Duality and Convex Programming

Jonathan M. Borwein
School of Mathematical and Physical Sciences, University of Newcastle, Australia
Jonathan.Borwein@newcastle.edu.au

D. Russell Luke*
Institut für Numerische und Angewandte Mathematik, Universität Göttingen
r.luke@math.uni-goettingen.de

April 21, 2010

Key words: convex analysis, variational analysis, duality.
AMS 2010 Subject Classification: 49M29, 49M20, 65K10, 90C25, 90C46

Abstract

This chapter surveys key concepts in convex duality theory and their application to the analysis and numerical solution of problem archetypes in imaging.

1 Introduction

An image is worth a thousand words, the joke goes, but takes up a million times the memory. All the more reason to be efficient when computing with images. Whether one is determining a "best" image according to some criteria, or applying a "fast" iterative algorithm for processing images, the theory of optimization and variational analysis lies at the heart of achieving these goals. Success or failure hinges on the abstract structure of the task at hand. For many practitioners, these details are secondary: if the picture looks good, it *is* good. For the specialist in variational analysis and optimization, however, it is what is went into constructing the image that matters: if it is *convex*, it is good.

This chapter surveys more than a half-a-century of work in convex analysis that has played a fundamental role in the development of computational imaging, and to bring to light as many of the contributors to this field as possible. There is no shortage of good books on convex and variational analysis; we point interested readers to the modern references [3, 5, 15, 16, 19, 31, 42, 48, 49, 57, 64, 69, 73, 76, 77, 80, 81, 87]. References focused more on imaging and signal processing,

*This author's work was supported in part by NSF grant DMS-0712796.

but with a decidedly variational flavor, include [2, 24, 79]. For general references on numerical optimization, see [11, 20, 26, 58, 68, 85].

For many years the dominant distinction in applied mathematics between problem types has rested upon linearity, or lack thereof. This categorization still holds sway today with nonlinear problems largely regarded as "hard" while linear problems are generally considered "easy". But since the advent of the *interior point revolution* [67], at least in linear optimization, it is more or less agreed that *nonconvexity*, not nonlinearity, more accurately delineates hard from easy. The goal of this chapter is to make this case more broadly. Indeed, for convex sets topological, algebraic and geometric notions often coincide, and so the tools of convex analysis provide not only for a tremendous synthesis of ideas, but for key insights whose dividends are efficient algorithms for solving large (*infinite dimensional*) problems, and indeed even large *nonlinear* problems.

We consider different instances of a single optimization model. This model accounts for the vast majority of variational problems appearing in imaging science:

$$\begin{array}{ll} \underset{x\in C\subset X}{\text{minimize}} & I_\varphi(x) \\ \text{subject to} & Ax \in D. \end{array} \tag{1.1}$$

Here X and Y are real Banach spaces with continuous duals X^* and Y^*, C and D are closed and convex, $A : X \to Y$ is a continuous linear operator, and the integral functional $I_\varphi(x) := \int_T \varphi(x(t))\mu(dt)$ is defined on some vector subspace $L_p(T,\mu)$ of X for μ a complete totally finite measure on some measure space T. The integral operator I_φ is an *entropy* with integrand $\varphi : \mathbb{R} \to]-\infty, +\infty]$ a closed convex function. This provides an extremely flexible framework that specializes to most of the instances of interest, and is general enough to extend results to non-Hilbert space settings. The most common examples are

$$\text{Burg entropy:} \quad \varphi(x) := -\ln(x) \tag{1.2}$$

$$\text{Shannon-Boltzmann entropy:} \quad \varphi(x) := x\ln(x) \tag{1.3}$$

$$\text{Fermi-Dirac entropy :} \quad \varphi(x) := x\ln(x) + (1-x)\ln(1-x) \tag{1.4}$$

$$L_p\text{-norm} \quad \varphi(x) := \frac{\|x\|^p}{p} \tag{1.5}$$

$$L_p \text{ entropy} \quad \varphi(x) := \begin{cases} \frac{x^p}{p} & x \geq 0 \\ +\infty & \text{else} \end{cases} \tag{1.6}$$

$$\text{total variation} \quad \varphi(x) := |\nabla x|. \tag{1.7}$$

See [10, 13, 14, 18, 22, 27, 28, 37, 44, 82] for these and other entropies.

There is a rich correspondence between the algorithmic approach to applications implicit in the variational formulation (1.1) and the prevalent *feasibility* approach to problems. Here one considers the problem of finding the point x that lies in the intersection of the constraint sets:

$$\text{find } x \in C \cap S \quad \text{where} \quad S := \{x \in X \mid Ax \in D\}.$$

In the case where the intersection is quite large, one might wish to find the point in the intersection in some sense closest to a reference point x_0 (frequently the origin). It is the job of the objective in (1.1) to pick the element of $C \cap S$ that has the desired properties, that is, to pick the *best approximation*. The feasibility formulation suggests very naturally projection algorithms for finding the intersection whereby one applies the constraints one at a time in some fashion, e.g. cyclically, or at random [4,26,33,86]. This is quite a powerful framework as it provides a great deal of flexibility and is amenable to parallelization for large-scale problems. Many of the algorithms for feasibility problems have counterparts for the more general best approximation problems [5, 8, 39, 60]. For studies of these algorithms in nonconvex settings see [6, 7, 23, 35, 55, 56, 59–61]. The projection algorithms that are central to convex feasibility and best approximation problems play a key role in algorithms for solving the problems we will consider here.

Before detailing specific applications we state a general duality result for problem (1.1) that motivates many of the tools we use. One of the more central tools we make use of is the *Fenchel conjugate* [43] of a mapping $f : X \to [-\infty, +\infty]$, denoted $f^* : X^* \to [-\infty, +\infty]$ and defined by

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

The conjugate is always convex (as a supremum of affine functions) while $f = f^{**}$ exactly if f is convex, proper (not everywhere infinite) and lower semi-continuous (lsc) [19, 42]. Here and below, unless otherwise specified X is a normed space with dual X^*. The following theorem uses constraint qualifications involving the concept of the *core* of a set, the *effective domain* of a function (dom f), and the points of continuity of a function (cont f).

Definition 1.1 (core). *The core of a set $F \subset X$ is defined by $x \in \operatorname{core} F$ if for each $h \in \{x \in X \mid \|x\| = 1\}$, there exists $\delta > 0$ so that $x + th \in F$ for all $0 \le t \le \delta$.*

It is clear from the definition that, $\operatorname{int} F \subset \operatorname{core} F$. If F is a convex subset of a Euclidean space, or if F is closed, then the core and the interior are *identical* [15, Theorem 4.1.4].

Theorem 1.2 (Fenchel duality - Theorems 2.3.4 and 4.4.18 of [19])**.** *Let X and Y be Banach spaces, let $f : X \to (-\infty, +\infty]$ and $g : Y \to (-\infty, +\infty]$ and let $A : X \to Y$ be a bounded linear map. Define the primal and dual values p, $d \in [-\infty, +\infty]$ by the Fenchel problems*

$$\begin{aligned} p &= \inf_{x \in X}\{f(x) + g(Ax)\} \\ d &= \sup_{y^* \in Y^*}\{-f^*(A^*y^*) - g^*(-y^*)\}. \end{aligned} \tag{1.8}$$

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

If f, g are convex and satisfy either

$$0 \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} f) \quad \text{with } f \text{ and } g \text{ lsc}, \tag{1.9}$$

or

$$A \operatorname{dom} f \cap \operatorname{cont} g \neq \emptyset, \tag{1.10}$$

then $p = d$, and the supremum to the dual problem is attained if finite.

Applying Theorem 1.2 to problem (1.1) we have $f(x) = I_\varphi(x) + \iota_C(x)$ and $g(y) = \iota_D(y)$ where ι_F is the *indicator function* of the set F:

$$\iota_F(x) := \begin{cases} 0 & \text{if } x \in F \\ +\infty & \text{else.} \end{cases} \tag{1.11}$$

The tools of convex analysis and the phenomenon of duality are central to formulating, analyzing and solving application problems. Already apparent from the general application above is the necessity for a calculus of Fenchel conjugation in order to compute the conjugate of sums of functions. In some specific cases, one can arrive at the same conclusion with less theoretical overhead, but this is at the cost of missing out on more general structures that are not necessarily automatic in other settings.

Duality has a long-established place in economics where primal and dual problems have direct interpretations in the context of the theory of zero-sum games, or where Lagrange multipliers and dual variables are understood, for instance, as shadow prices. In imaging there is not as often an easy interplay between the physical interpretation of primal and dual problems. Duality has been used towards a variety of ends in contemporary image and signal processing, the majority of them, however, having to do with algorithms [9,17,18,27–29,34,36,38,47,50,62,84]. Nevertheless, the dual perspective yields new statistical or information theoretic insight into image processing problems, in addition to faster algorithms. Five main applications illustrate the variational analytical approach to problem solving: linear inverse problems with convex constraints, compressive imaging, image denoising and deconvolution, nonlinear inverse scattering, and finally Fredholm integral equations. We briefly review these applications below. In subsequent sections we develop the tools for their analysis. At the end of the chapter we revisit these applications in light of the convex analytical tools collected along the way.

