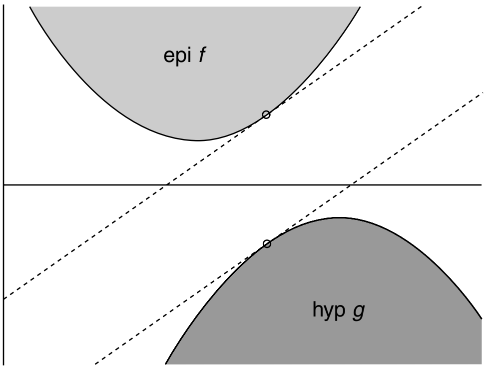

Figure 1.2 A sketch of the sandwich theorem.

A final key result is the capability to reconstruct a convex set from a well defined set of boundary points, just as one can reconstruct a convex polytope from its corners (extreme points). The basic result in this area is:

Theorem 1.2.11 (Minkowski)**.** *Let E be a Euclidean space. Any compact convex set $C \subset E$ is the convex hull of its extreme points. In Banach space it is typically necessary to take the closure of the convex hull of the extreme points.*

Proof. This theorem is proved in Euclidean spaces in Theorem 2.7.2. □

With these building blocks in place, we use the following sections to illustrate some diverse examples where convex functions and convexity play a crucial role.

1.3 Some mathematical illustrations

Perhaps the most forcible illustration of the power of convexity is the degree to which the theory of *best approximation,* i.e. existence of *nearest points* and the study of nonexpansive mappings, can be subsumed as a convex optimization problem. For a closed set S in a Hilbert space X we write $d_S(x) := \inf_{x \in S} \|x - s\|$ and call d_S the (metric) *distance function* associated with the set S. A set C in X such that each $x \in X$ has a unique nearest point in C is called a *Čebyšev set*.

Theorem 1.3.1. *Let X be a Euclidean (resp. Hilbert) space and suppose C is a nonempty (weakly) closed subset of X. Then the following are equivalent.*

(a) C is convex.
(b) C is a Čebyšev set.
(c) d_C^2 is Fréchet differentiable.
(d) d_C^2 is Gâteaux differentiable.

Proof. See Theorem 4.5.9 for the proof. □

We shall use the necessary condition for $\inf_C f$ to deduce that the projection on a convex set is nonexpansive; this and some other properties are described in Exercise 2.3.17.

Example 1.3.2 (Algebra)**.** *Birkhoff's theorem* [57] says the doubly stochastic matrices (those with nonnegative entries whose row and column sum equal one) are convex combinations of permutation matrices (their extreme points).

A proof using convexity is requested in Exercise 2.7.5 and sketched in detail in [95, Exercise 22, p. 74].

Example 1.3.3 (Real analysis)**.** The following very general construction links convex functions to nowhere differentiable continuous functions.

Theorem 1.3.4 (Nowhere differentiable functions [145])**.** *Let $a_n > 0$ be such that $\sum_{n=1}^{\infty} a_n < \infty$. Let $b_n < b_{n+1}$ be integers such that $b_n | b_{n+1}$ for each n, and the* *sequence $a_n b_n$ does not converge to 0. For each index $j \geq 1$, let f_j be a continuous function mapping the real line onto the interval $[0, 1]$ such that $f_j = 0$ at each even integer and $f_j = 1$ at each odd integer. For each integer k and each index j, let f_j be convex on the interval $(2k, 2k + 2)$.*

Then the continuous function $\sum_{j=1}^{\infty} a_j f_j(b_j x)$ has neither a finite left-derivative nor a finite right-derivative at any point.

In particular, for a convex nondecreasing function f mapping $[0, 1]$ to $[0, 1]$, define $f(x) = f(2 - x)$ for $1 < x < 2$ and extend f periodically. Then $F_f(x) := \sum_{j=1}^{\infty} 2^{-j} f(2^j x)$ defines a continuous nowhere differentiable function.

Example 1.3.5 (Operator theory)**.** The Riesz–Thorin convexity theorem informally says that if T induces a bounded linear operator between Lebesgue spaces L^{p_1} and L^{p_2} and also between L^{q_1} and L^{q_2} for $1 < p_1, p_2 < \infty$ and $1 < q_1, q_2 < \infty$ then it also maps L^{r_1} to L^{r_2} whenever $(1/r_1, 1/r_2)$ is a convex combination of $(1/p_1, 1/p_2)$ and $(1/q_1, 1/q_2)$ (all three pairs lying in the unit square).

A precise formulation is given by Zygmund in [451, p. 95].

