Applications of Convex Analysis within Mathematics

Francisco J. Aragón Artacho*, Jonathan M. Borwein†, Victoria Martín-Márquez‡, and Liangjin Yao§

July 19, 2013

Abstract

In this paper, we study convex analysis and its theoretical applications. We first apply important tools of convex analysis to Optimization and to Analysis. We then show various deep applications of convex analysis and especially infimal convolution in Monotone Operator Theory. Among other things, we recapture the Minty surjectivity theorem in Hilbert space, and present a new proof of the sum theorem in reflexive spaces. More technically, we also discuss autoconjugate representers for maximally monotone operators. Finally, we consider various other applications in mathematical analysis.

2010 Mathematics Subject Classification:
Primary 47N10, 90C25; Secondary 47H05, 47A06, 47B65

Keywords: Adjoint, Asplund averaging, autoconjugate representer, Banach limit, Chebyshev set, convex functions, Fenchel duality, Fenchel conjugate, Fitzpatrick function, Hahn–Banach extension theorem, infimal convolution, linear relation, Minty surjectivity theorem, maximally monotone operator, monotone operator, Moreau's decomposition, Moreau envelope, Moreau's max formula, Moreau–Rockafellar duality, normal cone operator, renorming, resolvent, Sandwich theorem, subdifferential operator, sum theorem, Yosida approximation.

1 Introduction

While other articles in this collection look at the applications of Moreau's seminal work, we have opted to illustrate the power of his ideas theoretically within optimization theory and within mathematics more generally. Space constraints preclude being comprehensive, but we think the presentation made shows how elegantly much of modern analysis can be presented thanks to the work of Jean-Jacques Moreau and others.

*Centre for Computer Assisted Research Mathematics and its Applications (CARMA), University of Newcastle, Callaghan, NSW 2308, Australia. E-mail: francisco.aragon@ua.es

†Centre for Computer Assisted Research Mathematics and its Applications (CARMA), University of Newcastle, Callaghan, NSW 2308, Australia. E-mail: jonathan.borwein@newcastle.edu.au. Laureate Professor at the University of Newcastle and Distinguished Professor at King Abdul-Aziz University, Jeddah.

‡Departamento de Análisis Matemático, Universidad de Sevilla, Spain. E-mail: victoriam@us.es

§Centre for Computer Assisted Research Mathematics and its Applications (CARMA), University of Newcastle, Callaghan, NSW 2308, Australia. E-mail: liangjin.yao@newcastle.edu.au.

1.1 Preliminaries

Let X be a real Banach space with norm $\|\cdot\|$ and dual norm $\|\cdot\|_*$. When there is no ambiguity we suppress the $*$. We write X^* and $\langle\,\cdot\,,\cdot\,\rangle$ for the real dual space of continuous linear functions and the duality paring, respectively, and denote the closed unit ball by $B_X := \{x \in X \mid \|x\| \leq 1\}$ and set $\mathbb{N} := \{1, 2, 3, \ldots\}$. We identify X with its canonical image in the bidual space X^{**}. A set $C \subseteq X$ 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 $0 \leq \lambda \leq 1$.

Given a subset C of X, $\operatorname{int} C$ is the *interior* of C and $\overline{C}$ is the *norm closure* of C. For a set $D \subseteq X^*$, $\overline{D}^{\mathrm{w}^*}$ is the weak* closure of D. The *indicator function* of C, written as ι_C, is defined at $x \in X$ by

$$\iota_C(x) := \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise.} \end{cases} \tag{1}$$

The *support function* of C, written as σ_C, is defined by $\sigma_C(x^*) := \sup_{c \in C}\langle c, x^*\rangle$. There is also a naturally associated (metric) *distance function*, that is,

$$\mathrm{d}_C(x) := \inf\{\|x - y\| \mid y \in C\}\,. \tag{2}$$

Distance functions play a central role in convex analysis, both theoretically and algorithmically.

Let $f\colon X \to \left]-\infty, +\infty\right]$ be a function. Then $\operatorname{dom} f := f^{-1}(\mathbb{R})$ is the *domain* of f, and the *lower level sets* of a function $f : X \to \left]-\infty, +\infty\right]$ are the sets $\{x \in X \mid f(x) \leq \alpha\}$ where $\alpha \in \mathbb{R}$. The *epigraph* of f is $\operatorname{epi} f := \{(x, r) \in X \times \mathbb{R} \mid f(x) \leq r\}$. We will denote the set of points of continuity of f by $\operatorname{cont} f$. The function f is said to be *convex* if for any $x, y \in \operatorname{dom} f$ and any $\lambda \in [0, 1]$, one has

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

We say f is proper if $\operatorname{dom} f \neq \varnothing$. Let f be proper. The *subdifferential* of f is defined by

$$\partial f\colon X \rightrightarrows X^*\colon x \mapsto \{x^* \in X^* \mid \langle x^*, y - x\rangle \leq f(y) - f(x), \text{ for all } y \in X\}.$$

By the definition of ∂f, even when $x \in \operatorname{dom} f$, it is possible that $\partial f(x)$ may be empty. For example $\partial f(0) = \varnothing$ for $f(x) := -\sqrt{x}$ whenever $x \geq 0$ and $f(x) := +\infty$ otherwise. If $x^* \in \partial f(x)$ then x^* is said to be a *subgradient* of f at x. An important example of a subdifferential is the *normal cone* to a convex set $C \subseteq X$ at a point $x \in C$ which is defined as $N_C(x) := \partial\iota_C(x)$.

Let $g\colon X \to \left]-\infty, +\infty\right]$. Then the *inf-convolution* $f \Box g$ is the function defined on X by

$$f \Box g\colon x \mapsto \inf_{y \in X}\big\{f(y) + g(x - y)\big\}.$$

(In [45] Moreau studied inf-convolution when X is an arbitrary commutative semigroup.) Notice that, if both f and g are convex, so it is $f \Box g$ (see, e.g., [49, p. 17]).

We use the convention that $(+\infty) + (-\infty) = +\infty$ and $(+\infty) - (+\infty) = +\infty$. We will say a function $f : X \to \left]-\infty, +\infty\right]$ is *Lipschitz on a subset* D of $\operatorname{dom} f$ if there is a constant $M \geq 0$ so that $|f(x) - f(y)| \leq M\|x - y\|$ for all $x, y \in D$. In this case M is said to be a *Lipschitz constant* for f on D. If for each $x_0 \in D$, there is an open set $U \subseteq D$ with $x_0 \in U$ and a constant M so that

$|f(x) - f(y)| \le M\|x - y\|$ for all $x, y \in U$, we will say f is *locally Lipschitz on* D. If D is the entire space, we simply say f is Lipschitz or locally Lipschitz respectively.

Consider a function $f : X \to]-\infty, +\infty]$; we say f is *lower-semicontinuous* (lsc) if $\liminf_{x \to \bar{x}} f(x) \ge f(\bar{x})$ for all $\bar{x} \in X$, or equivalently, if $\operatorname{epi} f$ is closed. The function f is said to be *sequentially weakly lower semi-continuous* if for every $\bar{x} \in X$ and every sequence $(x_n)_{n \in \mathbb{N}}$ which is weakly convergent to $\bar{x}$, one has $\liminf_{n \to \infty} f(x_n) \ge f(\bar{x})$. This is a useful distinction since there are infinite dimensional Banach spaces (*Schur spaces* such as ℓ^1) in which weak and norm convergence coincide for sequences, see [22, p. 384, esp. Thm 8.2.5].

1.2 Structure of this paper

The remainder of this paper is organized as follows. In Section 2, we describe results about Fenchel conjugates and the subdifferential operator, such as Fenchel duality, the Sandwich theorem, etc. We also look at some interesting convex functions and inequalities. In Section 3, we discuss the Chebyshev problem from abstract approximation. In Section 4, we show applications of convex analysis in Monotone Operator Theory. We reprise such results as the Minty surjectivity theorem, and present a new proof of the sum theorem in reflexive spaces. We also discuss Fitzpatrick's problem on so called autoconjugate representers for maximally monotone operators. In Section 5 we discuss various other applications.

2 Subdifferential operators, conjugate functions & Fenchel duality

We begin with some fundamental properties of convex sets and convex functions. While many results hold in all locally convex spaces, some of the most important such as (iv)(b) in the next Fact do not.

Fact 2.1 (Basic properties [22, Ch. 2 and 4].) *The following hold.*

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

(ii) *A function f is convex if and only if* $\operatorname{epi} f$ *is convex if and only if $\iota_{\operatorname{epi} f}$ is convex.*

(iii) *Global minima and local minima in the domain coincide for proper convex functions.*

(iv) *Let f be a proper convex function and let $x \in \operatorname{dom} f$. (a) f is locally Lipschitz at x if and only f is continuous at x if and only if f is locally bounded at x. (b) Additionally, if f is lower semicontinuous, then f is continuous at every point in* $\operatorname{int} \operatorname{dom} f$.

(v) *A proper lower semicontinuous and convex function is bounded from below by a continuous affine function.*

(vi) *If C is a nonempty set, then $\mathrm{d}_C(\cdot)$ is non-expansive (i.e., is a Lipschitz function with constant one). Additionally, if C is convex, then $\mathrm{d}_C(\cdot)$ is a convex function.*

(vii) *If C is a convex set, then C is weakly closed if and only if it is norm closed.*

(viii) Three-slope inequality*: Suppose* $f : \mathbb{R} \to]-\infty, \infty]$ *is convex and* $a < b < c$. *Then*

$$\frac{f(b) - f(a)}{b - a} \leq \frac{f(c) - f(a)}{c - a} \leq \frac{f(c) - f(b)}{c - b}.$$

The following trivial fact shows the fundamental significance of subgradients in optimization.

Proposition 2.2 (Subdifferential at optimality) *Let* $f\colon X \to]-\infty, +\infty]$ *be a proper convex function. Then the point* $\bar{x} \in \operatorname{dom} f$ *is a (global) minimizer of* f *if and only if* $0 \in \partial f(\bar{x})$.

The *directional derivative* of f at $\bar{x} \in \operatorname{dom} f$ in the direction d is defined by

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

if the limit exists. If f is convex, the directional derivative is everywhere finite at any point of $\operatorname{int}\operatorname{dom} f$, and it turns out to be Lipschitz at $\operatorname{cont} f$. We use the term directional derivative with the understanding that it is actually a *one-sided* directional derivative.

If the directional derivative $f'(\bar{x}, d)$ exists for all directions d and the operator $f'(\bar{x})$ defined by $\langle f'(\bar{x}), \cdot \rangle := f'(\bar{x}; \cdot\,)$ is linear and bounded, then we say that f is *Gâteaux differentiable* at $\bar{x}$, and $f'(\bar{x})$ is called the *Gâteaux derivative.* Every function $f : X \to]-\infty, +\infty]$ which is lower semicontinuous, convex and Gâteaux differentiable at x, it is continuous at x. Additionally, the following properties are relevant for the existence and uniqueness of the subgradients.

Proposition 2.3 (See [22, Fact 4.2.4 and Corollary 4.2.5].) *Suppose* $f : X \to]-\infty, +\infty]$ *is convex.*

(i) *If* f *is Gâteaux differentiable at* $\bar{x}$*, then* $f'(\bar{x}) \in \partial f(\bar{x})$.

(ii) *If* f *is continuous at* $\bar{x}$*, then* f *is Gâteaux differentiable at* $\bar{x}$ *if and only if* $\partial f(\bar{x})$ *is a singleton.*

Example 2.4 We show that part (ii) in Proposition 2.3 is not always true in infinite dimensions without continuity hypotheses.

(a) The indicator of the Hilbert cube $C := \{x = (x_1, x_2, \dots) \in \ell^2 : |x_n| \leq 1/n, \forall n \in \mathbb{N}\}$ at zero or any other non-support point has a unique subgradient but is nowhere Gâteaux differentiable.

(b) *Boltzmann-Shannon entropy* $x \mapsto \int_0^1 x(t)\log(x(t))\mathrm{d}t$ viewed as a lower semicontinuous and convex function on $L^1[0, 1]$ has unique subgradients at $x(t) > 0$ a.e. but is nowhere Gâteaux differentiable (which for a lower semicontinuous and convex function in Banach space implies continuity).

That Gâteaux differentiability of a convex and lower semicontinuous function implies continuity at the point is a consequence of the Baire category theorem. $\Diamond$

The next result proved by Moreau in 1963 establishes the relationship between subgradients and directional derivatives, see also [49, page 65]. Proofs can be also found in most of the books in variational analysis, see e.g. [25, Theorem 4.2.7].

Theorem 2.5 (Moreau's max formula [46]) *Let $f\colon X \to \,]-\infty, +\infty]$ be a convex function and let $d \in X$. Suppose that f is continuous at $\bar{x}$. Then, $\partial f(\bar{x}) \neq \varnothing$ and*

$$f'(\bar{x}; d) = \max\{\langle x^*, d\rangle \mid x^* \in \partial f(\bar{x})\}. \tag{3}$$

Let $f : X \to [-\infty, +\infty]$. The *Fenchel conjugate* (also called the *Legendre-Fenchel conjugate*[1] or *transform*) of f is the function $f^* : X^* \to [-\infty, +\infty]$ defined by

$$f^*(x^*) := \sup_{x\in X}\{\langle x^*, x\rangle - f(x)\}.$$

We can also consider the conjugate of f^* called the *biconjugate* of f and denoted by f^{**}. This is a convex function on X^{**} satisfying $f^{**}|_X \leq f$. A useful and instructive example is $\sigma_C = \iota_C^*$.

Example 2.6 Let $1 < p < \infty$. If $f(x) := \frac{\|x\|^p}{p}$ for $x \in X$ then $f^*(x^*) = \frac{\|x^*\|_*^q}{q}$, where $\frac{1}{p} + \frac{1}{q} = 1$. Indeed, for any $x^* \in X^*$, one has

$$f^*(x^*) = \sup_{\lambda\in\mathbb{R}_+}\sup_{\|x\|=1}\left\{\langle x^*, \lambda x\rangle - \frac{\|\lambda x\|^p}{p}\right\} = \sup_{\lambda\in\mathbb{R}_+}\left\{\lambda\|x^*\|_* - \frac{\lambda^p}{p}\right\} = \frac{\|x^*\|_*^q}{q}.$$

$\Diamond$

By direct construction and Fact 2.1 (i), for any function f, the conjugate function f^* is always convex and lower semicontinuous, and if the domain of f is nonempty, then f^* never takes the value $-\infty$. The conjugate plays a role in convex analysis in many ways analogous to the role played by the Fourier transform in harmonic analysis with infimal convolution, see below, replacing integral convolution and sum replacing product [22, Chapter 2.].

2.1 Inequalities and their applications

An immediate consequence of the definition is that for $f, g : X \to [-\infty, +\infty]$, the inequality $f \geq g$ implies $f^* \leq g^*$. An important result which is straightforward to prove is the following.

Proposition 2.7 (Fenchel–Young) *Let $f : X \to \,]-\infty, +\infty]$. All points $x^* \in X^*$ and $x \in \operatorname{dom} f$ satisfy the inequality*

$$f(x) + f^*(x^*) \geq \langle x^*, x\rangle. \tag{4}$$

Equality holds if and only if $x^* \in \partial f(x)$.

Example 2.8 (Young's inequality) By taking f as in Example 2.6, one obtains directly from Proposition 2.7

$$\frac{\|x\|^p}{p} + \frac{\|x^*\|_*^q}{q} \geq \langle x^*, x\rangle,$$

for all $x \in X$ and $x^* \in X^*$, where $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. When $X = \mathbb{R}$ one recovers the original Young inequality. $\Diamond$

[1] Originally the connection was made between a monotone function on an interval and its inverse. The convex functions then arise by integration.