1.1 Linear inverse problems with convex constraints

Let X be a Hilbert space and $\varphi(x) := \frac{1}{2}\|x\|^2$. The integral functional I_φ is the usual L_2 norm and the solution is the closest feasible point to the origin:

$$\begin{array}{ll} \underset{x \in C \subset X}{\text{minimize}} & \frac{1}{2}\|x\|^2 \\ \text{subject to} & Ax = b. \end{array} \tag{1.12}$$

Levi, for instance, used this variational formulation to determine the minimum energy band-limited signal that matched N measurements $b \in \mathbb{R}^n$ with the model $A : X \to \mathbb{R}^n$ [54]. Note that the signal space is infinite dimensional while the measurement space is finite dimensional, a common situation in practice. Potter and Arun [70] recognized a much broader applicability of this variational formulation to remote sensing and medical imaging, and applied duality theory to characterize solutions to (1.12) by $\overline{x} = P_C A^*(\overline{y})$, where $\overline{y} \in Y$ satisfies $b = AP_C A^* \overline{y}$ [70, Theorem 1]. Particularly attractive is the feature that when Y is finite dimensional, these formulas yield a finite dimensional approach to an infinite dimensional problem . The numerical algorithm suggested by Potter and Arun is an iterative procedure in the dual variables:

$$y_{j+1} = y_j + \gamma(b - AP_C A^* y_j) \quad j = 0, 1, 2, \ldots \tag{1.13}$$

The optimality condition and numerical algorithms are explored at the end of this chapter.

As satisfying as this theory is, there is a crucial assumption in the theorem of Potter and Arun about the existence of $\overline{y} \in Y$ such that $b = AP_C A^* \overline{y}$; one need only consider linear least squares for an example where the primal problem is well-posed but no such $\overline{y}$ exists [12]. To facilitate the argument we specialize Theorem 1.2 to the case of linear constraints. The next corollary is a specialization of Theorem 1.2 where g is the indicator function of the point b in the linear constraint.

Corollary 1.3 (Fenchel duality for linear constraints). *Given any $f : X \to (-\infty, \infty]$, any bounded linear map $A : X \to Y$, and any element $b \in Y$, the following weak duality inequality holds:*

$$\inf_{x \in X} \{f(x) \mid Ax = b\} \geq \sup_{y^* \in Y^*} \{\langle b, \ y^* \rangle - f^*(A^* y^*)\}.$$

If f is lsc and convex and $b \in \operatorname{core}(A \operatorname{dom} f)$, then equality holds and the supremum is attained if finite.

Suppose that $C = X$, a Hilbert space and $A : X \to X$. The Fenchel dual to (1.12) is

$$\underset{y \in X}{\text{maximize}} \ \langle y, \ b \rangle - \frac{1}{2} \|A^* y\|^2. \tag{1.14}$$

(The L_2 norm is self-dual.) Suppose that the primal problem (1.12) is *feasible*, that is, $b \in \text{range}(A)$. The objective in (1.14) is convex and differentiable, so elementary calculus (Fermat's rule) yields the optimal solution $\overline{y}$ with $AA^* \overline{y} = b$, assuming $\overline{y}$ exists. If the range of A is strictly larger than that of AA^*, however, it is possible to have $b \in \text{range}(A)$ but $b \notin \text{range}(AA^*)$, in which case the optimal solution $\overline{x}$ to (1.12) is *not* equal to $A^* \overline{y}$, since $\overline{y}$ is not attained. For a concrete example see [12, Example 2.1].

1.2 Imaging with missing data

Let $X = \mathbb{R}^n$, and $\varphi(x) := \|x\|_p$ for $p = 0$ or $p = 1$. The case $p = 1$ is the ℓ_1 norm, and by $\|x\|_0$ we mean the function

$$\|x\|_0 := \sum_j |\operatorname{sign}(x_j)|$$

where $\operatorname{sign}(0) := 0$. One then has the optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & \|x\|_p \\ \text{subject to} & Ax = b. \end{array} \tag{1.15}$$

This model has received a great deal of attention recently in applications where the number of measurements is much smaller than the dimension of the signal space, that is $b \in \mathbb{R}^m$ for $m \ll n$. This problem is well-known in statistics as the missing data problem.

For ℓ_1 optimization ($p = 1$), the seminal work of Candés and Tao establishes probabilistic criteria for when the solution to (1.15) is unique and exactly matches the true signal x_* [25].

Sparsity of the original signal x_* and the algebraic structure of the matrix A are key requirements. Convex analysis easily yields a geometric interpretation of these facts. We develop the tools to show that the dual to this problem is the linear program

$$\begin{array}{ll} \underset{y\in\mathbb{R}^m}{\text{maximize}} & b^T y \\ \text{subject to} & (A^*y)_j \in [-1,1] \quad j=1,2,\ldots,n. \end{array} \tag{1.16}$$

Elementary facts from linear programming guarantee that the solution includes a vertex of the polyhedron described by the constraints, and hence, assuming A is full rank, there can be at most m active constraints. The number of active constraints in the dual problem provides an upper bound on the number of nonzero elements in the primal variable – the signal to be recovered. Unless the number of nonzero elements of x_* is less than the number of measurements m, there is no hope of uniquely recovering x_*. The uniqueness of solutions to the primal problem is easily understood in terms of the geometry of the dual problem, that is, whether or not solutions to the dual problem reside along the edges or faces of the polyhedron. More refined details about *how* sparse x_* need be in order to have a reasonable hope of exact recovery requires more work, but elementary convex analysis already provides the essential intuition.

For the function $\|x\|_0$ ($p = 0$ in (1.15)) the equivalence of the primal and dual problems is lost due to the nonconvexity of the objective. The theory of Fenchel duality still yields *weak duality*, but this is of limited use in this instance. The Fenchel dual to (1.15) is

$$\begin{array}{ll} \underset{y\in\mathbb{R}^m}{\text{maximize}} & b^T y \\ \text{subject to} & (A^*y)_j = 0 \quad j=1,2,\ldots,n. \end{array} \tag{1.17}$$

If we denote the *values* of the primal (1.15) and dual problems (1.17) by p and d respectively, then these values satisfy the *weak duality inequality* $p \geq d$. The primal problem is a combinatorial optimization problem, and hence NP-hard; the dual problem, however, is a linear program, which is finitely terminating. Relatively elementary variational analysis provides a lower bound on the sparsity of signals x that satisfy the measurements. In this instance, however, the lower bound only reconfirms what we already know. Indeed, if A is full rank, then the only solution to the dual problem is $y = 0$. In other words, the minimal sparsity of the solution to the primal problem is zero, which is obvious. The loss of information in passing from primal to dual formulations of nonconvex problems is a common phenomenon and underscores the importance of convexity.

The Fenchel conjugates of the ℓ_1 norm and the function $\|\cdot\|_0$ are given respectively by

$$\varphi_1^*(y) := \begin{cases} 0 & \|y\|_\infty \leq 1 \\ +\infty & \text{else} \end{cases} \qquad (\varphi_1(x) := \|x\|_1) \tag{1.18}$$

$$\varphi_0^*(y) := \begin{cases} 0 & y = 0 \\ +\infty & \text{else} \end{cases} \qquad (\varphi_0(x) := \|x\|_0) \tag{1.19}$$

It is not uncommon to consider the function $\|\cdot\|_0$ as the limit of $\left(\sum_j |x_j|^p\right)^{1/p}$ as $p \to 0$. We present an alternative approach based on regularization of the conjugates. for L and $\epsilon > 0$ define

$$\varphi_{\epsilon,L}(y) := \begin{cases} \epsilon\left(\frac{(L+y)\ln(L+y)+(L-y)\ln(L-y)}{2L\ln(2)} - \frac{\ln(L)}{\ln(2)}\right) & (y \in [-L, L]) \\ +\infty & \text{for } |y| > L. \end{cases} \tag{1.20}$$

This is a scaled and shifted Fermi-Dirac entropy (1.4). It is also a smooth convex function on the interior of its domain and so elementary calculus can be used to calculate the Fenchel conjugate,

$$\varphi^*_{\epsilon,L}(x) = \frac{\epsilon}{\ln(2)} \ln\left(4^{xL/\epsilon} + 1\right) - xL - \epsilon. \tag{1.21}$$

For $L > 0$ fixed, in the limit as $\epsilon \to 0$ we have

$$\lim_{\epsilon\to 0} \varphi_{\epsilon,L}(y) = \begin{cases} 0 & y \in [-L, L] \\ +\infty & \text{else} \end{cases} \quad \text{and} \quad \lim_{\epsilon\to 0} \varphi^*_{\epsilon,L}(x) = L|x|.$$

For $\epsilon > 0$ fixed we have

$$\lim_{L\to 0} \varphi_{\epsilon,L}(x) = \begin{cases} 0 & y = 0 \\ +\infty & \text{else} \end{cases} \quad \text{and} \quad \lim_{L\to 0} \varphi^*_{\epsilon,L}(x) := 0.$$

Note that $\|\cdot\|_0$ and $\varphi^*_{\epsilon,0} := 0$ have the same conjugate, but unlike $\|\cdot\|_0$ the biconjugate of $\varphi^*_{\epsilon,0}$ is itself. Also note that $\varphi_{\epsilon,L}$ and $\varphi^*_{\epsilon,L}$ are convex and smooth on the interior of their domains for all $\epsilon, L > 0$. This is in contrast to metrics of the form $\left(\sum_j |x_j|^p\right)$ which are nonconvex for $p < 1$. We therefore propose solving

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & I_{\varphi^*_{\epsilon,L}}(x) \\ \text{subject to} & Ax = b \end{array} \tag{1.22}$$

as a smooth convex relaxation of the conventional ℓ_p optimization for $0 \le p \le 1$.