Example 1.3.6 (Real analysis)**.** The Bohr–Mollerup theorem characterizes the gamma-function $x \mapsto \int_0^{\infty} t^{x-1} \exp(-t)\, dt$ as the unique function f mapping the positive half line to itself such that (a) $f(1) = 1$, (b) $xf(x) = f(x + 1)$ and (c) $\log f$ is convex function

A proof of this is outlined in Exercise 2.1.24; Exercise 2.1.25 follows this by outlining how this allows for computer implementable proofs of results such as $\beta(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x, y)$ where β is the classical beta-function. A more extensive discussion of this topic can be found in [73, Section 4.5].

Example 1.3.7 (Complex analysis)**.** *Gauss's theorem* shows that the roots of the derivative of a polynomial lie inside the convex hull of the zeros.

More precisely one has the *Gauss–Lucas theorem*: For an arbitrary not identically constant polynomial, the zeros of the derivative lie in the smallest convex polygon containing the zeros of the original polynomial. While Gauss originally observed: *Gauss's theorem*: The zeros of the derivative of a polynomial P that are not multiple zeros of P are the positions of equilibrium in the field of force due to unit particles situated at the zeros of P, where each particle repels with a force equal to the inverse distance. *Jensen's sharpening* states that if P is a real polynomial not identically constant, then all nonreal zeros of P' lie inside the Jensen disks determined by all pairs of conjugate nonreal zeros of P. See Pólya–Szegő [273].

Example 1.3.8 (Levy–Steinitz theorem (combinatorics))**.** The rearrangements of a series with values in Euclidean space always is an affine subspace (also called a flat).

Riemann's rearrangement theorem is the one-dimensional version of this lovely result. See [382], and also *Pólya-Szegő* [272] for the complex (planar) case.

We finish this section with an interesting example of a convex function whose convexity, established in [74, §1.9], seems hard to prove directly (a proof is outlined in Exercise 4.4.10):

Example 1.3.9 (Concave reciprocals)**.** Let $g(x) > 0$ for $x > 0$. Suppose $1/g$ is concave (which implies $\log g$ and hence g are convex) then

$$(x,y) \mapsto \frac{1}{g(x)} + \frac{1}{g(y)} - \frac{1}{g(x+y)},$$

$$(x,y,z) \mapsto \frac{1}{g(x)} + \frac{1}{g(y)} + \frac{1}{g(z)} - \frac{1}{g(x+y)} - \frac{1}{g(y+z)} - \frac{1}{g(x+z)} + \frac{1}{g(x+y+z)}$$

and all similar n-fold alternating combinations are reciprocally concave on the strictly positive orthant. The foundational case is $g(x) := x$. Even computing the Hessian in a computer algebra system in say six dimensions is a Herculean task.

1.4 Some more applied examples

Another lovely advertisement for the power of convexity is the following reduction of the classical *Brachistochrone problem* to a tractable convex equivalent problem. As Balder [29] recalls

> 'Johann Bernoulli's famous 1696 brachistochrone problem asks for the optimal shape of a metal wire that connects two fixed points A and B in space. A bead of unit mass falls along this wire, without friction, under the sole influence of gravity. The shape of the wire is defined to be optimal if the bead falls from A to B in as short a time as possible.'

Example 1.4.1 (Calculus of variations)**.** *Hidden convexity* in the Brachistochrone problem. The standard formulation, requires one to minimize

$$T(f) := \int_0^{x_1} \frac{\sqrt{1+f'^2(x)}}{\sqrt{g f(x)}}\, \mathrm{d}x \tag{1.4.1}$$

over all positive smooth arcs f on $(0, x_1)$ which extend continuously to have $f(0) = 0$ and $f(x_1) = y_1$, and where we let $A = (0,0)$ and $B := (x_1, y_1)$, with $x_1 > 0, y_1 \geq 0$. Here g is the gravitational constant.

A priori, it is not clear that the minimum even exists – and many books slough over all of the hard details. Yet, it is an easy exercise to check that the substitution $\phi := \sqrt{f}$ makes the integrand *jointly convex*. We obtain

$$S(\phi) := \sqrt{2g}T(\phi^2) = \int_0^{x_1} \sqrt{1/\phi^2(x) + 4\phi'^2(x)}\, \mathrm{d}x. \tag{1.4.2}$$

One may check elementarily that the solution ψ on $(0, x_1)$ of the differential equation

$$\left(\psi'(x)\right)^2 \psi^2(x) = C/\psi(x)^2 - 1, \qquad \psi(0) = 0,$$

where C is chosen to force $\psi(x_1) = \sqrt{y_1}$, exists and satisfies $S(\phi) > S(\psi)$ for all other feasible ϕ. Finally, one unwinds the transformations to determine that the original problem is solved by a cardioid.