This in turn leads to one of the workhorses of modern analysis:

Example 2.9 (Hölder's inequality) Let f and g be measurable on a measure space (X, μ). Then

$$\int_X fg \, \mathrm{d}\mu \le \|f\|_p \|g\|_q, \tag{5}$$

where $1 < p < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$. Indeed, by rescaling, we may assume without loss of generality that $\|f\|_p = \|g\|_q = 1$. Then Young's inequality in Example 2.8 yields

$$|f(x)g(x)| \le \frac{|f(x)|^p}{p} + \frac{|g(x)|^q}{q} \quad \text{for } x \in X,$$

and (5) follows by integrating both sides. The result holds true in the limit for $p = 1$ or $p = \infty$. $\diamondsuit$

We next take a brief excursion into special function theory and normed space geometry to emphasize that "convex functions are everywhere."

Example 2.10 (Bohr–Mollerup theorem) The *Gamma function* defined for $x > 0$ as

$$\Gamma(x) := \int_0^\infty e^{-t} t^{x-1} \mathrm{d}t = \lim_{n \to \infty} \frac{n!\, n^x}{x(x+1)\cdots(x+n)}$$

is the unique function f mapping the positive half-line to itself and such that (a) $f(1) = 1$, (b) $xf(x) = f(x+1)$ and (c) $\log f$ is a convex function.

Indeed, clearly $\Gamma(1) = 1$, and it is easy to prove (b) for Γ by using integration by parts. In order to show that $\log \Gamma$ is convex, pick any $x, y > 0$ and $\lambda \in (0,1)$ and apply Hölder's inequality (5) with $p = 1/\lambda$ to the functions $t \mapsto e^{-\lambda t} t^{\lambda(x-1)}$ and $t \mapsto e^{-(1-\lambda)t} t^{(1-\lambda)(y-1)}$. For the converse, let $g := \log f$. Then (a) and (b) imply $g(n+1+x) = \log\left[x(1+x)\dots(n+x)f(x)\right]$ and thus $g(n+1) = \log(n!)$. Convexity of g together with the three-slope inequality, see Fact 2.1(viii), implies that

$$g(n+1) - g(n) \le \frac{g(n+1+x) - g(n+1)}{x} \le g(n+2+x) - g(n+1+x),$$

and hence,

$$x \log(n) \le \log\left(x(x+1)\cdots(x+n)f(x)\right) - \log(n!) \le x \log(n+1+x);$$

whence,

$$0 \le g(x) - \log\left(\frac{n!\, n^x}{x(x+1)\cdots(x+n)}\right) \le x \log\left(1 + \frac{1+x}{n}\right).$$

Taking limits when $n \to \infty$ we obtain

$$f(x) = \lim_{n \to \infty} \frac{n!\, n^x}{x(x+1)\cdots(x+n)} = \Gamma(x).$$

As a bonus we recover a classical and important limit formula for $\Gamma(x)$.

Application of the Bohr–Mollerup theorem is often *automatable* in a computer algebra system, as we now illustrate. Consider the *beta function*

$$\beta(x,y) \;:=\; \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}\,t \tag{6}$$

for $\mathrm{Re}(x), \mathrm{Re}(y) > 0$. As is often established using polar coordinates and double integrals

$$\beta(x,y) \;=\; \frac{\Gamma(x)\,\Gamma(y)}{\Gamma(x+y)}. \tag{7}$$

We may use the Bohr–Mollerup theorem with

$$f := x \to \beta(x,y)\,\Gamma(x+y)/\Gamma(y)$$

to prove (7) for real x, y.

Now (a) and (b) from Example 2.10 are easy to verify. For (c) we again use Hölder's inequality to show f is log-convex. Thus, $f = \Gamma$ as required. $\Diamond$

Example 2.11 (Blaschke–Santaló theorem) The volume of a unit ball in the $\|\cdot\|_p$-norm, $V_n(p)$ is

$$V_n(p) \;=\; 2^n \frac{\Gamma(1+\frac{1}{p})^n}{\Gamma(1+\frac{n}{p})}. \tag{8}$$

as was first determined by Dirichlet. When $p = 2$, this gives

$$V_n \;=\; 2^n \frac{\Gamma(\frac{3}{2})^n}{\Gamma(1+\frac{n}{2})} \;=\; \frac{\Gamma(\frac{1}{2})^n}{\Gamma(1+\frac{n}{2})},$$

which is more concise than that usually recorded in texts.

Let C in $\mathbb{R}^n$ be a *convex body* which is symmetric around zero, that is, a closed bounded convex set with nonempty interior. Denoting n-dimensional Euclidean volume of $S \subseteq \mathbb{R}^n$ by $V_n(S)$, the *Blaschke–Santaló* inequality says

$$V_n(C)\,V_n(C^\circ) \;\le\; V_n(E)\,V_n(E^\circ) \;=\; V_n^2(B_n(2)) \tag{9}$$

where maximality holds (only) for *any symmetric* ellipsoid E and $B_n(2)$ is the Euclidean unit ball. It is *conjectured* the minimum is attained by the 1-ball and the ∞-ball. Here as always the polar set is defined by $C^\circ \;:=\; \{y \in \mathbb{R}^n \colon \langle y, x\rangle \le 1 \text{ for all } x \in C\}$.

The p-ball case of (9) follows by proving the following convexity result:

Theorem 2.12 (Harmonic-arithmetic log-concavity) The function

$$V_\alpha(p) := 2^\alpha \Gamma\left(1+\frac{1}{p}\right)^\alpha \Big/ \Gamma\left(1+\frac{\alpha}{p}\right)$$

satisfies

$$V_\alpha(p)^\lambda\, V_\alpha(q)^{1-\lambda} \;<\; V_\alpha\left(\frac{1}{\frac{\lambda}{p}+\frac{1-\lambda}{q}}\right), \tag{10}$$

for all $\alpha > 1$, if $p, q > 1$, $p \neq q$, and $\lambda \in (0,1)$.

Set $\alpha := n$, $\frac{1}{p}+\frac{1}{q} = 1$ with $\lambda = 1-\lambda = 1/2$ to recover the $p-$norm case of the Blaschke–Santaló inequality. It is amusing to deduce the corresponding lower bound. This technique extends to various substitution norms. Further details may be found in [16, §5.5]. Note that we may easily explore $V_\alpha(p)$ graphically. ♢

2.2 The biconjugate and duality

The next result has been associated by different authors with the names of Legendre, Fenchel, Moreau and Hörmander; see, e.g., [22, Proposition 4.4.2].

Proposition 2.13 (**Hörmander**[2])*(See [66, Theorem 2.3.3] or [22, Proposition 4.4.2(a)].)* Let $f : X \to\ ]-\infty, +\infty]$ be a proper function. Then

$$f \text{ is convex and lower semicontinuous } \Leftrightarrow f = f^{**}|_X.$$

Example 2.14 (Establishing convexity) (See [12, Theorem 1].) We may compute conjugates by hand or using the software *SCAT* [20]. This is discussed further in Section 5.3. Consider $f(x) := e^x$. Then $f^*(x) = x\log(x) - x$ for $x \geq 0$ (taken to be zero at zero) and is infinite for $x < 0$. This establishes the convexity of $x\log(x) - x$ in a way that takes no knowledge of $x\log(x)$.

A more challenging case is the following (slightly corrected) conjugation formula [21, p. 94, Ex. 13] which can be computed algorithmically: Given real $\alpha_1, \alpha_2, \ldots, \alpha_m > 0$, define $\alpha := \sum_i \alpha_i$ and suppose a real μ satisfies $\mu > \alpha + 1$. Now define a function $f : \mathbb{R}^m \times \mathbb{R} \mapsto\ ]-\infty, +\infty]$ by

$$f(x,s) := \begin{cases} \mu^{-1} s^\mu \prod_i x_i^{-\alpha_i} & \text{if } x \in \mathbb{R}^m_{++},\ s \in \mathbb{R}_+; \\ \\ 0 & \text{if } \exists x_i = 0,\ x \in \mathbb{R}^m_+,\ s = 0; \\ +\infty & \text{otherwise.} \end{cases}, \qquad \forall x := (x_n)_{n=1}^m \in \mathbb{R}^m,\ s \in \mathbb{R}.$$

It transpires that

$$f^*(y,t) = \begin{cases} \rho\nu^{-1} t^\nu \prod_i (-y_i)^{-\beta_i} & \text{if } y \in \mathbb{R}^m_{--},\ t \in \mathbb{R}_+ \\ 0 & \text{if } y \in \mathbb{R}^m_-,\ t \in \mathbb{R}_- \\ +\infty & \text{otherwise} \end{cases}, \qquad \forall y := (y_n)_{n=1}^m \in \mathbb{R}^m,\ t \in \mathbb{R}.$$

for constants

$$\nu := \frac{\mu}{\mu - (\alpha+1)}, \quad \beta_i := \frac{\alpha_i}{\mu - (\alpha+1)}, \quad \rho := \prod_i \left(\frac{\alpha_i}{\mu}\right)^{\beta_i}.$$

[2] Hörmander first proved the case of support and indicator functions in [38] which led to discovery of general result.

We deduce that $f = f^{**}$, whence f (and f^*) is (essentially strictly) convex. For attractive alternative proof of convexity see [42]. Many other substantive examples are to be found in [21, 22]. $\diamondsuit$

The next theorem gives us a remarkable sufficient condition for convexity of functions in terms of the Gâteaux differentiability of the conjugate. There is a simpler analogue for the Fréchet derivative.

Theorem 2.15 (See [22, Corollary 4.5.2].) *Suppose $f : X \to\ ]-\infty, +\infty]$ is such that f^{**} is proper. If f^* is Gâteaux differentiable at all $x^* \in \operatorname{dom} \partial f^*$ and f is sequentially weakly lower semicontinuous, then f is convex.*

Let $f : X \to\ ]-\infty, +\infty]$. We say f is *coercive* if $\lim_{\|x\|\to\infty} f(x) = +\infty$. We say f is *supercoercive* if $\lim_{\|x\|\to\infty} \frac{f(x)}{\|x\|} = +\infty$.

Fact 2.16 (See [22, Fact 4.4.8].) *If f is proper convex and lower semicontinuous at some point in its domain, then the following statements are equivalent.*

(i) *f is coercive.*

(ii) *There exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $f \geq \alpha\|\cdot\| + \beta$.*

(iii) $\liminf_{\|x\|\to\infty} f(x)/\|x\| > 0$.

(iv) *f has bounded lower level sets.*

Because a convex function is continuous at a point if and only if it is bounded above on a neighborhood of that point (Fact 2.1(iv)), we get the following result; see also [38, Theorem 7] for the case of the indicator function of a bounded convex set.

Theorem 2.17 (Hörmander–Moreau–Rockafellar) *Let $f : X \to\ ]-\infty, +\infty]$ be convex and lower semicontinuous at some point in its domain, and let $x^* \in X^*$. Then $f - x^*$ is coercive if and only if f^* is continuous at x^*.*

Proof. "$\Rightarrow$": By Fact 2.16, there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $f \geq x^* + \alpha\|\cdot\| + \beta$. Then $f^* \leq -\beta + \iota_{\{x^*+\alpha B_{X^*}\}}$, from where $x^* + \alpha B_{X^*} \subseteq \operatorname{dom} f^*$. Therefore, f^* is continuous at x^* by Fact 2.1(iv).

"$\Leftarrow$": By the assumption, there exists $\beta \in \mathbb{R}$ and $\delta > 0$ such that

$$f^*(x^* + z^*) \leq \beta, \quad \forall z^* \in \delta B_{X^*}.$$

Thus, by Proposition 2.7,

$$\langle x^* + z^*, y\rangle - f(y) \leq \beta, \quad \forall z^* \in \delta B_{X^*}, \forall y \in X;$$

whence, taking the supremum with $z^* \in \delta B_{X^*}$,

$$\delta\|y\| - \beta \leq f(y) - \langle x^*, y\rangle, \quad \forall y \in X.$$

Then, by Fact 2.16, $f - x^*$ is coercive. ■

Example 2.18 Given a set C in X, recall that the negative polar cone of C is the convex cone

$$C^- := \{x^* \in X^* \mid \sup\langle x^*, C\rangle \leq 0\}.$$

Suppose that X is reflexive and let $K \subseteq X$ be a closed convex cone. Then K^- is another nonempty closed convex cone with $K^{--} := (K^-)^- = K$. Moreover, the indicator function of K and K^- are conjugate to each other. If we set $f := \iota_{K^-}$, the indicator function of the negative polar cone of K, Theorem 2.17 applies to get that

$$x \in \operatorname{int} K \text{ if and only if the set } \{x^* \in K^- \mid \langle x^*, x\rangle \geq \alpha\} \text{ is bounded for any } \alpha \in \mathbb{R}.$$

Indeed, since $x \in \operatorname{int} K = \operatorname{int}\operatorname{dom}\iota^*_{K^-}$ if and only if $\iota^*_{K^-}$ is continuous at x, from Theorem 2.17 we have that this is true if and only if the function $\iota_{K^-} - x$ is coercive. Now, Fact 2.16 assures us that coerciveness is equivalent to boundedness of the lower level sets, which implies the assertion. $\diamondsuit$

Theorem 2.19 (Moreau–Rockafellar duality [47]) *Let $f : X \to (-\infty, +\infty]$ be a lower semi-continuous convex function. Then f is continuous at 0 if and only if f^* has weak*-compact lower level sets.*

Proof. Observe that f is continuous at 0 if and only if f^{**} is continuous at 0 ([22, Fact 4.4.4(b)])if and only if f^* is coercive (Theorem 2.17) if and only if f^* has bounded lower level sets (Fact 2.16) if and only if f^* has weak*-compact lower level sets by the Banach-Alaoglu theorem (see [59, Theorem 3.15]). ■

Theorem 2.20 (Conjugates of supercoercive functions) *Suppose $f : X \to\ ]-\infty, +\infty]$ is a lower semicontinuous and proper convex function. Then*

(a) f is supercoercive if and only if f^* is bounded (above) on bounded sets.

(b) f is bounded (above) on bounded sets if and only if f^* is supercoercive.

Proof. (a) "$\Rightarrow$": Given any $\alpha > 0$, there exists M such that $f(x) \geq \alpha\|x\|$ if $\|x\| \geq M$. Now there exists $\beta \geq 0$ such that $f(x) \geq -\beta$ if $\|x\| \leq M$ by Fact 2.1(v). Therefore $f \geq \alpha\|\cdot\| + (-\alpha M - \beta)$. Thus, it implies that $f^* \leq \alpha(\|\cdot\|)^*(\frac{\cdot}{\alpha}) + \alpha M + \beta$ and hence $f^* \leq \alpha M + \beta$ on αB_{X^*}.

"$\Leftarrow$": Let $\gamma > 0$. Now there exists K such that $f^* \leq K$ on γB_{X^*}. Then $f \geq \gamma\|\cdot\| - K$ and so $\liminf_{\|x\|\to\infty} \frac{f(x)}{\|x\|} \geq \gamma$. Hence $\liminf_{\|x\|\to\infty} \frac{f(x)}{\|x\|} = +\infty$.

(b): According to (a), f^* is supercoercive if and only if f^{**} is bounded on bounded sets. By [22, Fact 4.4.4(a)] this holds if and only if f is bounded (above) on bounded sets. ■

We finish this subsection by recalling some properties of infimal convolutions. Some of their many applications include smoothing techniques and approximation. We shall meet them again in Section 4. Let $f, g : X \to\ ]-\infty, +\infty]$. Geometrically, the infimal convolution of f and g is the largest extended real-valued function whose epigraph contains the sum of epigraphs of f and g (see example in Figure 1), consequently it is a convex function. The following is a useful result concerning the conjugate of the infimal convolution.