1.3 Image Denoising and Deconvolution

We consider next problems of the form

$$\underset{x\in X}{\text{minimize}}\; I_\varphi(x) + \frac{1}{2\lambda}\|Ax - y\|^2 \tag{1.23}$$

where X is a Hilbert space, $I_\varphi : X \to (-\infty, +\infty]$ is a semi-norm on X, and $A : X \to Y$, is a bounded linear operator. This problem is explored in [9] as a general framework that includes total variation minimization [78], wavelet shrinkage [40] and basis pursuit [30]. When A is the identity, problem (1.23) amounts to a technique for denoising; here y is the received, noisy signal, and the solution $\overline{x}$ is an approximation with the desired statistical properties promoted by the objective I_φ. When the linear mapping A is not the identity (for instance, A models convolution against the point spread function of an optical system) problem (1.23) is a variational formulation of *deconvolution*, that is, recovering the true signal from the image y. The focus here is on total variation minimization.

Total variation was first introduced by Rudin, Osher and Fatemi [78] as a regularization technique for denoising images while preserving edges and, more precisely, the statistics of the noisy image. The *total variation* of an image $x \in X = L_1(T)$ – for T and open subset of $\mathbb{R}^2$ – is defined by

$$I_{TV}(x) := \sup\left\{ \int_T x(t)\,\mathrm{div}\,\xi(t)dt \mid \xi \in C_c^1(T,\mathbb{R}^2), |\xi(t)| \le 1\ \forall t \in T \right\}.$$

The integral functional I_{TV} is finite if and only if the distributional derivative Dx of x is a finite Radon measure in T, in which case we have $I_{TV}(x) = |Dx|(T)$. If, moreover, x has a gradient $\nabla x \in L_1(T,\mathbb{R}^2)$, then $I_{TV}(x) = \int |\nabla x(t)|dt$, or, in the context of the general framework established at the beginning of this chapter, $I_{TV}(x) = I_\varphi(x)$ where $\varphi(x(t)) := |\nabla x(t)|$. The goal of the original *total variation denoising problem* proposed in [78] is then to

$$\begin{array}{lll} \underset{x\in X}{\text{minimize}} & I_{TV}(x) & \\ \text{subject to} & \int_T Ax = \int_T x_0 \quad \text{and} \quad \int_T |Ax - x_0|^2 = \sigma^2. \end{array} \tag{1.24}$$

The first constraint corresponds to the assumption that the noise has zero mean and the second assumption requires the denoised image to have a predetermined standard deviation σ. Under reasonable assumptions [28], this problem is equivalent to the convex optimization problem

$$\begin{array}{ll} \underset{x\in X}{\text{minimize}} & I_{TV}(x) \\ \text{subject to} & \|Ax - x_0\|^2 \le \sigma^2. \end{array} \tag{1.25}$$

Several authors have exploited duality in total variation minimization for efficient algorithms to solve the above problem [27,29,38,47]. One can "compute" the Fenchel conjugate of I_{TV} indirectly by using the already mentioned property that the *biconjugate* of a proper, convex lsc function is the function itself: $f^{**}(x) = f(x)$ if (and only if) f is proper, convex and lsc at x. Rewriting I_{TV} as the Fenchel conjugate of some function, we have

$$I_{TV}(x) = \sup_v \langle x,\ v\rangle - \iota_K(v)$$

where

$$K := \overline{\{\mathrm{div}\,\xi \mid \xi \in C_c^1(T,\mathbb{R}^2) \quad \text{and} \quad |\xi(t)| \le 1\ \forall t \in T\}}.$$

From this it is then clear that the Fenchel conjugate of I_{TV} is the indicator function of the convex set K, ι_K.

In [27] duality is used to develop an algorithm, with proof of convergence, for the problem

$$\underset{x\in X}{\text{minimize}}\ I_{TV}(x) + \frac{1}{2\lambda}\|x - x_0\|^2 \tag{1.26}$$

with X a Hilbert space. First-order optimality conditions for this unconstrained problem are

$$0 \in x - x_0 + \lambda \partial I_{TV}(x) \tag{1.27}$$

where $\partial I_{TV}(x)$ is the *subdifferential* of I_{TV} at x defined by

$$v \in \partial I_{TV}(x) \iff I_{TV}(y) \geq I_{TV}(x) + \langle v,\ y - x\rangle \quad \forall y.$$

The optimality condition (1.27) is equivalent to [19, Prop. 4.4.5]

$$x \in \partial I_{TV}^*\left((x_0 - x)/\lambda\right) \tag{1.28}$$

or, since $I_{TV}^* = \iota_K$,

$$\frac{x_0}{\lambda} \in \left(I + \frac{1}{\lambda}\partial\iota_K\right)(z)$$

where $z = (x_0 - x)/\lambda$. (For the finite dimensional statement see [48, Prop. I.6.1.2].) Since K is convex, standard facts from convex analysis determine that $\partial\iota_K(z)$ is the *normal cone mapping* to K at z, denoted $N_K(z)$ and defined by

$$N_K(z) := \begin{cases} \{v \in X \mid \langle v,\ x - z\rangle \leq 0 \quad \text{for all } x \in K\} & z \in K \\ \emptyset & z \notin K. \end{cases}$$

Note that this is a set-valued mapping. The *resolvent* $\left(I + \frac{1}{\lambda}\partial\iota_K\right)^{-1}$ evaluated at x_0/λ is the orthogonal *projection* of x_0/λ onto K. That is, the solution to (1.26) is

$$x_* = x_0 - P_K(x_0/\lambda) = x_0 - P_{\lambda K}(x_0).$$

The inclusions disappear from the formulation due to convexity of K: the resolvent of the normal cone mapping of a convex set is single-valued. The numerical algorithm for solving (1.26) then amounts to an algorithm for computing the projection $P_{\lambda K}$. We develop below the tools from convex analysis used in this derivation.

1.4 Inverse scattering

An important problem in applications involving scattering is the determination of the shape and location of scatterers from measurements of the scattered field at a distance. Modern techniques for solving this problem use *indicator functions* to detect the inconsistency or insolubility of an Fredholm integral equation of the first kind parameterized by points in space. The shape and location of the object is determined by those points where the auxiliary problem is solvable. Equivalently, the technique determines the shape and location of the scatterer by determining whether a sampling function, parameterized by points in space, is in the range of a compact linear operator constructed from the scattering data.

These methods have enjoyed great practical success since their introduction in the later half of the 1990's. Recently Kirsch and Grinberg [51] established a variational interpretation of these ideas. They observe that the range of a linear operator $G:\ X \to Y$ (X and Y are reflexive Banach spaces) can be characterized by the infimum of the mapping $h(\psi):\ Y^* \to \mathbb{R} \cup \{-\infty, +\infty\} := |\langle\psi, F\psi\rangle|$, where $F := GSG^*$ for $S:\ X^* \to X$ a coercive bounded linear operator. Specifically, they establish

Theorem 1.4 (Thoerem 1.16 of [51]). *Let X, Y be reflexive Banach spaces with duals X^* and Y^*. Let $F : Y^* \to Y$ and $G : X \to Y$ be bounded linear operators with $F = GSG^*$ for $S : X^* \to X$ a bounded linear operator satisfying the coercivity condition*

$$|\langle \varphi,\ S\varphi \rangle| \geq c\|\varphi\|_{X^*}^2 \quad \text{for some } c > 0 \text{ and all } \varphi \in \text{range}(G^*) \subset X^*.$$

Then for any $\phi \in Y \setminus \{0\}$ $\phi \in \text{range}(G)$ if and only if

$$\inf\{h(\psi) \mid \psi \in Y^*, \langle \phi,\ \psi \rangle = 1\} > 0.$$

It is shown below that the infimal characterization above is equivalent to the computation of the effective domain of the Fenchel conjugate of h,

$$h^*(\phi) := \sup_{\psi \in Y^*} \{\langle \phi,\ \psi \rangle - h(\psi)\}. \tag{1.29}$$

In the case of scattering, the operator F above is an integral operator whose kernel is made up of the "measured" field on a surface surrounding the scatterer. When the measurement surface is a sphere at infinity, the corresponding operator is known as the *far field operator*. The factor G maps the boundary condition of the governing PDE (the Helmholtz equation) to the *far field pattern*, that is, the kernel of the far field operator. Given the right choice of spaces, the mapping G is compact, one-to-one and dense. There are two keys to using the above facts for determining the shape and location of scatterers: first, the construction of the test function ϕ and, second, the connection of the range of G to that of some operator easily computed from the far field operator F. The secret behind the success of these methods in inverse scattering is, first, that the construction of ϕ is trivial and, second, that there is (usually) a simpler object to work with than the infimum in Theorem 1.4 that depends only on the far field operator (usually the only thing that is known). Indeed, the test functions ϕ are simply far field patterns due to point sources: $\phi_z := e^{-ik\widehat{x}\cdot z}$ where $\widehat{x}$ is a point on the unit sphere (the direction of the incident field), k is a nonnegative integer (the wavenumber of the incident field), and z is some point in space.