Fact 2.21 (See [22, Lemma 4.4.15] and [49, pp. 37-38].) *If f and g are proper functions on X, then $(f\Box g)^* = f^* + g^*$. Additionally, suppose f, g are convex and bounded below. If $f : X \to \mathbb{R}$ is continuous (resp. bounded on bounded sets, Lipschitz), then $f\Box g$ is a convex function that is continuous (resp. bounded on bounded sets, Lipschitz).*

Remark 2.22 Suppose C is a nonempty convex set. Then $\mathrm{d}_C = \|\cdot\|\Box\iota_C$, implying that d_C is a Lipschitz convex function. $\Diamond$

Example 2.23 Consider $f, g : \mathbb{R} \to\ ]-\infty, +\infty]$ given by

$$f(x) := \begin{cases} -\sqrt{1-x^2}, & \text{for } -1 \le x \le 1, \\ +\infty & \text{otherwise,} \end{cases} \qquad \textit{and} \quad g(x) := |x|.$$

The infimal convolution of f and g is

$$(f\Box g)(x) = \begin{cases} -\sqrt{1-x^2}, & -\frac{\sqrt{2}}{2} \le x \le -\frac{\sqrt{2}}{2}; \\ |x| - \sqrt{2}, & \text{otherwise.} \end{cases},$$

as shown in Figure 1. $\Diamond$

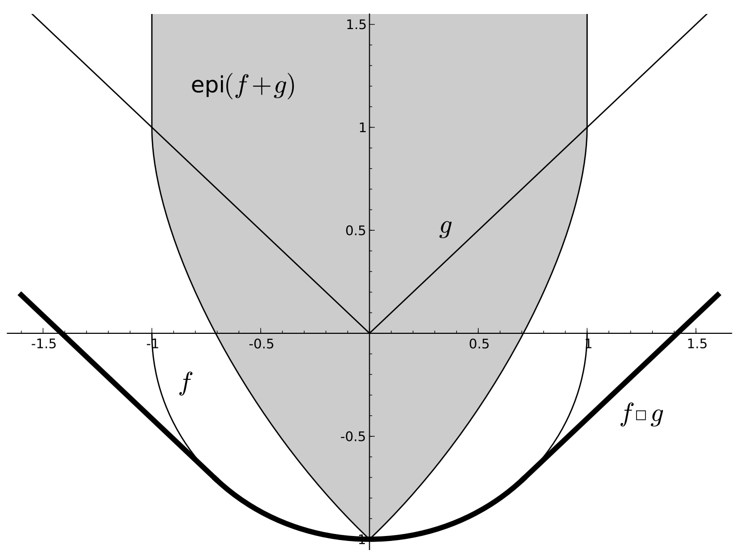

Figure 1: Infimal convolution of $f(x) = -\sqrt{1-x^2}$ and $g(x) = |x|$.

2.3 The Hahn-Banach circle

Let $T : X \to Y$ be a linear mapping between two Banach spaces X and Y. The *adjoint* of T is the linear mapping $T^* : Y^* \to X^*$ defined, for $y^* \in Y^*$, by

$$\langle T^* y^*, x\rangle = \langle y^*, Tx\rangle \quad \text{for all } x \in X.$$

A flexible modern version of Fenchel's celebrated duality theorem is:

Theorem 2.24 (Fenchel duality) *Let Y be another Banach space, let $f\colon X \to\,]-\infty, +\infty]$ and $g\colon Y \to\,]-\infty, +\infty]$ be convex functions and let $T\colon X \to Y$ be a bounded linear operator. Define the primal and dual values $p, d \in [-\infty, +\infty]$ by solving the* Fenchel problems

$$\begin{aligned} p &:= \inf_{x\in X}\{f(x) + g(Tx)\} \\ d &:= \sup_{y^*\in Y^*}\{-f^*(T^*y^*) - g^*(-y^*)\}. \end{aligned} \tag{11}$$

Then these values satisfy the weak duality *inequality $p \geq d$.*

Suppose further that f, g and T satisfy either

$$\bigcup_{\lambda>0} \lambda\,[\operatorname{dom} g - T \operatorname{dom} f] = Y \ \text{ and both } f \text{ and } g \text{ are lower semicontinuous,} \tag{12}$$

or the condition

$$\operatorname{cont} g \cap T \operatorname{dom} f \neq \varnothing. \tag{13}$$

Then $p = d$, and the supremum in the dual problem (11) *is attained when finite. Moreover, the perturbation function $h(u) := \inf_x f(x) + g(Tx + u)$ is convex and continuous at zero.*

Generalizations of Fenchel duality Theorem can be found in [27, 26]. An easy consequence is:

Corollary 2.25 (Infimal convolution) *Under the hypotheses of the Fenchel duality theorem 2.24 $(f + g)^*(x^*) = (f^*\Box g^*)(x^*)$ with attainment when finite.*

Another nice consequence of Fenchel duality is the ability to obtain primal solutions from dual ones, as we now record.

Corollary 2.26 *Suppose the conditions for equality in the Fenchel duality Theorem 2.24 hold, and that $\bar{y}^* \in Y^*$ is an optimal dual solution. Then the point $\bar{x} \in X$ is optimal for the primal problem if and only if it satisfies the two conditions $T^*\bar{y}^* \in \partial f(\bar{x})$ and $-\bar{y}^* \in \partial g(T\bar{x})$.*

The regularity conditions in Fenchel duality theorem can be weakened when each function is *polyhedral*, i.e., when their epigraph is polyhedral.

Theorem 2.27 (Polyhedral Fenchel duality) (See [21, Corollary 5.1.9].) *Suppose that X is a finite-dimensional space. The conclusions of the Fenchel duality Theorem 2.24 remain valid if the regularity condition* (12) *is replaced by the assumption that the functions f and g are polyhedral with*

$$\operatorname{dom} g \cap T \operatorname{dom} f \neq \varnothing.$$

Fenchel duality applied to a linear programming program yields the well-known Lagrangian duality.

Corollary 2.28 (Linear programming duality) *Given $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ and A an $m \times n$ real matrix, one has*

$$\inf_{x\in\mathbb{R}^n}\{c^T x \mid Ax \le b\} \ge \sup_{\lambda\in\mathbb{R}^m_+}\{-b^T\lambda \mid A^T\lambda = -c\}, \tag{14}$$

where $\mathbb{R}^m_+ := \big\{(x_1, x_2, \cdots, x_m) \mid x_i \ge 0,\ i = 1, 2, \cdots, m\big\}$. Equality in (14) holds if $b \in \operatorname{ran} A + \mathbb{R}^m_+$. Moreover, both extrema are obtained when finite.

Proof. Take $f(x) := c^T x$, $T := A$ and $g(y) := \iota_{b_\ge}(y)$ where $b_\ge := \{y \in \mathbb{R}^m \mid y \le b\}$. Then apply the polyhedral Fenchel duality Theorem 2.27 observing that $f^* = \iota_{\{c\}}$, and for any $\lambda \in \mathbb{R}^m$,

$$g^*(\lambda) = \sup_{y\le b} y^T\lambda = \begin{cases} b^T\lambda, & \text{if } \lambda \in \mathbb{R}^m_+; \\ +\infty, & \text{otherwise;} \end{cases}$$

and (14) follows, since $\operatorname{dom} g \cap A \operatorname{dom} f = \{Ax \in \mathbb{R}^m \mid Ax \le b\}$. ■

One can easily derive various relevant results from Fenchel duality, such as the Sandwich theorem, the subdifferential sum rule, and the Hahn-Banach extension theorem, among many others.

Theorem 2.29 (Extended sandwich theorem) *Let X and Y be Banach spaces and let $T : X \to Y$ be a bounded linear mapping. Suppose that $f : X \to\ ]-\infty, +\infty]$, $g : Y \to\ ]-\infty, +\infty]$ are proper convex functions which together with T satisfy either* (12) *or* (13). *Assume that $f \ge -g \circ T$. Then there is an affine function $\alpha : X \to \mathbb{R}$ of the form $\alpha(x) = \langle T^*y^*, x\rangle + r$ satisfying $f \ge \alpha \ge -g \circ T$. Moreover, for any $\bar{x}$ satisfying $f(\bar{x}) = (-g \circ T)(\bar{x})$, we have $-y^* \in \partial g(T\bar{x})$.*

Proof. With notation as in the Fenchel duality Theorem 2.24, we know $d = p$, and since $p \ge 0$ because $f(x) \ge -g(Tx)$, the supremum in d is attained. Therefore there exists $y^* \in Y^*$ such that

$$0 \le p = d = -f^*(T^*y^*) - g^*(-y^*).$$

Then, by Fenchel-Young inequality (4), we obtain

$$0 \le p \le f(x) - \langle T^*y^*, x\rangle + g(y) + \langle y^*, y\rangle, \tag{15}$$

for any $x \in X$ and $y \in Y$. For any $z \in X$, setting $y = Tz$ in the previous inequality, we obtain

$$a := \sup_{z\in X}[-g(Tz) - \langle T^*y^*, z\rangle] \le b := \inf_{x\in X}[f(x) - \langle T^*y^*, x\rangle]$$

Now choose $r \in [a, b]$. The affine function $\alpha(x) := \langle T^*y^*, x\rangle + r$ satisfies $f \ge \alpha \ge -g\circ T$, as claimed.

The last assertion follows from (15) simply by setting $x = \bar{x}$, where $\bar{x}$ satisfies $f(\bar{x}) = (-g\circ T)(\bar{x})$. Then we have $\sup_{y\in Y}\{\langle -y^*, y\rangle - g(y)\} \le (-g\circ T)(\bar{x}) - \langle T^*y^*, \bar{x}\rangle$. Thus $g^*(-y^*) + g(T\bar{x}) \le -\langle y^*, T\bar{x}\rangle$ and hence $-y^* \in \partial g(T\bar{x})$. ■

When $X = Y$ and T is the identity we recover the classical Sandwich theorem. The next example shows that without a constraint qualification, the sandwich theorem may fail.

Example 2.30 Consider $f, g : \mathbb{R} \to\ ]-\infty, +\infty]$ given by

$$f(x) := \begin{cases} -\sqrt{-x}, & \text{for } x \le 0, \\ +\infty & \text{otherwise,} \end{cases} \quad \textit{and} \quad g(x) := \begin{cases} -\sqrt{x}, & \text{for } x \ge 0, \\ +\infty & \text{otherwise.} \end{cases}$$

In this case, $\bigcup_{\lambda>0} \lambda\,[\operatorname{dom} g - \operatorname{dom} f] = [0, +\infty[\neq \mathbb{R}$ and it is not difficult to prove there is not any affine function which separates f and $-g$, see Figure 2. $\Diamond$

The prior constraint qualifications are sufficient but not necessary for the sandwich theorem as we illustrate in the next example.

Example 2.31 Let $f, g : \mathbb{R} \to\ ]-\infty, +\infty]$ be given by

$$f(x) := \begin{cases} \frac{1}{x}, & \text{for } x > 0, \\ +\infty & \text{otherwise,} \end{cases} \qquad \textit{and} \quad g(x) := \begin{cases} -\frac{1}{x}, & \text{for } x < 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Despite that $\bigcup_{\lambda>0} \lambda\,[\operatorname{dom} g - \operatorname{dom} f] = \ ]-\infty, 0[\ \neq \mathbb{R}$, the affine function $\alpha(x) := -x$ satisfies $f \geq \alpha \geq -g$, see Figure 2. $\Diamond$

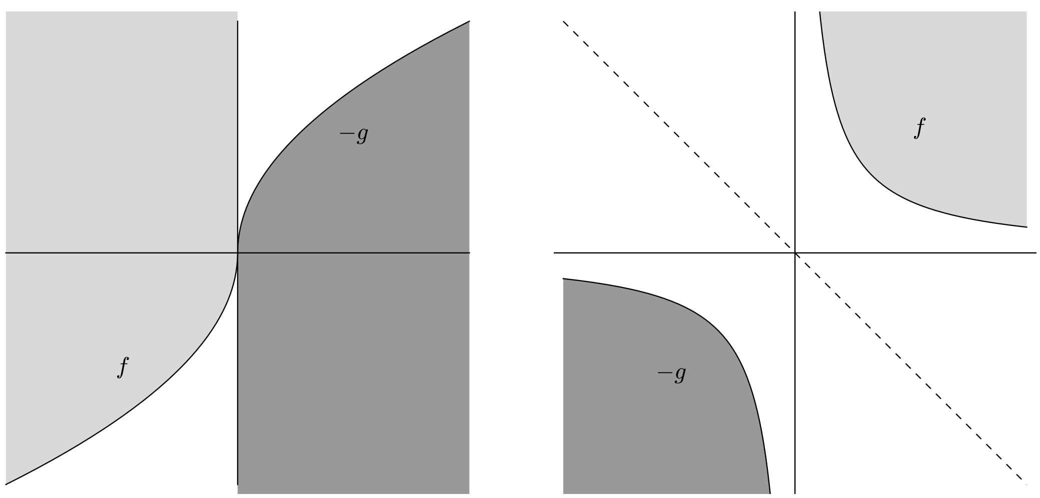

Figure 2: On the left we show the failure of the sandwich theorem in the absence of the constraint qualification; of the right we show that the constraint qualification is not necessary.

Theorem 2.32 (Subdifferential sum rule) *Let X and Y be Banach spaces, and let $f : X \to\ ]-\infty, +\infty]$ and $g : Y \to\ ]-\infty, +\infty]$ be convex functions and let $T : X \to Y$ be a bounded linear mapping. Then at any point $x \in X$ we have the sum rule*

$$\partial(f + g \circ T)(x) \supseteq \partial f(x) + T^*(\partial g(Tx))$$

with equality if (12) or (13) hold.

Proof. The inclusion is straightforward by using the definition of the subdifferential, so we prove the reverse inclusion. Fix any $x \in X$ and let $x^* \in \partial(f + g \circ T)(x)$. Then $0 \in \partial(f - \langle x^*, \cdot\rangle + g \circ T)(x)$.

Conditions for the equality in Theorem 2.24 are satisfied for the functions $f(\cdot)-\langle x^*,\cdot\rangle$ and g. Thus, there exists $y^* \in Y^*$ such that

$$f(x) - \langle x^*, x\rangle + g(Tx) = -f^*(T^*y^* + x^*) - g^*(-y^*).$$

Now set $z^* := T^*y^* + x^*$. Hence, by the Fenchel-Young inequality (4), one has

$$0 \leq f(x) + f^*(z^*) - \langle z^*, x\rangle = -g(Tx) - g^*(-y^*) - \langle T^*y^*, x\rangle \leq 0;$$

whence,

$$\begin{aligned} f(x) + f^*(z^*) &= \langle z^*, x\rangle \\ g(Tx) + g^*(-y^*) &= \langle -y^*, Tx\rangle. \end{aligned}$$

Therefore equality in Fenchel-Young occurs, and one has $z^* \in \partial f(x)$ and $-y^* \in \partial g(Tx)$, which completes the proof. ■

The subdifferential sum rule for two convex functions with a finite common point where one of them is continuous was proved by Rockafellar in 1966 with an argumentation based on Fenchel duality, see [55, Th. 3]. In an earlier work in 1963, Moreau [46] proved the subdifferential sum rule for a pair of convex and lsc functions, in the case that infimal convolution of the conjugate functions is achieved, see [49, p. 63] for more details. Moreau actually proved this result for functions which are the supremum of a family of affine continuous linear functions, a set which agrees with the convex and lsc functions when X is a locally convex vector space, see [44] or [49, p. 28]. See also [36, 37, 27, 19] for more information about the subdifferential calculus rule.

Theorem 2.33 (Hahn–Banach extension) *Let X be a Banach space and let $f\colon X \to \mathbb{R}$ be a continuous sublinear function with* dom $f = X$. *Suppose that L is a linear subspace of X and the function $h\colon L \to \mathbb{R}$ is linear and* dominated *by f, that is, $f \geq h$ on L. Then there exists $x^* \in X^*$, dominated by f, such that*

$$h(x) = \langle x^*, x\rangle, \text{ for all } x \in L.$$

Proof. Take $g := -h + \iota_L$ and apply Theorem 2.24 to f and g with T the identity mapping. Then, there exists $x^* \in X^*$ such that

$$\begin{aligned} 0 &\leq \inf_{x\in X} \{f(x) - h(x) + \iota_L(x)\} \\ &= -f^*(x^*) - \sup_{x\in X}\{\langle -x^*, x\rangle + h(x) - \iota_L(x)\} \\ &= -f^*(x^*) + \inf_{x\in L}\{\langle x^*, x\rangle - h(x)\}; \end{aligned} \tag{16}$$

whence,

$$f^*(x^*) \leq \langle x^*, x\rangle - h(x), \quad \text{for all } x \in L.$$

Observe that $f^*(x^*) \geq 0$ since $f(0) = 0$. Thus, being L a linear subspace, we deduce from the above inequality that

$$h(x) = \langle x^*, x\rangle, \quad \text{for all } x \in L.$$

Then (16) implies $f^*(x^*) = 0$, from where

$$f(x) \geq \langle x^*, x\rangle, \quad \text{for all } x \in X,$$

and we are done. ■

Remark 2.34 (Moreau's max formula, Theorem 2.5)—a true child of Cauchy's principle of steepest descent—can be also derived from Fenchel duality. In fact, the non-emptiness of the subdifferential at a point of continuity, Moreau's max formula, Fenchel duality, the Sandwich theorem, the subdifferential sum rule, and Hahn-Banach extension theorem are all equivalent, in the sense that they are easily inter-derivable.

In outline, one considers $h(u) := \inf_x \big(f(x) + g(Ax + u)\big)$ and checks that $\partial h(0) \neq \emptyset$ implies the Fenchel and Lagrangian duality results; while condition (12) or (13) implies h is continuous at zero and thus Theorem 2.5 finishes the proof. Likewise, the polyhedral calculus [21, §5.1] implies h is polyhedral when f and g are and shows that polyhedral functions have $\operatorname{dom} h = \operatorname{dom} \partial h$. This establishes Theorem 2.27. This also recovers abstract LP duality (e.g., semidefinite programming and conic duality) under condition (12). See [21, 22] for more details. ♢

Let us turn to two illustrations of the power of convex analysis within functional analysis.

A *Banach limit* is a bounded linear functional Λ on the space of bounded sequences of real numbers ℓ^∞ such that

(i) $\Lambda((x_{n+1})_{n\in\mathbb{N}}) = \Lambda((x_n)_{n\in\mathbb{N}})$ (so it only depends on the sequence's tail),

(ii) $\liminf_k x_k \leq \Lambda\big((x_n)_{n\in\mathbb{N}}\big) \leq \limsup_k x_k$

where $(x_n)_{n\in\mathbb{N}} = (x_1, x_2, \ldots) \in \ell^\infty$ and $(x_{n+1})_{n\in\mathbb{N}} = (x_2, x_3, \ldots)$. Thus Λ agrees with the limit on c, the subspace of sequences whose limit exists. Banach limits care peculiar objects!

The Hahn-Banach extension theorem can be used show the existence of Banach limits (see Sucheston [65] or [22, Exercise 5.4.12]). Many of its earliest applications were to summability theory and related fields. We sketch Sucheston's proof as follows.

Theorem 2.35 (Banach limits) (See [65].) *Banach limits exist.*

Proof. Let c be the subspace of convergent sequences in ℓ^∞. Define $f : \ell^\infty \to \mathbb{R}$ by

$$x := (x_n)_{n\in\mathbb{N}} \mapsto \lim_{n\to\infty} \left(\sup_j \frac{1}{n} \sum_{i=1}^{n} x_{i+j} \right). \tag{17}$$

Then f is sublinear with full domain, since the limit in (17) always exists (see [65, p. 309]). Define h on c by $h := \lim_n x_n$ for every $x := (x_n)_{n\in\mathbb{N}}$ in c. Hence h is linear and agrees with f on c. Applying the Hahn-Banach extension Theorem 2.33, there exists $\Lambda \in (\ell^\infty)^*$, dominated by f, such that $\Lambda = h$ on c. Thus Λ extends the limit linearly from c to ℓ^∞. Let S denote the forward shift defined as $S((x_n)_{n\in\mathbb{N}}) := (x_{n+1})_{n\in\mathbb{N}}$. Note that $f(Sx - x) = 0$, since

$$|f(Sx-x)| = \left| \lim_{n\to\infty} \left(\sup_j \frac{1}{n} (x_{j+n+1} - x_{j+1}) \right) \right| \leq \lim_{n\to\infty} \frac{2}{n} \sup_j |x_j| = 0.$$

Thus, $\Lambda(Sx) - \Lambda(x) = \Lambda(Sx - x) \le 0$, and $\Lambda(x) - \Lambda(Sx) = \Lambda(x - Sx) \le f(x - Sx) = 0$; that is, Λ is indeed a Banach limit. ∎

Remark 2.36 One of the referees kindly pointed out that in the proof of Theorem 2.35, the function h can be simply defined by $h : \{0\} \to \mathbb{R}$ with $h(0) = 0$.

Theorem 2.37 (Principle of uniform boundedness) (See ([22, Example 1.4.8].) *Let Y be another Banach space and $T_\alpha \colon X \to Y$ for $\alpha \in \mathcal{A}$ be bounded linear operators. Assume that $\sup_{\alpha \in A} \|T_\alpha(x)\| < +\infty$ for each x in X. Then $\sup_{\alpha \in A} \|T_\alpha\| < +\infty$.*

Proof. Define a function f_A by

$$f_A(x) := \sup_{\alpha \in A} \|T_\alpha(x)\|$$

for each x in X. Then, as observed in Fact 2.1(i), f_A is convex. It is also lower semicontinuous since each mapping $x \mapsto \|T_\alpha(x)\|$ is continuous. Hence f_A is a finite, lower semicontinuous and convex (actually sublinear) function. Now Fact 2.1(iv) ensures f_A is continuous at the origin. Select $\varepsilon > 0$ with $\sup\{f_A(x) \mid \|x\| \le \varepsilon\} \le 1 + f_A(0) = 1$. It follows that

$$\sup_{\alpha \in A} \|T_\alpha\| = \sup_{\alpha \in A} \frac{1}{\varepsilon} \sup_{\|x\| \le \varepsilon} \|T_\alpha(x)\| = \frac{1}{\varepsilon} \sup_{\|x\| \le \varepsilon} \sup_{\alpha \in A} \|T_\alpha(x)\| \le \frac{1}{\varepsilon}.$$

Thus, uniform boundedness is revealed to be continuity of f_A. ∎

3 The Chebyshev problem

Let C be a nonempty subset of X. We define the *nearest point mapping* by

$$P_C(x) := \{v \in C \mid \|v - x\| = \mathrm{d}_C(x)\}.$$

A set C is said to be a *Chebyshev set* if $P_C(x)$ is a singleton for every $x \in X$. If $P_C(x) \neq \varnothing$ for every $x \in X$, then C is said to be *proximal*; the term *proximinal* is also used.

In 1961 Victor Klee [39] posed the following fundamental question: Is every Chebyshev set in a Hilbert space convex? At this stage, it is known that the answer is affirmative for weakly closed sets. In what follows we will present a proof of this fact via convex duality. To this end, we will make use of the following fairly simple lemma.

Lemma 3.1 (See [22, Proposition 4.5.8].) *Let C be a weakly closed Chebyshev subset of a Hilbert space H. Then the nearest point mapping P_C is continuous.*

Theorem 3.2 *Let C be a nonempty weakly closed subset of a Hilbert space H. Then C is convex if and only if C is a Chebyshev set.*

Proof. For the direct implication, we will begin by proving that C is proximal. We can and do suppose that $0 \in C$. Pick any $x \in H$. Consider the convex and lsc functions $f(z) := -\langle x, z\rangle + \iota_{B_H}(z)$ and $g(z) := \sigma_C(z)$. Notice that $\bigcup_{\lambda > 0} \lambda\,[\mathrm{dom}\, g - \mathrm{dom}\, f] = H$ (in fact f is continuous at $0 \in \mathrm{dom}\, f \cap \mathrm{dom}\, g$). With the notation of Theorem 2.24, one has $p = d$, and the supremum of the dual problem is attained if finite. Since $f^*(y) = \|x + y\|$ and $g^*(y) = \iota_C(y)$, as C is closed, the dual problem (11) takes the form

$$d = \sup_{y \in H} \{-\|x + y\| - \iota_C(-y)\} = -\operatorname{d}_C(x).$$

Choose any $c \in C$. Observe that $0 \leq \operatorname{d}_C(x) \leq \|x - c\|$. Therefore the supremum must be attained, and $P_C(x) \neq \varnothing$. Uniqueness follows easily from the convexity of C.

For the converse, consider the function $f := \frac{1}{2}\| \cdot \|^2 + \iota_C$. We first show that

$$\partial f^*(x) = \{P_C(x)\}, \text{ for all } x \in H. \tag{18}$$

Indeed, for $x \in H$,

$$\begin{aligned}
f^*(x) &= \sup_{y \in C} \left\{ \langle x, y \rangle - \frac{1}{2} \langle y, y \rangle \right\} \\
&= \frac{1}{2} \langle x, x \rangle + \frac{1}{2} \sup_{y \in C} \{-\langle x, x \rangle + 2\langle x, y \rangle - \langle y, y \rangle\} \\
&= \frac{1}{2}\|x\|^2 - \frac{1}{2} \inf_{y \in C} \|x - y\|^2 = \frac{1}{2}\|x\|^2 - \frac{1}{2} d_C^2(x) \\
&= \frac{1}{2}\|x\|^2 - \frac{1}{2}\|x - P_C(x)\|^2 = \langle x, P_C(x) \rangle - \frac{1}{2}\|P_C(x)\|^2 \\
&= \langle x, P_C(x) \rangle - f(P_C(x)).
\end{aligned}$$

Consequently, by Proposition 2.7, $P_C(x) \in \partial f^*(x)$ for $x \in X$. Now suppose $y \in \partial f^*(x)$, and define $x_n = x + \frac{1}{n}(y - P_C(x))$. Then $x_n \to x$, and hence $P_C(x_n) \to P_C(x)$ by Lemma 3.1. Using the subdifferential inequality, we have

$$0 \leq \langle x_n - x, P_C(x_n) - y \rangle = \frac{1}{n} \langle y - P_C(x), P_C(x_n) - y \rangle.$$

This now implies:

$$0 \leq \lim_{n \to \infty} \langle y - P_C(x), P_C(x_n) - y \rangle = -\|y - P_C(x)\|^2.$$

Consequently, $y = P_C(x)$ and so (18) is established.

Since f^* is continuous and we just proved that ∂f^* is a singleton, Proposition 2.3 implies that f^* is Gâteaux differentiable. Now $-\infty < f^{**}(x) \leq f(x) = \frac{1}{2}\|x\|^2$ for all $x \in C$. Thus, f^{**} is a proper function. One can easily check that f is sequentially weakly lsc, C being weakly closed. Therefore, Theorem 2.15 implies that f is convex; whence, $\operatorname{dom} f = C$ must be convex. ■

Observe that we have actually proved that *every Chebyshev set with a continuous projection mapping is convex* (and closed). We finish the section by recalling a simple but powerful "hidden convexity" result.

Remark 3.3 (See [5].) Let C be a closed subset of a Hilbert space H. Then there exists a continuous and convex function f defined on H such that $\operatorname{d}_C^2(x) = \|x\|^2 - f(x)$, $\forall x \in H$. Precisely, f can be taken as $x \mapsto \sup_{c \in C}\{2\langle x, c \rangle - \|c\|^2\}$.

4 Monotone operator theory

Let $A\colon X \rightrightarrows X^*$ be a *set-valued operator* (also known as a relation, point-to-set mapping or multifunction), i.e., for every $x \in X$, $Ax \subseteq X^*$, and let $\operatorname{gra} A := \big\{(x,x^*) \in X \times X^* \mid x^* \in Ax\big\}$ be the *graph* of A. The *domain* of A is $\operatorname{dom} A := \big\{x \in X \mid Ax \neq \varnothing\big\}$ and $\operatorname{ran} A := A(X)$ is the *range* of A. We say that A is *monotone* if

$$\langle x - y, x^* - y^* \rangle \geq 0, \quad \text{for all } (x,x^*), (y,y^*) \in \operatorname{gra} A, \tag{19}$$

and *maximally monotone* if A is monotone and A has no proper monotone extension (in the sense of graph inclusion). Given A monotone, we say that $(x,x^*) \in X \times X^*$ is *monotonically related to* $\operatorname{gra} A$ if

$$\langle x - y, x^* - y^* \rangle \geq 0, \quad \text{for all } (y,y^*) \in \operatorname{gra} A.$$

Monotone operators have frequently shown themselves to be a key class of objects in both modern Optimization and Analysis; see, e.g., [13, 14, 15, 24], the books [7, 22, 28, 53, 61, 62, 58, 66, 67, 68] and the references given therein.

Given sets $S \subseteq X$ and $D \subseteq X^*$, we define $S^\perp$ by $S^\perp := \{x^* \in X^* \mid \langle x^*, x\rangle = 0, \quad \forall x \in S\}$ and $D_\perp$ by $D_\perp := \{x \in X \mid \langle x, x^*\rangle = 0, \quad \forall x^* \in D\}$ [54]. Then the *adjoint* of A is the operator $A^* : X^{**} \rightrightarrows X^*$ such that

$$\operatorname{gra} A^* := \big\{(x^{**}, x^*) \in X^{**} \times X^* \mid (x^*, -x^{**}) \in (\operatorname{gra} A)^\perp\big\}.$$

Note that the adjoint is always a *linear relation*, i.e. its graph is a linear subspace.

The *Fitzpatrick function* [33] associated with an operator A is the function $F_A : X \times X^* \to \left]-\infty, +\infty\right]$ defined by

$$F_A(x,x^*) := \sup_{(a,a^*) \in \operatorname{gra} A} \Big(\langle x, a^*\rangle + \langle a, x^*\rangle - \langle a, a^*\rangle\Big). \tag{20}$$

Fitzpatrick functions have been proved to be an important tool in modern monotone operator theory. One of the main reasons is shown in the following result.

Fact 4.1 (Fitzpatrick) (See ([33, Propositions 3.2&4.2, Theorem 3.4 and Corollary 3.9].) *Let $A\colon X \rightrightarrows X^*$ be monotone with $\operatorname{dom} A \neq \varnothing$. Then F_A is proper lower semicontinuous in the norm $\times$ weak*-topology $\omega(X^*, X)$, convex, and $F_A = \langle \cdot, \cdot \rangle$ on $\operatorname{gra} A$. Moreover, if A is maximally monotone, for every $(x,x^*) \in X \times X^*$, the inequality*

$$\langle x, x^*\rangle \leq F_A(x,x^*) \leq F_A^*(x^*, x)$$

is true, and the first equality holds if and only if $(x,x^*) \in \operatorname{gra} A$.