The crucial observation of Kirsch is that ϕ_z is in the range of G *if and only if* z is a point *inside* the scatterer. If one does not know where the scatter is, let alone its shape, then one does not know G, however, the Fenchel conjugate depends not on G but on the operator F which is constructed from measured data. In general, the Fenchel conjugate, and hence the Kirsch-Grinberg infimal characterization, is difficult to compute, but depending on the physical setting, there is a functional U of F under which the ranges of $U(F)$ and G coincide. In the case where F is a normal operator, $U(F) = (F^*F)^{1/4}$; for non-normal F the functional U depends more delicately on the physical problem at hand and is only known in a handful of cases. So the algorithm for determining the shape and location of a scatterer amounts to determining those points z where $e^{-ik\widehat{x}\cdot z}$ is in the range of $U(F)$ and where U and F are known and easily computed.

1.5 Fredholm integral equations

In the scattering application of the previous section the prevailing numerical technique is not to calculate the Fenchel conjugate of $h(\psi)$ but rather to explore the range of some functional of F.

Ultimately, the computation involves solving a Fredholm integral equation of the first kind. This brings us back to the more general setting with which we began. Let

$$(Ax)(s) = \int_T a(s,t)\mu(dt) = b(s)$$

for reasonable kernels and operators. If A is compact, for instance, as in most deconvolution problems of interest, the problem is *ill-posed* in the sense of Hadamard. Some sort of *regularization* technique is therefore required for numerical solutions [41,45,46,53,83]. We explore regularization in relation to the constraint qualifications (1.9) or (1.10).

Formulating the integral equation as an entropy minimization problem we have

$$\begin{array}{ll} \underset{x\in X}{\text{minimize}} & I_\varphi(x) \\ \text{subject to} & Ax = b. \end{array} \tag{1.30}$$

Following [12, Example 2.2], let T and S be the interval $[0,1]$ with Lebesgue measures μ and ν, and let $a(s,t)$ be a continuous kernel of the Fredholm operator A mapping $X := C([0,1])$ to $Y := C([0,1])$, both equipped with the supremum norm. The adjoint operator is given by $A^*y = \left\{\int_S a(s,t)\lambda(ds)\right\}\mu(dt)$ where the dual spaces are the spaces of Borel measures, $X^* = M([0,1])$ and $Y^* = M([0,1])$. Every element of the range is therefore μ-absolutely continuous and A^* can be viewed as having its range in $L_1([0,1],\mu)$. It follows from [75] that the Fenchel dual of (1.30) for the operator A is therefore

$$\max_{y^*\in Y^*} \langle b,\ y^*\rangle - I_{\varphi^*}(A^*y^*). \tag{1.31}$$

Note that the dual problem, unlike the primal, is *unconstrained*. Suppose that A is injective and that $b \in \text{range}(A)$. Assume also that φ^* is everywhere finite and differentiable. Assuming the solution $\overline{y}$ to the dual is attained, then naive application of calculus provides that

$$b = A\left(\frac{\partial\varphi^*}{\partial r}(A^*\overline{y})\right) \quad \text{and} \quad x_\varphi = \left(\frac{\partial\varphi^*}{\partial r}(A^*\overline{y})\right). \tag{1.32}$$

Similar to the counterexample explored in subsection 1.1, it is quite likely that $A\left(\frac{\partial\varphi^*}{\partial r}(\text{range}(A^*))\right)$ is smaller than the range of A, hence it is possible to have $b \in \text{range}(A)$ but not in $A\left(\frac{\partial\varphi^*}{\partial r}(\text{range}(A^*))\right)$. Thus the assumption that the solution to the dual problem is attained cannot hold and the primal-dual relationship is broken.

For a specific example, following [12, Example 2.2], consider the Laplace transform restricted to $[0,1]$: $a(s,t) := e^{-st}$ $(s \in [0,1])$, and let φ be either the Boltzmann-Shannon entropy, Fermi-Dirac entropy, or an L_p norm with $p \in (1,2)$, equations (1.3)-(1.5) respectively. Take $b(s) := \int_{[0,1]} e^{-st}\overline{x}(t)dt$ for $\overline{x} := \alpha|t-\frac{1}{2}| + \beta$, a solution to (1.30). It can be shown that the restricted Laplace operator defines an injective linear operator from $C([0,1])$ to $C([0,1])$. However, x_φ given by (1.32) is continuously differentiable, and thus cannot match the known solution $\overline{x}$ which is not differentiable. Indeed, in the case of the Boltzmann-Shannon entropy, the conjugate function and $A^*\overline{y}$ are entire hence the ostensible solution x_φ must be *infinitely* differentiable on $[0,1]$. One could guarantee that the solution to the primal problem (1.30) is attained by replacing $C([0,1])$ with $L_p([0,1])$, but this does not resolve the problem of attainment in the dual problem.

To recapture the correspondence between primal and dual problems it is necessary to regularize or, alternatively, relax the problem, or to require the constraint qualification $b \in \operatorname{core}(A \operatorname{dom} \varphi)$. Such conditions usually require A to be surjective, or at least to have closed range.

2 Background

As this is meant to be a survey of some of the more useful milestones in convex analysis, the focus is more on the connections between ideas than their proofs. For the proofs we point the reader to a variety of sources for the sake of diversity. The presentation is by default in a normed space X with dual X^*, though if statements become too technical we will specialize to Euclidean space. E denotes a finite-dimensional real vector space $\mathbb{R}^n$ for some $n \in \mathbb{N}$ endowed with the usual norm. Typically, X will be reserved for a real infinite-dimensional Banach space. A common convention in convex analysis is to include one or both of $-\infty$ and $+\infty$ in the range of functions (typically only $+\infty$). This is denoted by the (semi-) closed interval, $(-\infty, +\infty]$ or $[-\infty, +\infty]$.

A set $C \subset X$ is said to be convex if it contains all line segments between any two points in C: $\lambda x + (1-\lambda)y \in C$ for all $\lambda \in [0,1]$ and $x, y \in C$. Much of the theory of convexity is centered on the analysis of convex sets, however sets and functions are treated interchangeably through the use of level sets, epigraphs and indicator functions. The *lower level sets* of a function $f : X \to [-\infty, +\infty]$ are denoted $\operatorname{lev}_{\leq\alpha} f$ and defined by $\operatorname{lev}_{\alpha} f := \{x \in X \mid f(x) \leq \alpha\}$ where $\alpha \in \mathbb{R}$. The *epigraph* of a function $f : X \to [-\infty, +\infty]$ is defined by

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

This leads to the very natural definition of a convex function as one whose epigraph is a convex set. More directly, a convex function is defined as a mapping $f : X \to [-\infty, +\infty]$ with convex domain and

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \qquad \text{for any } x, y \in \operatorname{dom} f \text{ and } \lambda \in [0,1].$$

A proper convex function $f : X \to [-\infty, +\infty]$ is *strictly* convex if the above inequality is strict for all distinct x and y in the domain of f and all $0 < \lambda < 1$. A function is said to be *closed* if its epigraph is closed; whereas a *lower semi-continuous* (lsc) function f satisfies $\liminf_{x \to \overline{x}} f(x) \geq f(\overline{x})$ for all $\overline{x} \in X$. These properties are in fact equivalent:

Proposition 2.1. *The following properties of a function $f : X \to [-\infty, +\infty]$ are equivalent:*

(i) f is lsc;

(ii) $\operatorname{epi} f$ is closed in $X \times \mathbb{R}$;

(iii) the level sets $\operatorname{lev}_{\leq\alpha} f$ are closed on X for each $\alpha \in \mathbb{R}$.

Guide. For Euclidean spaces, this is shown in [77, Theorem 1.6]. In the Banach space setting this is [19, Proposition 4.1.1]. This is left as an exercise for the Hilbert space setting in [32, Exercise 2.1]. □

Our principal focus is on *proper* functions, that is $f : E \to [-\infty, +\infty]$ with nonempty domain. One passes from sets to functions through the indicator function

$$\iota_C(x) := \begin{cases} 0 & x \in C \\ +\infty & \text{else.} \end{cases}$$

For $C \subset X$ convex, we may refer to $f : C \to [-\infty, +\infty]$ as a convex function if the extended function

$$\overline{f}(x) := \begin{cases} f(x) & x \in C \\ +\infty & \text{else} \end{cases}$$

is convex.

2.1 Lipschitzian Properties

Convex functions have the remarkable, yet elementary, property that local boundedness and local Lipschitz properties are *equivalent* without any additional assumptions on the function. In the following statement of this fact, we denote the unit ball by $B_X := \{x \in X \mid \|x\| \leq 1\}$.

Lemma 2.2. *Let $f : X \to (-\infty, +\infty]$ be a convex function and suppose that $C \subset X$ is a bounded convex set. If f is bounded on $C + \delta B_X$ for some $\delta > 0$, then f is Lipschitz on C.*

Guide. See [19, Lemma 4.1.3]. □

With this fact, one can easily establish the following.