The next result is central to maximal monotone operator theory and algorithmic analysis. Originally it was proved by more extended direct methods than the concise convex analysis argument we present next.

Theorem 4.2 (Local boundedness) (See [53, Theorem 2.2.8].) *Let* $A : X \rightrightarrows X^*$ *be monotone with* $\operatorname{int}\operatorname{dom} A \neq \varnothing$. *Then* A *is locally bounded at* $x \in \operatorname{int}\operatorname{dom} A$, *i.e., there exist* $\delta > 0$ *and* $K > 0$ *such that*

$$\sup_{y^* \in Ay} \|y^*\| \leq K, \quad \forall y \in x + \delta B_X.$$

Proof. Let $x \in \operatorname{int}\operatorname{dom} A$. After translating the graphs if necessary, we can and do suppose that $x = 0$ and $(0,0) \in \operatorname{gra} A$. Define $f : X \to \left]-\infty, +\infty\right]$ by

$$y \mapsto \sup_{(a,a^*) \in \operatorname{gra} A,\, \|a\| \leq 1} \langle y - a, a^* \rangle.$$

By Fact 2.1(i), f is convex and lower semicontinuous. Since $0 \in \operatorname{int}\operatorname{dom} A$, there exists $\delta_1 > 0$ such that $\delta_1 B_X \subseteq \operatorname{dom} A$. Now we show that $\delta_1 B_X \subseteq \operatorname{dom} f$. Let $y \in \delta_1 B_X$ and $y^* \in Ay$. Thence, we have

$$\begin{aligned}
&\langle y - a, y^* - a^* \rangle \geq 0, \quad \forall (a, a^*) \in \operatorname{gra} A,\, \|a\| \leq 1 \\
&\Rightarrow \langle y - a, y^* \rangle \geq \langle y - a, a^* \rangle, \quad \forall (a, a^*) \in \operatorname{gra} A,\, \|a\| \leq 1 \\
&\Rightarrow +\infty > (\|y\| + 1) \cdot \|y^*\| \geq \langle y - a, a^* \rangle, \quad \forall (a, a^*) \in \operatorname{gra} A,\, \|a\| \leq 1 \\
&\Rightarrow f(y) < +\infty \quad \Rightarrow y \in \operatorname{dom} f.
\end{aligned}$$

Hence $\delta_1 B_X \subseteq \operatorname{dom} f$ and thus $0 \in \operatorname{int}\operatorname{dom} f$. By Fact 2.1(iv), there is $\delta > 0$ with $\delta \leq \min\{\frac{1}{2}, \frac{1}{2}\delta_1\}$ such that

$$f(y) \leq f(0) + 1, \quad \forall y \in 2\delta B_X.$$

Now we show that $f(0) = 0$. Since $(0,0) \in \operatorname{gra} A$, then $f(0) \geq 0$. On the other hand, by the monotonicity of A, $\langle a, a^* \rangle = \langle a - 0, a^* - 0 \rangle \geq 0$ for every $(a, a^*) \in \operatorname{gra} A$. Then we have $f(0) = \sup_{(a,a^*) \in \operatorname{gra} A,\, \|a\| \leq 1} \langle 0 - a, a^* \rangle \leq 0$. Thence $f(0) = 0$.

Thus,

$$\langle y, a^* \rangle \leq \langle a, a^* \rangle + 1, \quad \forall y \in 2\delta B_X, (a, a^*) \in \operatorname{gra} A,\, \|a\| \leq \delta,$$

whence, taking the supremum with $y \in 2\delta B_X$,

$$\begin{aligned}
&2\delta \|a^*\| \leq \|a\| \cdot \|a^*\| + 1 \leq \delta \|a^*\| + 1, \quad \forall (a, a^*) \in \operatorname{gra} A,\, a \in \delta B_X \\
&\Rightarrow \|a^*\| \leq \frac{1}{\delta}, \quad \forall (a, a^*) \in \operatorname{gra} A,\, a \in \delta B_X.
\end{aligned}$$

Setting $K := \frac{1}{\delta}$, we get the desired result. ■

Generalizations of Theorem 4.2 can be found in [62, 18] and [23, Lemma 4.1].

4.1 Sum theorem and Minty surjectivity theorem

In the early 1960s, Minty [43] presented an important characterization of maximally monotone operators in a Hilbert space; which we now reestablish. The proof we give of Theorem 4.3 is due to Simons and Zălinescu [63, Theorem 1.2]. We denote by Id the *identity mapping* from H to H.

Theorem 4.3 (Minty) *Suppose that H is a Hilbert space. Let $A : H \rightrightarrows H$ be monotone. Then A is maximally monotone if and only if* $\operatorname{ran}(A + \operatorname{Id}) = H$.

Proof. "⇒": Fix any $x_0^* \in H$, and let $B : H \rightrightarrows H$ be given by $\operatorname{gra} B := \operatorname{gra} A - \{(0, x_0^*)\}$. Then B is maximally monotone. Define $F : H \times H \to\ ]-\infty, +\infty]$ by

$$(x, x^*) \mapsto F_B(x, x^*) + \frac{1}{2}\|x\|^2 + \frac{1}{2}\|x^*\|^2. \tag{21}$$

Fact 4.1 together with Fact 2.1(v) implies that F is coercive. By [66, Theorem 2.5.1(ii)], F has a minimizer. Assume that $(z, z^*) \in H \times H$ is a minimizer of F. Then we have $(0, 0) \in \partial F(z, z^*)$. Thus, $(0, 0) \in \partial F_B(z, z^*) + (z, z^*)$ and $(-z, -z^*) \in \partial F_B(z, z^*)$. Then

$$\big\langle(-z, -z^*), (b, b^*) - (z, z^*)\big\rangle \leq F_B(b, b^*) - F_B(z, z^*), \quad \forall (b, b^*) \in \operatorname{gra} B,$$

and by Fact 4.1,

$$\big\langle(-z, -z^*), (b, b^*) - (z, z^*)\big\rangle \leq \langle b, b^*\rangle - \langle z, z^*\rangle, \quad \forall (b, b^*) \in \operatorname{gra} B;$$

that is,

$$0 \leq \langle b, b^*\rangle - \langle z, z^*\rangle + \langle z, b\rangle + \langle z^*, b^*\rangle - \|z\|^2 - \|z^*\|^2, \quad \forall (b, b^*) \in \operatorname{gra} B. \tag{22}$$

Hence,

$$\big\langle b + z^*, b^* + z\big\rangle = \langle b, b^*\rangle + \langle z, b\rangle + \langle z^*, b^*\rangle + \langle z, z^*\rangle \geq \|z + z^*\|^2 \geq 0, \quad \forall (b, b^*) \in \operatorname{gra} B,$$

which implies that $(-z^*, -z) \in \operatorname{gra} B$, since B is maximally monotone. This combined with (22) implies $0 \leq -2\langle z, z^*\rangle - \|z\|^2 - \|z^*\|^2$. Then we have $z = -z^*$, and $(z, -z) = (-z^*, -z) \in \operatorname{gra} B$, whence $(z, -z) + (0, x_0^*) \in \operatorname{gra} A$. Therefor $x_0^* \in Az + z$, which implies $x_0^* \in \operatorname{ran}(A + \operatorname{Id})$.

"⇐": Let $(v, v^*) \in H \times H$ be monotonically related to $\operatorname{gra} A$. Since $\operatorname{ran}(A + \operatorname{Id}) = H$, there exists $(y, y^*) \in \operatorname{gra} A$ such that $v^* + v = y^* + y$. Then we have

$$-\|v - y\|^2 = \big\langle v - y, y^* + y - v - y^*\big\rangle = \big\langle v - y, v^* - y^*\big\rangle \geq 0.$$

Hence $v = y$, which also implies $v^* = y^*$. Thus $(v, v^*) \in \operatorname{gra} A$, and therefore A is maximally monotone. ■

Remark 4.4 The extension of Minty's theorem to reflexive spaces (in which case it asserts the surjectivity of $A + J_X$ for the normalized duality mapping J_X defined below) was originally proved by Rockafellar. The proof given in [22, Proposition 3.5.6, page 119] which uses Fenchel's duality theorem more directly than the one we gave here, is only slightly more complicated than that of Theorem 4.3.

Let A and B be maximally monotone operators from X to X^*. Clearly, the *sum operator* $A + B\colon X \rightrightarrows X^*\colon x \mapsto Ax + Bx := \big\{a^* + b^* \mid a^* \in Ax \text{ and } b^* \in Bx\big\}$ is monotone. Rockafellar established the following important result in 1970 [57], the so-called "sum theorem": Suppose that X is reflexive. If $\operatorname{dom} A \cap \operatorname{int} \operatorname{dom} B \neq \varnothing$, then $A + B$ is maximally monotone. We can weaken this constraint qualification to be that $\bigcup_{\lambda>0}\lambda\,[\operatorname{dom}A-\operatorname{dom}B]$ is a closed subspace (see [4, 62, 64, 22, 2]).

We turn to a new proof of this generalized result. To this end, we need the following fact along with the definition of the partial inf-convolution. Given two real Banach spaces X,Y and $F_1,F_2\colon X\times Y\to\,]-\infty,+\infty]$, the *partial inf-convolution* $F_1\Box_2F_2$ is the function defined on $X\times Y$ by

$$F_1\Box_2F_2\colon (x,y)\mapsto \inf_{v\in Y}\big\{F_1(x,y-v)+F_2(x,v)\big\}.$$

.

Fact 4.5 (Simons and Zălinescu) (See [64, Theorem 4.2] or [62, Theorem 16.4(a)].) *Let X,Y be real Banach spaces and $F_1,F_2\colon X\times Y\to\,]-\infty,+\infty]$ be proper lower semicontinuous and convex bifunctionals. Assume that for every $(x,y)\in X\times Y$,*

$$(F_1\Box_2F_2)(x,y)>-\infty$$

and that $\bigcup_{\lambda>0}\lambda\,[P_X\operatorname{dom}F_1-P_X\operatorname{dom}F_2]$ is a closed subspace of X. Then for every $(x^*,y^*)\in X^*\times Y^*$,

$$(F_1\Box_2F_2)^*(x^*,y^*)=\min_{u^*\in X^*}\left\{F_1^*(x^*-u^*,y^*)+F_2^*(u^*,y^*)\right\}.$$

We denote by J_X *the duality map* from X to X^*, which will be simply written as J, i.e., the subdifferential of the function $\frac{1}{2}\|\cdot\|^2$. Let $F\colon X\times Y\to\,]-\infty,+\infty]$ be a bifunctional defined on two real Banach spaces. Following the notation by Penot [51] we set

$$F^{\intercal}\colon Y\times X\colon (y,x)\mapsto F(x,y). \tag{23}$$

Theorem 4.6 (Sum theorem) *Suppose that X is reflexive. Let $A,B:X\rightrightarrows X$ be maximally monotone. Assume that $\bigcup_{\lambda>0}\lambda\,[\operatorname{dom}A-\operatorname{dom}B]$ is a closed subspace. Then $A+B$ is maximally monotone.*

Proof. Clearly, $A+B$ is monotone. Assume that $(z,z^*)\in X\times X^*$ is monotonically related to $\operatorname{gra}(A+B)$.

Let $F_1:=F_A\Box_2F_B$, and $F_2:=F_1^{*\intercal}$. By [9, Lemma 5.8], $\bigcup_{\lambda>0}\lambda\,[P_X(\operatorname{dom}F_A)-P_X(\operatorname{dom}F_B)]$ is a closed subspace. Then Fact 4.5 implies that

$$F_1^*(x^*,x)=\min_{u^*\in X^*}\left\{F_A^*(x^*-u^*,x)+F_B^*(u^*,x)\right\},\quad \text{for all } (x,x^*)\in X\times X^*. \tag{24}$$

Set $G:X\times X^*\to\,]-\infty,+\infty]$ by

$$(x,x^*)\mapsto F_2(x+z,x^*+z^*)-\langle x,z^*\rangle-\langle z,x^*\rangle+\frac{1}{2}\|x\|^2+\frac{1}{2}\|x^*\|^2.$$

Assume that $(x_0,x_0^*)\in X\times X^*$ is a minimizer of G. ([66, Theorem 2.5.1(ii)] implies that minimizers exist since G is coercive). Then we have $(0,0)\in\partial G(x_0,x_0^*)$. Thus, there exists $v^*\in Jx_0, v\in J_{X^*}x_0^*$ such that $(0,0)\in\partial F_2(x_0+z,x_0^*+z^*)+(v^*,v)+(-z^*,-z)$, and then

$$(z^*-v^*,z-v)\in\partial F_2(x_0+z,x_0^*+z^*).$$

Thence

$$\Big\langle (z^* - v^*, z - v), (x_0 + z, x_0^* + z^*) \Big\rangle = F_2(x_0 + z, x_0^* + z^*) + F_2^*(z^* - v^*, z - v). \tag{25}$$

Fact 4.1 and (24) show that

$$F_2 \geq \langle \cdot, \cdot \rangle, \quad F_2^{*\intercal} = \overline{F_1} \geq \langle \cdot, \cdot \rangle.$$

Then by (25),

$$\begin{aligned} \Big\langle (z^* - v^*, z - v), (x_0 + z, x_0^* + z^*) \Big\rangle &= F_2(x_0 + z, x_0^* + z^*) + F_2^*(z^* - v^*, z - v) \\ &\geq \big\langle x_0 + z, x_0^* + z^* \big\rangle + \big\langle z^* - v^*, z - v \big\rangle. \end{aligned} \tag{26}$$

Thus, since $v^* \in Jx_0, v \in J_{X^*} x_0^*$,

$$\begin{aligned} 0 \leq \delta &:= \Big\langle (z^* - v^*, z - v), (x_0 + z, x_0^* + z^*) \Big\rangle - \big\langle x_0 + z, x_0^* + z^* \big\rangle - \big\langle z^* - v^*, z - v \big\rangle \\ &= \big\langle - x_0 - v, x_0^* + v^* \big\rangle = \langle -x_0, x_0^* \rangle - \langle x_0, v^* \rangle - \langle v, x_0^* \rangle - \langle v, v^* \rangle \\ &= \langle -x_0, x_0^* \rangle - \frac{1}{2}\|x_0^*\|^2 - \frac{1}{2}\|x_0\|^2 - \frac{1}{2}\|v^*\|^2 - \frac{1}{2}\|v\|^2 - \langle v, v^* \rangle, \end{aligned}$$

which implies

$$\delta = 0 \quad \text{and} \quad \langle x_0, x_0^* \rangle + \frac{1}{2}\|x_0^*\|^2 + \frac{1}{2}\|x_0\|^2 = 0;$$

that is,

$$\delta = 0 \quad \text{and} \quad x_0^* \in -Jx_0. \tag{27}$$

Combining (26) and (27), we have $F_2(x_0 + z, x_0^* + z^*) = \big\langle x_0 + z, x_0^* + z^* \big\rangle$. By (24) and Fact 4.1,

$$(x_0 + z, x_0^* + z^*) \in \operatorname{gra}(A + B). \tag{28}$$

Since (z, z^*) is monotonically related to $\operatorname{gra}(A + B)$, it follows from (28) that

$$\big\langle x_0, x_0^* \big\rangle = \big\langle x_0 + z - z, x_0^* + z^* - z^* \big\rangle \geq 0,$$

and then by (27),

$$-\|x_0\|^2 = -\|x_0^*\|^2 \geq 0,$$

whence $(x_0, x_0^*) = (0, 0)$. Finally, by (28), one deduces that $(z, z^*) \in \operatorname{gra}(A + B)$ and $A + B$ is maximally monotone. ■

It is still unknown whether the reflexivity condition can be omitted in Theorem 4.6 though many partial results exist, see [14, 15] and [22, §9.7].

4.2 Autoconjugate functions

Given $F\colon X\times X^* \to \left]-\infty,+\infty\right]$, we say that F is *autoconjugate* if $F=F^{*\intercal}$ on $X\times X^*$. We say F is a *representer* for $\operatorname{gra} A$ if

$$\operatorname{gra} A = \big\{(x,x^*)\in X\times X^* \mid F(x,x^*)=\langle x,x^*\rangle\big\}. \tag{29}$$

Autoconjugate functions are the core of representer theory, which has been comprehensively studied in Optimization and Partial Differential Equations (see [8, 9, 52, 62, 22, 34]).

Fitzpatrick posed the following question in [33, Problem 5.5]:

> *If $A: X \rightrightarrows X^*$ is maximally monotone, does there necessarily exist an autoconjugate representer for A?*

Bauschke and Wang gave an affirmative answer to the above question in reflexive spaces by construction of the function $\mathcal{B}_A$ in Fact 4.7. The first construction of an autoconjugate representer for a maximally monotone operator satisfying a mild constraint qualification in a reflexive space was provided by Penot and Zălinescu in [52]. This naturally raises a question:

> Is $\mathcal{B}_A$ still an autoconjugate representer for a maximally monotone operator A in a general Banach space?

We give a negative answer to the above question in Example 4.12: in certain spaces, $\mathcal{B}_A$ fails to be autoconjugate.

Fact 4.7 (Bauschke and Wang) (See [8, Theorem 5.7].) *Suppose that X is reflexive. Let $A\colon X\rightrightarrows X^*$ be maximally monotone. Then*

$$\begin{aligned}&\mathcal{B}_A\colon X\times X^*\to\left]-\infty,+\infty\right]\\ &(x,x^*)\mapsto \inf_{(y,y^*)\in X\times X^*}\Big\{\tfrac12 F_A(x+y,x^*+y^*)+\tfrac12 F_A^{*\intercal}(x-y,x^*-y^*)+\tfrac12\|y\|^2+\tfrac12\|y^*\|^2\Big\}\end{aligned} \tag{30}$$

is an autoconjugate representer for A.

We will make use of the following result to prove Theorem 4.11 below.

Fact 4.8 (Simons) (See [62, Corollary 10.4].) *Let $f_1,f_2,g\colon X\to\left]-\infty,+\infty\right]$ be proper convex. Assume that g is continuous at a point of* $\operatorname{dom} f_1-\operatorname{dom} f_2$. *Suppose that*

$$h(x):=\inf_{z\in X}\big\{\tfrac12 f_1(x+z)+\tfrac12 f_2(x-z)+\tfrac14 g(2z)\big\}>-\infty,\quad \forall x\in X.$$

Then

$$h^*(x^*)=\min_{z^*\in X^*}\big\{\tfrac12 f_1^*(x^*+z^*)+\tfrac12 f_2^*(x^*-z^*)+\tfrac14 g^*(-2z^*)\big\},\quad \forall x^*\in X^*.$$

Let $A : X \rightrightarrows X^*$ be a linear relation. We say that A is *skew* if $\operatorname{gra} A \subseteq \operatorname{gra}(-A^*)$; equivalently, if $\langle x, x^*\rangle = 0,\ \forall (x,x^*) \in \operatorname{gra} A$. Furthermore, A is *symmetric* if $\operatorname{gra} A \subseteq \operatorname{gra} A^*$; equivalently, if $\langle x, y^*\rangle = \langle y, x^*\rangle$, $\forall (x,x^*), (y,y^*) \in \operatorname{gra} A$. We define the *symmetric part* and the *skew part* of A via

$$P := \tfrac{1}{2}A + \tfrac{1}{2}A^* \quad \text{and} \quad S := \tfrac{1}{2}A - \tfrac{1}{2}A^*, \tag{31}$$

respectively. It is easy to check that P is symmetric and that S is skew.

Fact 4.9 (See [6, Theorem 3.7].) *Let $A : X^* \to X^{**}$ be linear and continuous. Assume that* $\operatorname{ran} A \subseteq X$ *and that there exists $e \in X^{**}\backslash X$ such that*

$$\langle Ax^*, x^*\rangle = \langle e, x^*\rangle^2, \quad \forall x^* \in X^*.$$

Let P and S respectively be the symmetric part and skew part of A. Let $T : X \rightrightarrows X^*$ be defined by

$$\operatorname{gra} T := \big\{(-Sx^*, x^*) \mid x^* \in X^*, \langle e, x^*\rangle = 0\big\} = \big\{(-Ax^*, x^*) \mid x^* \in X^*, \langle e, x^*\rangle = 0\big\}. \tag{32}$$

Then the following hold.

(i) *A is a maximally monotone operator on X^* .*

(ii) $Px^* = \langle x^*, e\rangle e,\ \forall x^* \in X^*$.

(iii) *T is maximally monotone and skew on X.*

(iv) $\operatorname{gra} T^* = \{(Sx^* + re, x^*) \mid x^* \in X^*,\ r \in \mathbb{R}\}$.

(v) *$F_T = \iota_C$, where $C := \{(-Ax^*, x^*) \mid x^* \in X^*\}$.*

We next give concrete examples of A, T as in Fact 4.9.

Example 4.10 (c_0) (See [6, Example 4.1].) Let $X := c_0$, with norm $\|\cdot\|_\infty$ so that $X^* = \ell^1$ with norm $\|\cdot\|_1$, and $X^{**} = \ell^\infty$ with its second dual norm $\|\cdot\|_*$ (i.e., $\|y\|_* := \sup_{n\in\mathbb{N}} |y_n|$, $\forall y := (y_n)_{n\in\mathbb{N}} \in \ell^\infty$). Fix $\alpha := (\alpha_n)_{n\in\mathbb{N}} \in \ell^\infty$ with $\limsup \alpha_n \neq 0$, and let $A_\alpha : \ell^1 \to \ell^\infty$ be defined by

$$(A_\alpha x^*)_n := \alpha_n^2 x_n^* + 2\sum_{i>n} \alpha_n \alpha_i x_i^*, \quad \forall x^* = (x_n^*)_{n\in\mathbb{N}} \in \ell^1. \tag{33}$$

Now let P_α and S_α respectively be the symmetric part and skew part of A_α. Let $T_\alpha : c_0 \rightrightarrows X^*$ be defined by

$$\begin{aligned}\operatorname{gra} T_\alpha &:= \big\{(-S_\alpha x^*, x^*) \mid x^* \in X^*, \langle \alpha, x^*\rangle = 0\big\} = \big\{(-A_\alpha x^*, x^*) \mid x^* \in X^*, \langle \alpha, x^*\rangle = 0\big\} \\ &= \big\{\big((-\sum_{i>n} \alpha_n\alpha_i x_i^* + \sum_{i<n} \alpha_n\alpha_i x_i^*)_{n\in\mathbb{N}}, x^*\big) \mid x^* \in X^*, \langle \alpha, x^*\rangle = 0\big\}.\end{aligned} \tag{34}$$

Then

(i) $\langle A_\alpha x^*, x^*\rangle = \langle \alpha, x^*\rangle^2, \quad \forall x^* = (x^*_n)_{n\in\mathbb{N}} \in \ell^1$ and (34) is well defined.

(ii) A_α is a maximally monotone.

(iii) T_α is a maximally monotone operator.

(iv) Let $G : \ell^1 \to \ell^\infty$ be *Gossez's operator* [35] defined by

$$\big(G(x^*)\big)_n := \sum_{i>n} x^*_i - \sum_{i<n} x^*_i, \quad \forall (x^*_n)_{n\in\mathbb{N}} \in \ell^1.$$

Then $T_e : c_0 \rightrightarrows \ell^1$ as defined by

$$\operatorname{gra} T_e := \{(-G(x^*), x^*) \mid x^* \in \ell^1, \langle x^*, e\rangle = 0\}$$

is a maximally monotone operator, where $e := (1, 1, \ldots, 1, \ldots)$. $\Diamond$

We may now show that $\mathcal{B}_T$ need not be autoconjugate.

Theorem 4.11 *Let $A : X^* \to X^{**}$ be linear and continuous. Assume that* $\operatorname{ran} A \subseteq X$ *and that there exists $e \in X^{**}\backslash X$ such that $\|e\| < \frac{1}{\sqrt{2}}$ and*

$$\langle Ax^*, x^*\rangle = \langle e, x^*\rangle^2, \quad \forall x^* \in X^*.$$

Let P and S respectively be the symmetric part and skew part of A. Let T, C be defined as in Fact 4.9. Then

$$\mathcal{B}_T(-Aa^*, a^*) > \mathcal{B}^*_T(a^*, -Aa^*), \quad \forall a^* \notin \{e\}_\perp.$$

In consequence, $\mathcal{B}_T$ is not autoconjugate.

Proof. First we claim that

$$\iota^{*\intercal}_C|_{X\times X^*} = \iota_{\operatorname{gra} T}. \tag{35}$$

Clearly, if we set $D := \{(A^*x^*, x^*) \mid x^* \in X^*\}$, we have

$$\iota^{*\intercal}_C = \sigma^\intercal_C = \iota^\intercal_{C^\perp} = \iota_D, \tag{36}$$

where in the second equality we use the fact that C is a subspace. Additionally,

$$\begin{aligned}
A^*x^* \in X &\Leftrightarrow (S+P)^*x^* \in X \Leftrightarrow S^*x^* + P^*x^* \in X \Leftrightarrow -Sx^* + Px^* \in X\\
&\Leftrightarrow -Sx^* - Px^* + 2Px^* \in X \Leftrightarrow 2Px^* - Ax^* \in X \Leftrightarrow Px^* \in X \quad (\text{since } \operatorname{ran} A \subseteq X)\\
&\Leftrightarrow \langle x^*, e\rangle e \in X \quad (\text{by Fact 4.9(ii)})\\
&\Leftrightarrow \langle x^*, e\rangle = 0 \quad (\text{since } e \notin X).
\end{aligned} \tag{37}$$

Observe that $Px^* = 0$ for all $x^* \in \{e\}_\perp$ by Fact 4.9(ii). Thus, $A^*x^* = -Ax^*$ for all $x^* \in \{e\}_\perp$. Combining (36) and (37), we have

$$\iota^{*\intercal}_C|_{X\times X^*} = \iota_{D\cap(X\times X^*)} = \iota_{\operatorname{gra} T},$$

and hence (35) holds.

Let $a^* \notin \{e\}_\perp$. Then $\langle a^*, e\rangle \neq 0$. Now we compute $\mathcal{B}_T(-Aa^*, a^*)$. By Fact 4.9(v) and (35),

$$\begin{aligned} &\mathcal{B}_T(-Aa^*, a^*)\\ &= \inf_{(y,y^*)\in X\times X^*} \left\{\iota_C(-Aa^* + y, a^* + y^*) + \iota_{\operatorname{gra} T}(-Aa^* - y, a^* - y^*) + \tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}\|y^*\|^2\right\}. \end{aligned} \tag{38}$$

Thus

$$\begin{aligned} \mathcal{B}_T(-Aa^*, a^*) &= \inf_{y=-Ay^*} \left\{\iota_{\operatorname{gra} T}(-Aa^* - y, a^* - y^*) + \tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}\|y^*\|^2\right\}\\ &= \inf_{y=-Ay^*,\, \langle a^*-y^*, e\rangle=0} \left\{\tfrac{1}{2}\|y\|^2 + \tfrac{1}{2}\|y^*\|^2\right\} = \inf_{\langle a^*-y^*, e\rangle=0} \left\{\tfrac{1}{2}\|Ay^*\|^2 + \tfrac{1}{2}\|y^*\|^2\right\}\\ &\geq \inf_{\langle a^*-y^*, e\rangle=0} \langle Ay^*, y^*\rangle = \inf_{\langle a^*-y^*, e\rangle=0} \langle e, y^*\rangle^2\\ &= \langle e, a^*\rangle^2. \end{aligned} \tag{39}$$

Next we will compute $\mathcal{B}_T^*(a^*, -Aa^*)$. By Fact 4.8 and (38), we have

$$\begin{aligned} &\mathcal{B}_T^*(a^*, -Aa^*)\\ &= \min_{(y^*,y^{**})\in X^*\times X^{**}} \left\{\frac{1}{2}\iota_C^*(a^* + y^*, -Aa^* + y^{**}) + \frac{1}{2}\iota_{\operatorname{gra} T}^*(a^* - y^*, -Aa^* - y^{**}) + \tfrac{1}{2}\|y^{**}\|^2 + \tfrac{1}{2}\|y^*\|^2\right\}\\ &= \min_{(y^*,y^{**})\in X^*\times X^{**}} \left\{\iota_D(-Aa^* + y^{**}, a^* + y^*) + \iota_{(\operatorname{gra} T)^\perp}(a^* - y^*, -Aa^* - y^{**}) + \tfrac{1}{2}\|y^{**}\|^2 + \tfrac{1}{2}\|y^*\|^2\right\} \quad \text{(by (36))}\\ &\leq \iota_D(-Aa^* + 2Pa^*, a^*) + \iota_{(\operatorname{gra} T)^\perp}(a^*, -Aa^* - 2Pa^*) + \tfrac{1}{2}\|2Pa^*\|^2 \quad \text{(by taking } y^* = 0, y^{**} = 2Pa^*)\\ &= \iota_{\operatorname{gra}(-T^*)}(-Aa^* - 2Pa^*, a^*) + \tfrac{1}{2}\|2Pa^*\|^2\\ &= \tfrac{1}{2}\|2Pa^*\|^2 \quad \text{(by Fact 4.9(iv))}\\ &= \tfrac{1}{2}\|2\langle a^*, e\rangle e\|^2 \quad \text{(by Fact 4.9(ii))}\\ &= 2\langle a^*, e\rangle^2\|e\|^2. \end{aligned}$$

This inequality along with (39), $\langle e, a^*\rangle \neq 0$ and $\|e\| < \frac{1}{\sqrt{2}}$, yield

$$\mathcal{B}_T(-Aa^*, a^*) \geq \langle e, a^*\rangle^2 > 2\langle a^*, e\rangle^2\|e\|^2 \geq \mathcal{B}_T^*(a^*, -Aa^*), \quad \forall a^* \notin \{e\}_\perp.$$

Hence $\mathcal{B}_T$ is not autoconjugate. ■

Example 4.12 (Example 4.10 revisited) Let $X := c_0$, with norm $\|\cdot\|_\infty$ so that $X^* = \ell^1$ with norm $\|\cdot\|_1$, and $X^{**} = \ell^\infty$ with its second dual norm $\|\cdot\|_*$. Fix $\alpha := (\alpha_n)_{n\in\mathbb{N}} \in \ell^\infty$ with $\limsup \alpha_n \neq 0$ and $\|\alpha\|_* < \frac{1}{\sqrt{2}}$, and let $A_\alpha : \ell^1 \to \ell^\infty$ be defined by

$$(A_\alpha x^*)_n := \alpha_n^2 x_n^* + 2\sum_{i>n} \alpha_n\alpha_i x_i^*, \quad \forall x^* = (x_n^*)_{n\in\mathbb{N}} \in \ell^1. \tag{40}$$

Now let P_α and S_α respectively be the symmetric part and skew part of A_α. Let $T_\alpha : c_0 \rightrightarrows X^*$ be defined by

$$\operatorname{gra} T_\alpha := \left\{(-S_\alpha x^*, x^*) \mid x^* \in X^*, \langle \alpha, x^*\rangle = 0\right\} = \left\{(-A_\alpha x^*, x^*) \mid x^* \in X^*, \langle \alpha, x^*\rangle = 0\right\}$$

$$(41) \qquad = \big\{\big((-\sum_{i>n}\alpha_n\alpha_i x_i^* + \sum_{i<n}\alpha_n\alpha_i x_i^*)_{n\in\mathbb{N}}, x^*\big) \mid x^*\in X^*, \langle\alpha, x^*\rangle = 0\big\}.$$

Then, by Example 4.10 and Theorem 4.11,

$$\mathcal{B}_{T_\alpha}(-Aa^*, a^*) > \mathcal{B}^*_{T_\alpha}(a^*, -Aa^*), \quad \forall a^* \notin \{e\}_\perp.$$

In consequence, $\mathcal{B}_{T_\alpha}$ is not autoconjugate. ◊

The latter raises a very interesting question:

Problem 4.13 *Is there a maximally monotone operator on some (resp. every) non-reflexive Banach space that has no autoconjugate representer?*

4.3 The Fitzpatrick function and differentiability

The *Fitzpatrick function* introduced in [33] was discovered precisely to provide a more transparent convex alternative to the earlier saddle function construction due to Krauss [22]—we have not discussed saddle-functions but they produce interesting maximally monotone operators [57, §33 & §37]. At the time, Fitzpatrick's interests were more centrally in the differentiation theory for convex functions and monotone operators.