Proposition 2.3 (convexity and continuity in normed spaces)**.** *Let $f : X \to (-\infty, +\infty]$ be proper and convex, and let $x \in \operatorname{dom} f$. The following are equivalent:*

(i) *f is Lipschitz on some neighborhood of x;*

(ii) *f is continuous at x;*

(iii) *f is bounded on a neighborhood of x;*

(iv) *f is bounded above on a neighborhood of x.*

Guide. See [19, Proposition 4.1.4] or [16, Section 4.1.2]. □

In finite dimensions, convexity and continuity are much more tightly connected.

Proposition 2.4 (convexity and continuity in Euclidean spaces)**.** *Let $f : E \to (-\infty, +\infty]$ be convex. Then f is locally Lipschitz, and hence continuous, on the interior of its domain.*

Guide. See [19, Theorem 2.1.12], or [49, Theorem 3.1.2] □

Unlike finite-dimensions, in infinite-dimensions a convex function need not be continuous. A Hamel basis, for instance – i.e. an algebraic basis for the vector space – can be used to define discontinuous linear functionals [19, Exercise 4.1.21]. For lsc convex functions, however, the correspondence follows through. The following statement uses the notion of the *core* of a set given by Definition 1.1.

Example 2.5 (a discontinuous linear functional)**.** Let c_{00} denote the normed subspace of all finitely supported sequences in c_0, the vector space of sequences in X converging to 0; obviously c_{00} is open. Define $\Lambda : c_{00} \to \mathbb{R}$ by $\Lambda(x) = \sum x_j$ where $x = (x_j) \in c_{00}$. This is clearly a linear functional, and discontinuous at 0. Now extend Λ to a functional $\widehat{\Lambda}$ on the Banach space c_0 by taking a basis for c_0 considered as a vector space over c_{00}. In particular, $C := \widehat{\Lambda}^{-1}([-1, 1])$ is a convex set with empty interior for which 0 is a core point. Moreover, $\overline{C} = c_0$ and $\widehat{\Lambda}$ is certainly discontinuous. □

Proposition 2.6 (convexity and continuity in Banach spaces)**.** *Suppose X is a Banach space and $f : X \to (-\infty, +\infty]$ is lsc, proper and convex. Then the following are equivalent*

(i) *f is continuous at x;*

(ii) *$x \in \operatorname{int} \operatorname{dom} f$;*

(iii) *$x \in \operatorname{core} \operatorname{dom} f$.*

Guide. This is [19, Theorem 4.1.5]. See also [16, Theorem 4.1.3]. □

The above result is helpful since it is often easier to verify that a point is in the core of the domain of a convex function than in the interior.

2.2 Subdifferentials

The analog to the linear function in classical analysis is the *sublinear function* in convex analysis. A function $f : X \to [-\infty, +\infty]$ is said to be *sublinear* if

$$f(\lambda x + \gamma y) \le \lambda f(x) + \gamma f(y) \quad \text{for all } x, y \in X \text{ and } \lambda, \gamma \ge 0.$$

For this we use the convention that $0 \cdot (+\infty) = 0$. Sometimes sublinearity is defined as a function f that is *positively homogeneous (of degree 1)* – i.e. $0 \in \operatorname{dom} f$ and $f(\lambda x) = \lambda f(x)$ for all x and all $\lambda > 0$ – and is *subadditive*

$$f(x + y) \le f(x) + f(y) \quad \text{for all } x \text{ and } y.$$

Example 2.7 (norms)**.** A *norm* on a vector space is a sublinear function. Recall that a nonnegative function $\|\cdot\|$ on a vector space X is a norm if

(i) $\|x\| \geq 0$ for each $x \in X$;

(ii) $\|x\| = 0$ if and only if $x = 0$;

(iii) $\|\lambda x\| = |\lambda|\|x\|$ for every $x \in X$ and scalar λ;

(iv) $\|x + y\| \leq \|x\| + \|y\|$ for every $x, y \in X$.

A *normed space* is a vector space endowed with such a norm and is called a *Banach space* if it is *complete* which is to say that all Cauchy sequences converge. □

Another important sublinear function is the directional derivative of the function f at x in the direction d defined by

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

whenever this limit exists.

Proposition 2.8 (sublinearity of the directional derivative)**.** *Let X be a Banach space and let $f : X \to (-\infty, +\infty]$ be a convex function. Suppose that $\overline{x} \in \operatorname{core}(\operatorname{dom} f)$. Then the directional derivative $f'(\overline{x}; \cdot)$ is everywhere finite and sublinear.*

Guide. See [16, Proposition 4.2.4]. For the finite dimensional analog see [49, Proposition D.1.1.2] or [19, Proposition 2.1.17]. □

Another important instance of sublinear functions are *support functions* of convex sets which, in turn, permit local first order approximations to convex functions. A *support function* of a nonempty subset S of the dual space X^*, usually denoted σ_S, is defined by $\sigma_S(x) := \sup\{\langle s,\ x\rangle \mid s \in S\}$. The support function is convex, proper (not everywhere infinite), and $0 \in \operatorname{dom} \sigma_S$.

Example 2.9 (support functions and Fenchel conjugation)**.** From the definition of the support function it follows immediately that, for a closed convex set C,

$$\iota_C^* = \sigma_C \quad \text{and} \quad \iota_C^{**} = \iota_C.$$

□

A powerful observation is that any closed sublinear function can be viewed as a support function. To see this we represent closed convex functions via affine minorants. This is the content of the *Hahn-Banach* theorem, which we state in infinite dimensions as we will need this below.

Theorem 2.10 (Hahn-Banach: analytic form)**.** *Let X be a normed space and $\sigma : X \to \mathbb{R}$ be a continuous sublinear function with* dom $\sigma = X$*. Suppose that L is a linear subspace of X and that the linear function $h : L \to \mathbb{R}$ is dominated by σ on L, that is $\sigma \geq h$ on L. Then there is a linear function minorizing σ on X, that is, there exists a $x^* \in X^*$ dominated by σ such that $h(x) = \langle x^*, \; x\rangle \leq \sigma(x)$ for all $x \in L$.*

Guide. The proof can be carried out in finite dimensions with elementary tools, constructing x^* from h sequentially by one dimensional extensions from L. See [49, Theorem C.3.1.1], [19, Proposition 2.1.18]. The technique can be extended to Banach spaces using Zorn's lemma and a verification that the linear functionals so constructed are continuous (guaranteed by the domination property) [19, Theorem 4.1.7]. See also and [80, Theorem 1.11]. □

An important point in the Hahn-Banach extension theorem is the *existence* of a minorizing linear function, and hence the existence of the *set* of linear minorants. In fact, σ is the supremum of the linear functions minorizing it. In other words, σ is the support function of the nonempty set

$$S_\sigma := \{ s \in X^* \mid \langle s, \; x\rangle \leq \sigma(x) \quad \text{for all } x \in X \}.$$

A number of facts follow from Theorem 2.10, in particular the nonemptiness of the subdifferential, a sandwich theorem and, thence, Fenchel Duality (respectively Theorems 2.14, 2.17 and 3.10). It turns out that the converse also holds, and thus these facts are actually *equivalent* to nonemptiness of the subdifferential. This is the so-called *Hahn-Banach/Fenchel duality* circle.

As stated in Proposition 2.8, the directional derivative is everywhere finite and sublinear for a convex function f at points in the core of its domain. In light of the Hahn-Banach theorem, we then can express $f'(\overline{x}, \cdot)$ for all $d \in X$ in terms of its minorizing function:

$$f'(\overline{x}, d) = \sigma_S(d) = \max_{v \in S} \{\langle v, \; d\rangle\}.$$

The set S for which $f'(\overline{x}, d)$ is the support function has a special name: the *subdifferential* of f at $\overline{x}$. It is tempting to *define* the subdifferential this way, however there is a more elemental definition that does not rely on directional derivatives or support functions, or indeed even the convexity of f. We prove the correspondence between directional derivatives of convex functions and the subdifferential below as a consequence of the Hahn-Banach theorem.

Definition 2.11 (subdifferential)**.** *For a function $f : X \to (-\infty, +\infty]$ and a point $\overline{x} \in$ dom f, the* subdifferential *of f at $\overline{x}$, denoted $\partial f(\overline{x})$ is defined by*

$$\partial f(\overline{x}) := \{ v \in X^* \mid v(x) - v(\overline{x}) \leq f(x) - f(\overline{x}), \; \textit{for all } x \in X \}.$$

when $\overline{x} \notin$ dom f we define $\partial f(\overline{x}) = \emptyset$.

In Euclidean space the subdifferential is just

$$\partial f(\overline{x}) = \{ v \in E \mid \langle v, \; x\rangle - \langle v, \; \overline{x}\rangle \leq f(x) - f(\overline{x}), \; \text{for all } x \in E \}.$$

An element of $\partial f(x)$ is called a *subgradient* of f at x. See [16, 64, 77] for more in-depth discussion of the regular, or limiting subdifferential we have defined here, in addition to other useful varieties.