The search for results relating when a maximally monotone T is single-valued to differentiability of F_T did not yield fruit, and he put the function aside. This is still the one area where to the best of our knowledge F_T has proved of very little help—in part because generic properties of $\operatorname{dom} F_T$ and of $\operatorname{dom}(T)$ seem poorly related.

That said, monotone operators often provide efficient ways to prove differentiability of convex functions. The discussion of Mignot's theorem in[22] is somewhat representative of how this works as is the treatment in [53]. By contrast, as we have seen the Fitzpatrick function and its relatives now provide the easiest access to a gamut of solvability and boundedness results.

5 Other results

5.1 Renorming results: Asplund averaging

Edgar Asplund [3] showed how to exploit convex analysis to provide remarkable results on the existence of equivalent norms with nice properties. Most optimizers are unaware of his lovely idea which we recast in the language of inf-convolution. Our development is a reworking of that in Day [31]. Let us start with two equivalent norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a Banach space X. We consider the quadratic forms $p_0 := \|\cdot\|_1^2/2$ and $q_0 := \|\cdot\|_2^2/2$, and average for $n \geq 0$ by

$$(42) \qquad p_{n+1}(x) := \frac{p_n(x)+q_n(x)}{2} \text{ and } q_{n+1}(x) := \frac{(p_n \Box q_n)(2x)}{2}.$$

Let $C > 0$ be such that $q_0 \leq p_0 \leq (1+C)q_0$. By the construction of p_n and q_n, we have $q_n \leq p_n \leq (1+4^{-n}C)q_n$ ([3, Lemma]) and so the sequences $(p_n)_{n\in\mathbb{N}}$, $(q_n)_{n\in\mathbb{N}}$ converge to a common limit: a convex quadratic function p.

We shall show that the norm $\|\cdot\|_3 := \sqrt{2p}$ typically inherits the good properties of both $\|\cdot\|_1$ and $\|\cdot\|_2$. This is based on the following fairly straightforward result.

Theorem 5.1 (Asplund) (See [3, Theorem 1].) *If either p_0 or q_0 is strictly convex, so is p.*

We make a very simple application in the case that X is reflexive. In [41], Lindenstrauss showed that every reflexive Banach space has an equivalent strictly convex norm. The reader may consult [22, Chapter 4] for more general results. Now take $\|\cdot\|_1$ to be an equivalent strictly convex norm on X, and take $\|\cdot\|_2$ to be an equivalent smooth norm with its dual norm on X^* strictly convex. Theorem 5.1 shows that p is strictly convex. We note that by Corollary 2.25 and Fact 2.21

$$q^*_{n+1}(x^*) := \frac{q^*_n(x^*) + q_n(x^*)}{2} \quad \text{and } p^*_{n+1}(x^*) := \frac{(p^*_n \Box q^*_n)(2x^*)}{2}$$

so that Theorem 5.1 applies to p^*_0 and q^*_0. Hence p^* is strictly convex (see also [30, Proof of Corollary 1, page 111]). Hence $\|\cdot\|_3 (:= \sqrt{2p})$ and its dual norm $(:= \sqrt{2p^*})$ are equivalent strictly convex norms on X and X^* respectively.

Hence $\|\cdot\|_3$ is an equivalent strictly convex and smooth norm (since its dual is strictly convex). The existence of such a norm was one ingredient of Rockafellar's first proof of the Sum theorem.

5.2 Resolvents of maximally monotone operators and connection with convex functions

It is well known since Minty, Rockafellar, and Bertsekas-Eckstein that in Hilbert spaces, monotone operators can be analyzed from the alternative viewpoint of certain nonexpansive (and thus Lipschitz continuous) mappings, more precisely, the so-called resolvents. Given a Hilbert space H and a set-valued operator $A\colon H \rightrightarrows H$, the *resolvent* of A is

$$J_A := (\operatorname{Id} + A)^{-1}.$$

The history of this notion goes back to Minty [43] (in Hilbert spaces) and Brezis, Crandall and Pazy [29] (in Banach spaces). There exist more general notions of resolvents based on different tools, such as the normalized duality mapping, the Bregman distance or other maximally monotone operators (see [40, 1, 11]). For more details on resolvents on Hilbert spaces see [7].

The Minty surjectivity theorem (Theorem 4.3 [43]) implies that a monotone operator is maximally monotone if and only if the resolvent is single-valued with full domain. In fact, a classical result due to Eckstein-Bertsekas [32] says even more. Recall that a mapping $T : H \to H$ is *firmly nonexpansive* if for all $x, y \in H$, $\|Tx - Ty\| \le \langle Tx - ty, x - y\rangle$.

Theorem 5.2 *Let H be a Hilbert space. An operator $A\colon H \rightrightarrows H$ is (maximal) monotone if and only if J_A is firmly nonexpansive (with full domain).*

Example 5.3 Given a closed convex set $C \subseteq H$, the normal cone operator of C, N_C, is a maximally monotone operator whose resolvent can be proved to be the metric projection onto C. Therefore, Theorem 5.2 implies the firm nonexpansivity of the metric projection. $\Diamond$

In the particular case when A is the subdifferential of a possibly non-differentiable convex function in a Hilbert space, whose maximal monotonicity was established by Moreau [48] (in Banach spaces this is due to Rockafellar [56], see also [25, 22]), the resolvent turns into the proximal mapping in the following sense of Moreau. If $f : H \to \,]-\infty, +\infty]$ is a lower semicontinuous convex function defined on a Hilbert space H, the *proximal or proximity* mapping is the operator $\operatorname{prox}_f : H \to H$ defined by

$$\operatorname{prox}_f(x) := \operatorname*{argmin}_{y \in H} \left\{ f(y) + \frac{1}{2}\|x - y\|^2 \right\}.$$

This mapping is well-defined because $\operatorname{prox}_f(x)$ exists and is unique for all $x \in H$. Moreover, there exists the following subdifferential characterization: $u = \operatorname{prox}_f(x)$ if and only if $x - u \in \partial f(u)$.

Moreau's decomposition in terms of the proximal mapping is a powerful nonlinear analysis tool in the Hilbert setting that has been used in various areas of optimization and applied mathematics. Moreau established his decomposition motivated by problems in unilateral mechanics. It can be proved readily by using the conjugate and subdifferential.

Theorem 5.4 (Moreau decomposition) *Given a lower semicontinuous convex function* $f : H \to \,]-\infty, +\infty]$, *for all* $x \in H$,

$$x = \operatorname{prox}_f(x) + \operatorname{prox}_{f^*}(x).$$

Example 5.5 Note that for $f := \iota_C$, with C closed and convex, the proximal mapping turns into the projection onto a closed and convex set C. Therefore, this result generalizes the decomposition by orthogonal projection on subspaces. In particular, if K is a closed convex cone (thus $\iota_K^* = \iota_{K^-}$, see Example 2.18), Moreau's decomposition provides a characterization of the projection onto K:

$$x = y + z \text{ with } y \in K,\ z \in K^- \text{ and } \langle y, z \rangle = 0 \Leftrightarrow y = P_K x \text{ and } z = P_{K^-} x.$$

This illustrates that in Hilbert space, the Moreau decomposition can be thought of as generalizing the decomposition into positive and negative parts of a vector in a normed lattice [22, §6.7] to an arbitrary convex cone. $\Diamond$

There is another notion associated to an operator A, which is strongly related to the resolvent. That is the *Yosida approximation* of index $\lambda > 0$ or the *Yosida λ-regularization*:

$$A_\lambda := (\lambda \operatorname{Id} + A^{-1})^{-1} = \frac{1}{\lambda}(\operatorname{Id} - J_{\lambda A}).$$

If the operator A is maximally monotone, so is the Yosida approximation, and along with the resolvent they provide the so-called *Minty parametrization* of the graph of A that is Lipschitz continuous in both directions [58]:

$$(J_{\lambda A}(z), A_\lambda(z)) = (x, y) \Leftrightarrow z = x + y, (x, y) \in \operatorname{gra} A.$$

If $A = \partial f$ is the subdifferential of a proper lower semicontinuous convex function f, it turns out that the Yosida approximation of A is the gradient of the *Moreau envelope* of f $e_\lambda f$, defined as the infimal convolution of f and $\|\cdot\|^2/2\lambda$, that is,

$$e_\lambda f(x) := f \,\Box\, \frac{\|\cdot\|^2}{2\lambda} = \inf_{y \in H} \left\{ f(y) + \frac{1}{2\lambda}\|x - y\|^2 \right\}.$$

This justifies the alternative term Moreau-Yosida approximation for the mapping $(\partial f)_\lambda = (\lambda\,\mathrm{Id} + (\partial f)^{-1})^{-1}$. This allows to obtain a proof in Hilbert space of the connection between the convexity of the function and the monotonicity of the subdifferential (see [58]): *a proper lower semicontinuous function is convex if and only its* Clarke subdifferential *is monotone.*

It is worth mentioning that generally the role of the Moreau envelope is to approximate the function, with a regularizing effect since it is finite and continuous even though the function may not be so. This behavior has very useful implications in convex and variational analysis.

5.3 Symbolic convex analysis

The thesis work of Hamilton [20] has provided a conceptual and effective framework (the SCAT *Maple* software) for computing conjugates, subdifferentials and infimal convolutions of functions of several variables. Key to this is the notion of *iterated conjugation* (analogous to iterated integration) and a good data structure.

As a first example, with some care, the convex conjugate of the function

$$f : x \mapsto \log\left(\frac{\sinh(3\,x)}{\sinh x}\right)$$

can be symbolically nursed to obtain the result

$$g : y \mapsto \frac{y}{2} \cdot \log\left(\frac{y + \sqrt{16 - 3y^2}}{4 - 2y}\right) + \log\left(\frac{\sqrt{16 - 3y^2} - 2}{6}\right),$$

with domain $[-2, 2]$.

Since the conjugate of g is much more easily computed to be f, this produces a symbolic computational proof that f and g are convex and are mutually conjugate.

Similarly, *Maple* produces the conjugate of $x \mapsto \exp(\exp(x))$ as $y \mapsto y\,(\log(y) - W(y) - 1/W(y))$ in terms of the *Lambert's* W function—the multi-valued inverse of $z \mapsto ze^z$. This function is unknown to most humans but is built into both *Maple* and *Mathematica*. Thus Maple knows that to order five

$$g(y) = -1 + (-1 + \log y)\,y - \frac{1}{2}y^2 + \frac{1}{3}y^3 - \frac{3}{8}y^4 + O\left(y^5\right).$$

Figure 3 shows the *Maple*-computed conjugate after the *SCAT* package is loaded: There is a corresponding numerical program *CCAT* [20]. Current work is adding the capacity to symbolically compute convex compositions—and so in principle Fenchel duality.

5.4 Partial Fractions and Convexity

We consider a network *objective function* p_N given by

$$p_N(q) := \sum_{\sigma \in S_N} \left(\prod_{i=1}^{N} \frac{q_{\sigma(i)}}{\sum_{j=i}^{N} q_{\sigma(j)}} \right) \left(\sum_{i=1}^{N} \frac{1}{\sum_{j=i}^{N} q_{\sigma(j)}} \right),$$

```
> f11:=convert(exp(exp(x)),PWF);
```

$$f11 := \left\{ \mathrm{e}^{\mathrm{e}^x} \quad all(x) \right.$$

```
> g11:=Conj(f11,y);
```

$$g11 := \begin{cases} \infty & y < 0 \\ -1 & y = 0 \\ -\dfrac{y\left(\mathrm{LambertW}(y)^2 - \mathrm{LambertW}(y)\ln(y) + 1\right)}{\mathrm{LambertW}(y)} & 0 < y \end{cases}$$

```
> sdg11:=SubDiff(g11);
```

$$sdg11 := \begin{cases} \{\,\} & y < 0 \\ \{\,\} & y = 0 \\ \{-\mathrm{LambertW}(y) + \ln(y)\} & 0 < y \end{cases}$$

```
> Plot(sdg11,y=-1..1,view=[0..1,-5..0],axes=boxed,labels=["$y$",""]
  );
```

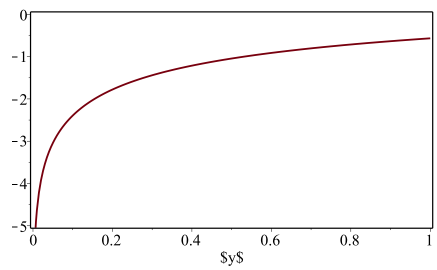

Figure 3: The conjugate and subdifferential of exp exp.

summed over *all* $N!$ permutations; so a typical term is

$$\left(\prod_{i=1}^{N} \frac{q_i}{\sum_{j=i}^{N} q_j}\right)\left(\sum_{i=1}^{N} \frac{1}{\sum_{j=i}^{n} q_j}\right).$$

For example, with $N = 3$ this is

$$q_1 q_2 q_3 \left(\frac{1}{q_1+q_2+q_3}\right)\left(\frac{1}{q_2+q_3}\right)\left(\frac{1}{q_3}\right)\left(\frac{1}{q_1+q_2+q_3} + \frac{1}{q_2+q_3} + \frac{1}{q_3}\right).$$

This arose as the objective function in research into coupon collection. The researcher, Ian Affleck, wished to show p_N was *convex* on the positive orthant.

First, we tried to simplify the expression for p_N. The *partial fraction decomposition* gives:

$$\begin{aligned} p_1(x_1) &= \frac{1}{x_1}, \\ p_2(x_1,x_2) &= \frac{1}{x_1} + \frac{1}{x_2} - \frac{1}{x_1+x_2}, \\ p_3(x_1,x_2,x_3) &= \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} - \frac{1}{x_1+x_2} - \frac{1}{x_2+x_3} - \frac{1}{x_1+x_3} + \frac{1}{x_1+x_2+x_3}. \end{aligned} \tag{43}$$

In [60], the simplified expression of P_N is given by

$$p(x_1, x_2, \cdots, x_N) := \sum_{i=1}^{N} \frac{1}{x_i} - \sum_{1 \le i < j \le N} \frac{1}{x_i + x_j} + \sum_{1 \le i < j < k \le N} \frac{1}{x_i + x_j + x_k}$$
$$- \ldots + (-1)^{N-1} \frac{1}{x_1 + x_2 + \ldots + x_N}.$$

Partial fraction decompositions are another arena in which computer algebra systems are hugely useful. The reader is invited to try performing the third case in (43) by hand. It is tempting to predict the "same" pattern will hold for $N = 4$. This is easy to confirm (by computer if not by hand) and so we are led to:

Conjecture 5.6 *For each $N \in \mathbb{N}$, the function*

$$p_N(x_1, \cdots, x_N) = \int_0^1 \left(1 - \prod_{i=1}^{N} (1 - t^{x_i})\right) \frac{dt}{t} \tag{44}$$

is convex; indeed $1/p_N$ is concave.