This is a generalization of the classical gradient. Just as the gradient need not exist, the subdifferential of a lsc convex function may be empty at some points in its domain. Take, for example, $f(x) = -\sqrt{1-x^2}$ for $-1 \le x \le 1$. Then $\partial f(x) = \emptyset$ for $x = \pm 1$.

Example 2.12 (common subdifferentials)**.**

(i) Gradients. A function $f : X \to \mathbb{R}$ is said to be *strictly differentiable* at $\overline{x}$ if

$$\lim_{x \to \overline{x}, u \to \overline{x}} \frac{f(x) - f(u) - \nabla f(\overline{x})(x-u)}{\|x-u\|} = 0.$$

This is a stronger differentiability property than Fréchet differentiability since it requires uniformity in *pairs* of points converging to $\overline{x}$. Luckily for convex functions the two notions agree. If f is convex and strictly differentiable at $\overline{x}$, then the subdifferential is exactly the gradient. (This follows from the equivalence of the subdifferential in Definition 2.11 and the basic limiting subdifferential defined in [64, Definition 1.77] for convex functions, and [64, Corollary 1.82].) In finite dimensions, at a point $\overline{x} \in \operatorname{dom} f$ for f convex, Fréchet and Gâteaux differentiability coincide, and the subdifferential is a singleton [19, Theorem 2.2.1]. In infinite dimensions, a convex function f that is continuous at $\overline{x}$ is Gâteaux differentiable at $\overline{x}$ if and only if the $\partial f(\overline{x})$ is a singleton [19, Corollary 4.2.11].

(ii) The subdifferential of the indicator function.

$$\partial \iota_C(\overline{x}) = N_C(\overline{x})$$

where $C \subset X$ is closed and convex, X is a Banach, and $N_C(\overline{x}) \subset X^*$ is the *normal cone mapping* to C at $\overline{x}$ defined by

$$N_C(\overline{x}) := \begin{cases} \{v \in X^* \mid \langle v,\ x - \overline{x} \rangle \le 0 \quad \text{for all } x \in C\} & \overline{x} \in C \\ \emptyset & \overline{x} \notin C. \end{cases} \tag{2.1}$$

See (3.6) for alternative definitions and further discussion of this important mapping.

(iii) Absolute value. For $x \in \mathbb{R}$,

$$\partial |\cdot|(x) = \begin{cases} -1 & x < 0 \\ [-1, 1] & x = 0 \\ 1 & x > 0 \end{cases}$$

□

The following elementary observation suggests the fundamental significance of subdifferential in optimization.

Theorem 2.13 (subdifferential at optimality: Fermat's rule)**.** *Let X be a normed space, and let $f : X \to (-\infty, +\infty]$ be proper and convex. Then f has a (global) minimum at $\overline{x}$ if and only if $0 \in \partial f(\overline{x})$.*

Guide. The first implication of the global result follows from a more general local result [64, Proposition 1.114] by convexity; the converse statement follows from the definition of the subdifferential and convexity. □

Returning now to the correspondence between the subdifferential and the directional derivative of a convex function $f'(x;d)$, on has the following fundamental result.

Theorem 2.14 (max formula - existence of ∂f)**.** *Let X be a normed space, $d \in X$ and let $f : X \to (-\infty, +\infty]$ be convex. Suppose that $\overline{x} \in \operatorname{cont} f$. Then $\partial f(\overline{x}) \neq \emptyset$ and*

$$f'(\overline{x}, d) = \max\{\langle x^*,\ d\rangle \mid x^* \in \partial f(\overline{x})\}$$

Proof. The tools are in place for a simple proof that synthesizes many of the facts tabulated so far. By Proposition 2.8 $f'(\overline{x};\cdot)$ is finite so, for fixed $d \in \{x \in X \mid \|x\| = 1\}$, let $\alpha = f'(\overline{x};d) < \infty$. The stronger assumption that $\overline{x} \in \operatorname{cont} f$ and the convexity of $f'(\overline{x};\cdot)$ yield that the directional derivative is Lipschitz continuous with constant K. Let $S := \{td \mid t \in \mathbb{R}\}$ and define the linear function $\Lambda : S \to \mathbb{R}$ by $\Lambda(td) := t\alpha$ for $t \in \mathbb{R}$. Then $\Lambda(\cdot) \leq f'(\overline{x};\cdot)$ on S. The Hahn-Banach theorem 2.10 then guarantees the existence of $\phi \in X^*$ such that

$$\phi = \Lambda \text{ on } S, \quad \phi(\cdot) \leq f'(\overline{x};\cdot) \text{ on } X.$$

Then $\phi \in \partial f(\overline{x})$ and $\phi(sd) = f'(\overline{x};sd)$ for all $s \geq 0$. □

A simple example on $\mathbb{R}$ illustrates the importance of the qualification $\overline{x} \in \operatorname{cont} f$. Let

$$f(x) : \mathbb{R} \to (-\infty, +\infty] := \begin{cases} -\sqrt{x}, & x \geq 0 \\ +\infty & \text{otherwise.} \end{cases}$$

For this example $\partial f(0) = \emptyset$.

An important application of the Max formula in finite dimensions is the mean value theorem for convex functions.

Theorem 2.15 (convex mean value theorem)**.** *Let $f : E \to (-\infty, +\infty]$ be convex and continuous. For $u, v \in E$ there exists a point $z \in E$ interior to the line segment $[u, v]$ with*

$$f(u) - f(v) \leq \langle w,\ u - v\rangle, \quad \textit{for all } w \in \partial f(z).$$

Guide. See [64, 77] for extensions of this result and detailed historical background. □

The next theorem is a key tool in developing a subdifferential calculus. It relies on assumptions that are used frequently enough that we present them separately.

Assumption 2.16. *Let X and Y be Banach spaces and let $T : X \to Y$ be a bounded linear mapping. Let $f : X \to (-\infty, +\infty]$ and $g : Y \to (-\infty, +\infty]$ satisfy one of*

$$0 \in \operatorname{core}(\operatorname{dom} g - T \operatorname{dom} f) \text{ and both } f \text{ and } g \text{ are lsc,} \tag{2.2}$$

or

$$T \operatorname{dom} f \cap \operatorname{cont} g \neq \emptyset. \tag{2.3}$$

The later assumption can be used in incomplete normed spaces as well.

Theorem 2.17 (sandwich theorem). *Let X and Y be Banach spaces and let $T : X \to Y$ be a bounded linear mapping. Suppose that $f : X \to (-\infty, +\infty]$ and $g : Y \to (-\infty, +\infty]$ are proper convex functions with $f \geq -g \circ T$ and which satisfy Assumption 2.16. Then there is an affine function $A : X \to \mathbb{R}$ defined by $Ax := \langle T^* y^*, x\rangle + r$ satisfying $f \geq A \geq -g \circ T$. Moreover, for any $\overline{x}$ satisfying $f(\overline{x}) = (-g \circ T)(\overline{x})$, we have $-y^* \in \partial g(T\overline{x})$.*

Guide. By our development to this point, we would use the Max formula [19, Theorem 4.1.18] to prove the result. For a vector space version see [80, Corollary 2.1]. Another route is via Fenchel duality which we explore in the next section. A third approach closely related to the Fenchel duality approach [16, Theorem 4.3.2] is based on a *decoupling* lemma presented in the next section (Lemma 3.9). □

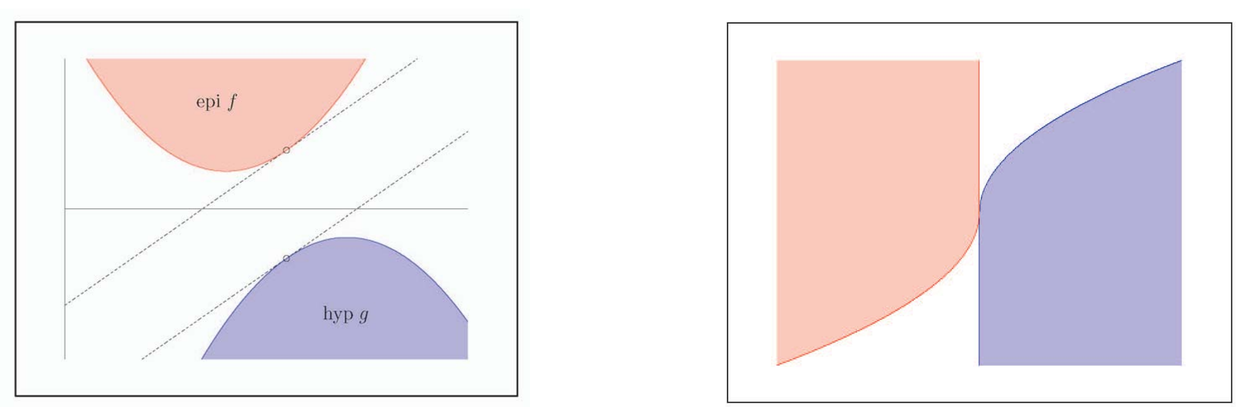
Corollary 2.18 (basic separation). *Let $C \subset X$ be a nonempty convex set with nonempty interior in a normed space, and suppose $x_0 \notin \operatorname{int} C$. Then there exists $\phi \in X^* \setminus \{0\}$ such that*

$$\sup_C \phi \leq \phi(x_0) \quad \text{and} \quad \phi(x) < \phi(x_0) \text{ for all } x \in \operatorname{int} C.$$

If $x_0 \notin \overline{C}$ then we may assume $\sup_C \phi < \phi(x_0)$.

Proof. Assume without loss of generality that $0 \in \operatorname{int} C$ and apply the sandwich theorem with $f = \iota_{\{x_0\}}$, T the identity mapping on X and $g(x) = \inf\{r > 0 \mid x \in rC\} - 1$. See [16, Theorem 4.3.8] and [19, Corollary 4.1.15]. □

The Hahn-Banach theorem 2.10 can be seen as an easy consequence of the sandwich theorem 2.17, which completes part of the circle. Figure 1 illustrates these ideas



(a) Success (b) Failure

Figure 1: Hahn Banach sandwich theorem and its failure.

In the next section we will add Fenchel duality to this cycle. Before doing so, we finish with a calculus of subdifferentials and a few fundamental results connecting the subdifferential to classical derivatives and *monotone operators*.

Theorem 2.19 (subdifferential sum rule)**.** *Let X and Y be Banach spaces, $T : X \to Y$ a bounded linear mapping and let $f : X \to (-\infty, +\infty]$ and $g : Y \to (-\infty, +\infty]$ be convex functions. Then at any point $x \in X$ we have*

$$\partial \left(f + g \circ T\right)(x) \supset \partial f(x) + T^*(\partial g(Tx)),$$

with equality if Assumption 2.16 holds.

Proof sketch. The inclusion is clear. Proving equality permits an elegant proof using the sandwich theorem [19, Theorem 4.1.19], which we sketch here. Take $\phi \in \partial \left(f + g \circ T\right)(\overline{x})$ and assume without loss of generality that

$$x \mapsto f(x) + g(Tx) - \phi(x)$$

attains a minimum of 0 at $\overline{x}$. By Theorem 2.17 there is an affine function $A := \langle T^* y^*, \ \cdot \rangle + r$ with $-y^* \in \partial g(T\overline{x})$ such that

$$f(x) - \phi(x) \geq Ax \geq -g(Ax).$$

Equality is attained at $x = \overline{x}$. It remains to check that $\phi + T^* y^* \in \partial f(\overline{x})$. □

The next result is a useful extension to Proposition 2.3.

Theorem 2.20 (convexity and regularity in normed spaces)**.** *Let $f : X \to (-\infty, +\infty]$ be proper and convex, and let $x \in \operatorname{dom} f$. The following are equivalent:*

(i) *f is Lipschitz on some neighborhood of x;*

(ii) *f is continuous at x;*

(iii) *f is bounded on a neighborhood of x;*

(iv) *f is bounded above on a neighborhood of x.*

(v) *∂f maps bounded subsets of X into bounded nonempty subsets of X^*.*

Guide. See [19, Theorem 4.1.25]. □

The next results relate to Example 2.12 and provide additional tools for verifying differentiability of convex functions. The notation $\to_{w^*}$ to denotes weak* convergence.

Theorem 2.21 (Šmulian)**.** *Let the convex function f be continuous at $\overline{x}$.*

(i) *The following are equivalent:*

(a) *f is Fréchet differentiable at $\overline{x}$.*

(b) *For each sequence $x_n \to \overline{x}$ and $\phi \in \partial f(\overline{x})$, there exist $\overline{n} \in \mathbb{N}$ and $\phi_n \in \partial f(x_n)$ for $n \geq \overline{n}$ such that $\phi_n \to \phi$.*

(c) *$\phi_n \to \phi$ whenever $\phi_n \in \partial f(x_n)$, $\phi \in \partial f(\overline{x})$.*

(ii) *The following are equivalent:*

(a) *f is Gâteaux differentiable at $\overline{x}$.*

(b) *For each sequence $x_n \to \overline{x}$ and $\phi \in \partial f(\overline{x})$, there exist $\overline{n} \in \mathbb{N}$ and $\phi_n \in \partial f(x_n)$ for $n \geq \overline{n}$ such that $\phi_n \to_{w^*} \phi$.*

(c) *$\phi_n \to_{w^*} \phi$ whenever $\phi_n \in \partial f(x_n)$, $\phi \in \partial f(\overline{x})$.*

A more complete statement of these facts and their provenance can be found in [19, Theorem 4.2.8-9]. In particular, in every infinite dimensional normed space there is a continuous convex function which is Gâteaux but not Fréchet differentiable at the origin.

An elementary but powerful observation about the subdifferential viewed as a multi-valued mapping will conclude this section. A multi-valued mapping T from X to X^* is denoted with double arrows, $T: X \rightrightarrows X^*$. Then T is *monotone* if

$$\langle v_2 - v_1, x_2 - x_1 \rangle \geq 0 \quad \text{whenever} \quad v_1 \in T(x_1),\ v_2 \in T(x_2).$$

Proposition 2.22 (monotonicity and convexity)**.** *Let $f: X \to (-\infty, +\infty]$ be proper and convex on a normed space. Then the subdifferential mapping $\partial f: X \rightrightarrows X^*$ is monotone.*

Proof. Add the subdifferential inequalities in the Definition 2.11 applied to $f(x_1)$ and $f(x_0)$ for $v_1 \in \partial f(x_1)$ and $v_0 \in \partial(f(x_0)$. □

3 Duality and Convex Analysis

The Fenchel conjugate is to convex analysis what the Fourier transform is to harmonic analysis. We begin by collecting some basic facts about this fundamental tool.

3.1 Fenchel Conjugation

The Fenchel conjugate, introduced in [43], of a mapping $f: X \to [-\infty, +\infty]$, as mentioned above is denoted $f^*: X^* \to [-\infty, +\infty]$ and defined by

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

The conjugate is always convex (as a supremum of affine functions). If the domain of f is nonempty, then f^* never takes the value $-\infty$.

Example 3.1 (important Fenchel conjugates)**.**

(i) Absolute value.

$$f(x) = |x| \quad (x \in \mathbb{R}), \qquad f^*(y) = \begin{cases} 0 & y \in [-1,1] \\ +\infty & \text{else.} \end{cases}$$

(ii) L_p norms ($p > 1$).

$$f(x) = \frac{1}{p}\|x\|^p \quad (p > 1), \qquad f^*(y) = \frac{1}{q}\|y\|^q \quad \left(\tfrac{1}{p} + \tfrac{1}{q} = 1\right).$$

In particular, note that the 2-norm is "self-conjugate".

(iii) Indicator functions.

$$f = \iota_C, \qquad f^* = \sigma_C$$

where σ_C is the *support function* of the set C. Note that if C is not closed and convex, then the conjugate of σ_C, that is the *biconjugate* of ι_C, is the *closed convex hull* of C. (See Proposition 3.3(ii) below.)

(iv) Boltzmann-Shannon entropy.

$$f(x) = \begin{cases} x \ln x - x & (x > 0) \\ 0 & (x = 0) \end{cases}, \qquad f^*(y) = e^y \quad (y \in \mathbb{R}).$$

(v) Fermi-Dirac entropy.

$$f(x) = \begin{cases} x \ln x + (1-x)\ln(1-x) & (x \in (0,1)) \\ 0 & (x = 0,1) \end{cases}, \qquad f^*(y) = \ln(1 + e^y) \quad (y \in \mathbb{R}).$$

□

Some useful properties of conjugate functions are tabulated below.

Proposition 3.2 (Fenchel-Young inequality)**.** *Let X be a normed space and let $f : X \to [-\infty, +\infty]$. Suppose that $x^* \in X^*$ and $x \in \operatorname{dom} f$. Then*

$$f(x) + f^*(x^*) \geq \langle x^*,\ x\rangle\,. \tag{3.1}$$

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

Proof sketch. The proof follows by elementary application of the definitions of the Fenchel conjugate and the subdifferential. See [73] for the finite dimensional case. The same proof works in the normed space setting. □

The conjugate, as the supremum of affine functions, is convex. In the following, we denote the closure of a function f by $\overline{f}$, and we let $\overline{\operatorname{conv}}\, f$ be the function whose epigraph is the closed convex hull of the epigraph of f.

Proposition 3.3. *Let X be a normed space and let $f : X \to [-\infty, +\infty]$.*

(i) *If $f \geq g$ then $g^* \geq f^*$.*

(ii) $f^* = \left(\overline{f}\right)^* = (\overline{\operatorname{conv}} f)^*$

Proof. The definition of the conjugate immediately implies (i). This immediately yields $f^* \leq \left(\overline{f}\right)^* \leq (\overline{\operatorname{conv}} f)^*$. To show (ii) it remains to show that $f^* \geq (\overline{\operatorname{conv}} f)^*$. Choose any $\phi \in X^*$. If $f^*(\phi) = +\infty$ the conclusion is clear, so assume $f^*(\phi) = \alpha$ for some $\alpha \in \mathbb{R}$. Then $\phi(x) - f(x) \leq \alpha$ for all $x \in X$. Define $g := \phi - f$. Then $g \leq \overline{\operatorname{conv}} f$ and, by (i) $(\overline{\operatorname{conv}} f)^* \leq g^*$. But $g^* = \alpha$, so $(\overline{\operatorname{conv}} f)^* \leq \alpha = f^*(\phi)$. □

Application of Fenchel conjugation twice, or *biconjugation* denoted by f^{**}, is a function on X^{**}. In certain instances biconjugation is the identity – in this way, the Fenchel conjugate resembles the Fourier transform. Indeed, Fenchel conjugation plays a role in convex analysis similar to the Fourier transform in harmonic analysis, and has a contemporaneous provenance dating back to Legendre.