One may check symbolically that this is true for $N < 5$ via a large Hessian computation. But this is impractical for larger N. That said, it is easy to numerically sample the Hessian for much larger N, and it is always positive definite. Unfortunately, while the integral is convex, the integrand is not, or we would be done. Nonetheless, the process was already a success, as the researcher was able to rederive his objective function in the form of (44).

A year after, Omar Hjab suggested re-expressing (44) as the *joint expectation* of Poisson distributions.[3] Explicitly, this leads to:

Lemma 5.7 [17, §1.7] *If $x = (x_1, \cdots, x_n)$ is a point in the positive orthant $\mathbb{R}^n_{++}$, then*

$$\int_0^\infty \left(1 - \prod_{i=1}^{n} (1 - e^{-t x_i})\right) dt = \left(\prod_{i=1}^{n} x_i\right) \int_{\mathbb{R}^n_{++}} e^{-\langle x, y \rangle} \max(y_1, \cdots, y_n)\, dy, \tag{45}$$

where $\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n$ is the Euclidean inner product.

It follows from the lemma—which is proven in [17] with no recourse to probability theory—that

$$p_N(x) = \int_{\mathbb{R}^N_{++}} e^{-(y_1 + \cdots + y_N)} \max\left(\frac{y_1}{x_1}, \cdots, \frac{y_N}{x_N}\right) dy,$$

and hence that p_N is positive, decreasing, and convex, as is the integrand. To derive the stronger result that $1/p_N$ is concave we refer to [17, §1.7]. Observe that since $\frac{2ab}{a+b} \le \sqrt{ab} \le (a+b)/2$, it follows from (45) that p_N is log-convex (and convex). A little more analysis of the integrand shows

[3] See "Convex, II" *SIAM Electronic Problems and Solutions* at `http://www.siam.org/journals/problems/downloadfiles/99-5sii.pdf`.

p_N is strictly convex on its domain. The same techniques apply when x_k is replaced in (43) or (44) by $g(x_k)$ for a concave positive function g.

Though much nice related work is to found in [60], there is still no truly direct proof of the convexity of p_N. Surely there should be! This development neatly shows both the power of computer assisted convex analysis and its current limitations.

Lest one think most results on the real line are easy, we challenge the reader to prove the empirical observation that

$$p \mapsto \sqrt{p} \int_0^\infty \left| \frac{\sin x}{x} \right|^p \, dx$$

is *difference convex* on $(1, \infty)$, i.e. it can be written as a difference of two convex functions [5].

6 Concluding comments

All researchers and practitioners in convex analysis and optimization owe a great debt to Jean-Jacques Moreau—whether they know so or not. We are delighted to help make his seminal role more apparent to the current generation of scholars. For those who read French we urge them to experience the pleasure of [44, 45, 46, 48] and especially [49]. For others, we highly recommend [50], which follows [48] and of which Zuhair Nashed wrote in his *Mathematical Review* MR0217617: "There is a great need for papers of this kind; the present paper serves as a model of clarity and motivation."

Acknowledgments The authors are grateful to the three anonymous referees for their pertinent and constructive comments. The authors also thank Dr. Hristo S. Sendov for sending them the manuscript [60]. The authors were all partially supported by various Australian Research Council grants.

References

[1] Y. Alber and D. Butnariu, "Convergence of Bregman projection methods for solving consistent convex feasibility problems in reflexive Banach spaces", *Journal of Optimization Theory and Applications*, vol. 92, pp. 33–61, 1997.

[2] M. Alimohammady and V. Dadashi, "Preserving maximal monotonicity with applications in sum and composition rules", *Optimization Letters*, vol. 7, pp. 511–517, 2013.

[3] E. Asplund, "Averaged norms", *Israel Journal of Mathematics* vol. 5, pp. 227–233, 1967.

[4] H. Attouch, H. Riahi, and M. Thera, "Somme ponctuelle d'operateurs maximaux monotones" [Pointwise sum of maximal monotone operators] Well-posedness and stability of variational problems. *Serdica. Mathematical Journal*, vol. 22, pp. 165–190, 1996.

[5] M. Bačák and J.M. Borwein, "On difference convexity of locally Lipschitz functions", *Optimization*, pp. 961–978, 2011.

[6] H.H. Bauschke, J.M. Borwein, X. Wang, and L. Yao, “Construction of pathological maximally monotone operators on non-reflexive Banach spaces”, *Set-Valued and Variational Analysis*, vol. 20, pp. 387–415, 2012.

[7] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.

[8] H.H. Bauschke and X. Wang, “The kernel average for two convex functions and its applications to the extension and representation of monotone operators”, *Transactions of the American Mathematical Society*, vol. 36, pp. 5947–5965, 2009.

[9] H.H. Bauschke, X. Wang, and L. Yao, “Monotone linear relations: maximality and Fitzpatrick functions”, *Journal of Convex Analysis*, vol. 16, pp. 673–686, 2009.

[10] H.H. Bauschke, X. Wang, and L. Yao, “Autoconjugate representers for linear monotone operators”, *Mathematical Programming (Series B)*, vol. 123, pp. 5-24, 2010.

[11] H.H. Bauschke, X. Wang, and L. Yao, “General resolvents for monotone operators: characterization and extension”, in *Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning and Inverse Problems*, Medical Physics Publishing, pp. 57–74, 2010.

[12] J.M. Borwein, “A generalization of Young’s ℓ^p inequality”, *Mathematical Inequalities & Applications*, vol. 1, pp. 131–136, 1998.

[13] J.M. Borwein, “Maximal monotonicity via convex analysis”, *Journal of Convex Analysis*, vol. 13, pp. 561–586, 2006.

[14] J.M. Borwein, “Maximality of sums of two maximal monotone operators in general Banach space”, *Proceedings of the American Mathematical Society*, vol. 135, pp. 3917–3924, 2007.

[15] J.M. Borwein, “Fifty years of maximal monotonicity”, *Optimization Letters*, vol. 4, pp. 473–490, 2010.

[16] J.M. Borwein and D.H. Bailey, *Mathematics by Experiment: Plausible Reasoning in the 21st Century*, A.K. Peters Ltd, Second expanded edition, 2008.

[17] J.M. Borwein, D.H. Bailey and R. Girgensohn, *Experimentation in Mathematics: Computational Paths to Discovery*, A.K. Peters Ltd, 2004. ISBN: 1-56881-211-6.

[18] J.M. Borwein and S. Fitzpatrick, “Local boundedness of monotone operators under minimal hypotheses”, *Bulletin of the Australian Mathematical Society*, vol. 39, pp. 439–441, 1989.

[19] J.M. Borwein, R.S Burachik, and L. Yao, “Conditions for zero duality gap in convex programming”, *Journal of Nonlinear and Convex Analysis*, in press; `http://arxiv.org/abs/1211.4953v2`.

[20] J.M. Borwein and C. Hamilton, “Symbolic Convex Analysis: Algorithms and Examples,” *Mathematical Programming*, **116** (2009), 17–35. *Maple* packages *SCAT* and *CCAT* available at `http://carma.newcastle.edu.au/ConvexFunctions/SCAT.ZIP`.

[21] J.M. Borwein and A.S. Lewis, *Convex Analyis andd Nonsmooth Optimization*, Second expanded edition, Springer, 2005.

[22] J.M. Borwein and J.D. Vanderwerff, *Convex Functions*, Cambridge University Press, 2010.

[23] J.M. Borwein and L. Yao, "Structure theory for maximally monotone operators with points of continuity", *Journal of Optimization Theory and Applications*, vol 157, pp. 1–24, 2013 (Invited paper).

[24] J.M. Borwein and L. Yao, "Recent progress on Monotone Operator Theory", *Infinite Products of Operators and Their Applications*, Contemporary Mathematics, in press; `http://arxiv.org/abs/1210.3401v2`.

[25] J.M. Borwein and Q.J. Zhu, *Techniques of variational analysis*, CMS Books in Mathematics/Ouvrages de Mathmatiques de la SMC, 20. Springer-Verlag, New York, 2005.

[26] R.I. Boţ S. Grad, and G. Wanka, *Duality in Vector Optimization*, Springer, 2009.

[27] R.I. Boţ and G. Wanka, "A weaker regularity condition for subdifferential calculus and Fenchel duality in infinite dimensional spaces", *Nonlinear Analysis*, vol. 64, pp. 2787–2804, 2006.

[28] R.S. Burachik and A.N. Iusem, *Set-Valued Mappings and Enlargements of Monotone Operators*, Springer, vol. 8, 2008.

[29] H. Brezis, G. Crandall and P. Pazy, Perturbations of nonlinear maximal monotone sets in Banach spaces, *Communications on Pure and Applied Mathematics*, vol. 23, pp. 123–144, 1970.

[30] J. Diestel, *Geometry of Banach spaces*, Springer-Verlag, 1975

[31] M.M. Day, *Normed linear spaces*, Third edition, Springer-Verlag, New York-Heidelberg, 1973.

[32] J. Eckstein and D.P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators", *Mathematical Programming*, vol. 55, pp. 293–318, 1992.

[33] S. Fitzpatrick, "Representing monotone operators by convex functions", in *Workshop/Miniconference on Functional Analysis and Optimization (Canberra 1988)*, Proceedings of the Centre for Mathematical Analysis, Australian National University, vol. 20, Canberra, Australia, pp. 59–65, 1988.

[34] N. Ghoussoub, *Self-dual partial differential systems and their variational principles. Springer Monographs in Mathematics*, Springer, 2009.

[35] J.-P. Gossez, "On the range of a coercive maximal monotone operator in a nonreflexive Banach space", *Proceedings of the American Mathematical Society*, vol. 35, pp. 88–92, 1972.

[36] J.-B. Hiriart-Urruty, M. Moussaoui, A. Seeger, and M. Volle, "Subdifferential calculus without qualification conditions, using approximate subdifferentials: a survey", *Nonlinear Analysis*, vol. 24, pp. 1727–1754, 1995.

[37] J.-B. Hiriart-Urruty and R. Phelps, "Subdifferential Calculus Using ε-Subdifferentials", *Journal of Functional Analysis* vol. 118, pp. 154–166, 1993.

[38] L. Hörmander, "Sur la fonction d'appui des ensembles convexes dans un espace localement convexe", *Arkiv för Matematik*, vol. 3, pp. 181–186, 1955.

[39] V. Klee, "Convexity of Chebysev sets", *Mathematische Annalen*, vol. 142, pp. 292–304, 1961.

[40] F. Kohsaka and W. Takahashi, "Existence and approximation of fixed points of firmly nonexpansivetype mappings in Banach spaces", *SIAM Journal on Optimization*, vol. 19, pp. 824–835, 2008.

[41] J. Lindenstrauss, "On nonseparable reflexive Banach spaces", *Bulletin of the American Mathematical Society*, vol. 72, pp. 967–970, 1966.

[42] P. Maréchal, "A convexity theorem for multiplicative functions", *Optimization Letters*, vol. 6, pp. 357–362, 2012.

[43] G. Minty, "Monotone (nonlinear) operators in a Hilbert space", *Duke Mathematical Journal*, vol. 29, pp. 341–346, 1962.

[44] J.J. Moreau, "Fonctions convexes en dualité", Faculté des Sciences de Montpellier, Séminaires de Mathématiques Université de Montpellier, Montpellier, 1962.

[45] J.J. Moreau, "Fonctions à valeurs dans $[-\infty, +\infty]$; notions algébriques", *Faculté des Sciences de Montpellier, Séminaires de Mathématiques*, Université de Montpellier, Montpellier, 1963.

[46] J.J. Moreau, "Étude locale d'une fonctionnelle convexe", Faculté des Sciences de Montpellier, Séminaires de Mathématiques Université de Montpellier, Montpellier, 1963.

[47] J.J. Moreau, "Sur la function polaire d'une fonctionelle semi-continue supérieurement", *Comptes Rendus de l'Académie des Sciences*, vol. 258, pp. 1128–1130, 1964.

[48] J.J. Moreau, "Proximité et dualité dans un espace hilbertien", *Bulletin de la Société Mathématique de France*, vol. 93, pp. 273–299, 1965.

[49] J.J. Moreau, *Fonctionnelles convexes*, Séminaire Jean Leray, College de France, Paris, pp. 1–108, 1966–1967. Available at `http://carma.newcastle.edu.au/ConvexFunctions/moreau66-67.pdf`.

[50] J.J. Moreau, "Convexity and duality", pp. 145–169 in *Functional Analysis and Optimization*, Academic Press, New York, 1966.

[51] J.-P. Penot, "The relevance of convex analysis for the study of monotonicity", *Nonlinear Analysis*, vol. 58, pp. 855–871, 2004.

[52] J.-P. Penot and C. Zălinescu, "Some problems about the representation of monotone operators by convex functions", *The Australian New Zealand Industrial and Applied Mathematics Journal*, vol. 47, pp. 1–20, 2005.

[53] R.R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, 2nd Edition, Springer-Verlag, 1993.

[54] R.R. Phelps and S. Simons, "Unbounded linear monotone operators on nonreflexive Banach spaces", *Journal of Nonlinear and Convex Analysis*, vol. 5, pp. 303–328, 1998.

[55] R.T. Rockafellar, "Extension of Fenchel's duality theorem for convex functions", *Duke Mathematical Journal*, vol. 33, pp. 81–89, 1966.

[56] R.T. Rockafellar, "On the maximal monotonicity of subdifferential mappings", *Pacific Journal of Mathematics*, vol. 33, pp. 209–216, 1970.

[57] R.T. Rockafellar, "On the maximality of sums of nonlinear monotone operators", *Transactions of the American Mathematical Society*, vol. 149, pp. 75–88, 1970.

[58] R.T. Rockafellar and R.J-B Wets, Variational analysis. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 317. Springer-Verlag, Berlin, 1998 (3rd Printing, 2009).

[59] R. Rudin, *Functional Analysis*, Second Edition, McGraw-Hill, 1991.

[60] H.S. Sendov and R. Zitikis, "The shape of the Borwein-Affleck-Girgensohn function generated by completely monotone and Bernstein functions", *Journal of Optimization Theory and Applications*, in press.

[61] S. Simons, *Minimax and Monotonicity*, Springer-Verlag, 1998.

[62] S. Simons, *From Hahn-Banach to Monotonicity*, Springer-Verlag, 2008.

[63] S. Simons and C. Zălinescu, "A new proof for Rockafellar's characterization of maximal monotone operators", *Proceedings of the American Mathematical Society*, vol. 132, pp. 2969–2972, 2004.

[64] S. Simons and C. Zălinescu, "Fenchel duality, Fitzpatrick functions and maximal monotonicity", *Journal of Nonlinear and Convex Analysis*, vol. 6, pp. 1–22, 2005.

[65] L. Sucheston, "Banach limits", *American Mathematical Monthly*, vol. 74, pp. 308–311, 1967.

[66] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing, 2002.

[67] E. Zeidler, *Nonlinear Functional Analysis and its Applications II/A: Linear Monotone Operators*, Springer-Verlag, 1990.

[68] E. Zeidler, *Nonlinear Functional Analysis and its Applications II/B: Nonlinear Monotone Operators*, Springer-Verlag, 1990.