Proposition 3.4 (biconjugation)**.** *Let $f : X \to (-\infty, +\infty]$, $x \in X$ and $x^* \in X^*$.*

(i) $f^{**}|_X \leq f$.

(ii) *If f is convex and proper, then $f^{**}(x) = f(x)$ at x if and only if f is lsc at x. In particular, f is lsc if and only if $f^{**}_X = f$.*

(iii) $f^{**}|_X = \overline{\operatorname{conv}} f$ *if* $\overline{\operatorname{conv}} f$ *is proper.*

Guide. (i) follows from Fenchel-Young, Theorem 3.2 and the definition of the conjugate. (ii) follows from (i) and an epi-separation property [19, Proposition 4.4.2]. (iii) follows from (ii) of this proposition and 3.3(ii). □

The next results highlight the relationship between the Fenchel conjugate and the subdifferential that we have already made use of in (1.28).

Proposition 3.5. *Let $f : X \to (-\infty, +\infty]$ be a function and $\overline{x} \in \operatorname{dom} f$. If $\phi \in \partial f(\overline{x})$ then $\overline{x} \in \partial f^*(\psi)$. If, additionally, f is convex and lsc at $\overline{x}$, then the converse holds, namely $\overline{x} \in \partial f^*(\phi)$ implies $\phi \in \partial f(\overline{x})$.*

Guide. See [49, Corollary 1.4.4] for the finite dimensional version of this fact that, with some modification, can be extended to normed spaces. □

To close this subsection we introduce *infimal convolutions*. Among their many applications are smoothing and approximation—just as is the case for integral convolutions.

Definition 3.6 (infimal convolution)**.** *Let f and g be proper extended real-valued functions on a normed space X. The* infimal convolution *of f and g is defined by*

$$(f\Box g)(x) := \inf_{y\in X} f(y) + g(x-y).$$

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; consequently it is a convex function when f and g are convex.

The next lemma follows directly from the definitions and careful application of the properties of suprema and infima.

Lemma 3.7. *Let X be a normed space and let f and g be proper functions on X , then $(f\Box g)^* = f^* + g^*$.*

An important example of infimal convolution is *Yosida approximation.*

Theorem 3.8 (Yosida approximation)**.** *Let $f : X \to \mathbb{R}$ be convex and bounded on bounded sets. Then both $f\Box n\|\cdot\|^2$ and $f\Box n\|\cdot\|$ converge uniformly to f on bounded sets.*

Guide. This follows from the above lemma and basic approximation facts. □

In the inverse problems literature $\left(f\Box n\|\cdot\|^2\right)(0)$ is often referred to as *Tikhonov regularization*; elsewhere, $f\Box n\|\cdot\|^2$ is referred to as *Moreau-Yosida regularization* because $f\Box\frac{1}{2}\|\cdot\|^2$, the *Moreau envelope*, was studied in depth by Moreau [65, 66]. The argmin mapping corresponding to the Moreau envelope—that is the mapping of $x \in X$ to the point $\overline{y} \in X$ at which the value of $f\Box\frac{1}{2}\|\cdot\|^2$ is attained—is called the *proximal mapping* [65, 66, 77]

$$\operatorname{prox}_{\lambda,f}(x) := \operatorname{argmin}_{y\in X} f(y) + \frac{1}{2\lambda}\|x-y\|^2. \tag{3.2}$$

When f is the indicator function of a closed convex set C, the proximal mapping is just the *metric projection* onto C, denoted by $P_C(x)$: $\operatorname{prox}_{\lambda,\iota_C}(x) = P_C(x)$.

3.2 Fenchel duality

Fenchel duality can be proved by Theorem 2.14 and the sandwich theorem 2.17 [19, Theorem 4.4.18]. According to our development, then, this places Fenchel duality as a consequence of the Hahn-Banach theorem. In order to close the Fenchel duality/Hahn-Banach circle of ideas, however, following [16] we prove the main duality result of this section using the Fenchel-Young inequality and the next important lemma.

Lemma 3.9 (decoupling)**.** *Let X and Y be Banach spaces and let $T : X \to Y$ be a bounded linear mapping. Suppose that $f : X \to (-\infty, +\infty]$ and $g : Y \to (-\infty, +\infty]$ are proper convex functions which satisfy Assumption 2.16. Then there is a $y^* \in Y^*$ such that for any $x \in X$ and $y \in Y$,*

$$p \le (f(x) - \langle y^*,\ Tx\rangle) + (g(y) + \langle y^*,\ y\rangle),$$

where $p := \inf_X\{f(x) + g(Tx)\}$.

Guide. Define the perturbed function $h : Y \to [-\infty, +\infty]$ by

$$h(u) := \inf_{x \in X} \{f(x) + g(Tx + u)\}$$

which has the property that h is convex, $\text{dom } h = \text{dom } g - T \text{ dom } f$ and (the most technical part of the proof) $0 \in \text{int}(\text{dom } h)$. This can be proved by assuming the first of the constraint qualifications (2.2). The second condition (2.3) implies (2.2). Then by Theorem 2.14 we have $\partial h(0) \neq \emptyset$, which guarantees the attainment of a minimum of the perturbed function. The decoupling is achieved through a particular choice of the perturbation u. See [16, Lemma 4.3.1]. ◻

One can now provide an elegant proof of Theorem 1.2, which is restated here for convenience.

Theorem 3.10 (Fenchel duality)**.** *Let X and Y be normed spaces, consider the functions $f : X \to (-\infty, +\infty]$ and $g : Y \to (-\infty, +\infty]$ and let $T : X \to Y$ be a bounded linear map. Define the primal and dual values p, $d \in [-\infty, +\infty]$ by the Fenchel problems*

$$p = \inf_{x \in X} \{f(x) + g(Tx)\} \tag{3.3}$$

$$d = \sup_{y^* \in Y^*} \{-f^*(T^* y^*) - g^*(-y^*)\}. \tag{3.4}$$

These values satisfy the weak duality inequality $p \geq d$.

If X, Y are Banach, f, g are convex and satisfy Assumption 2.16 then $p = d$, and the supremum to the dual problem is attained if finite.

Proof. Weak duality follows directly from the Fenchel-Young inequality.

For equality assume that $p \neq -\infty$ (this case is clear). Then Assumption 2.16 guarantees that $p < +\infty$, and by the decoupling lemma (Lemma 3.9) there is a $\phi \in Y^*$ such that for all $x \in X$ and $y \in Y$

$$p \leq (f(x) - \langle \phi,\ Tx \rangle) + (g(y) - \langle -\phi,\ y \rangle).$$

Taking the infimum over all x and then over all y yields

$$p \leq -f^*(T^*, \phi) - g^*(-\phi) \leq d \leq p$$

hence ϕ attains the supremum in (3.4), and $p = d$. ◻

Fenchel duality for *linear constraints*, Corollary 1.3, follows immediately by taking $g := \iota_{\{b\}}$.

3.3 Applications

Calculus. Fenchel duality is, in some sense, the dual space representation of the sandwich theorem. It is a straight forward exercise to derive Fenchel duality from the Theorem 2.17. Conversely, the existence of a point of attainment in Theorem 3.10 yields an explicit construction of the linear mapping in Theorem 2.17: $A := \langle T^* \phi,\ \cdot \rangle + r$ where ϕ is the point of attainment in (3.4) and

$r \in [a, b]$ where $a := \inf_{x \in X} f(x) - \langle T^*\phi,\ x\rangle$ and $b := \sup_{z \in X} -g(Tz) - \langle T^*\phi,\ z\rangle$. One could then derive all the theorems using the sandwich theorem, in particular the Hahn-Banach theorem 2.10 and the subdifferential sum rule, Theorem 2.19, as consequences of Fenchel duality instead. This establishes the *Hahn-Banach/Fenchel duality* circle: each of these facts is *equivalent* and easily interderivable with the nonemptiness of the subgradient of a function at a point of continuity.

An immediate consequence of Fenchel duality is a calculus of polar cones. Define the negative polar cone of a set K in a Banach space X by

$$K^- = \{x^* \in X^* \mid \langle x^*,\ x\rangle \le 0\ \forall\, x \in K\}. \tag{3.5}$$

An important example of a polar cone that we have seen in the applications is the *normal cone* of a convex set K at a point $x \in K$, defined by (2.1). Note that

$$N_K(\overline{x}) := (K - \overline{x})^- \tag{3.6}$$

Corollary 3.11 (polar cone calculus)**.** *Let X and Y be Banach spaces and $K \subset X$ and $H \subset Y$ be cones, and let $A : X \to Y$ be a bounded linear map. Then*

$$K^- + A^*H^- \subset \left(K + A^{-1}H\right)^-$$

where equality holds if K and H are closed convex cones which satisfy $H - AK = Y$.

This can be used to easily establish the normal cone calculus for closed convex sets C_1 and C_2 at a point $x \in C_1 \cap C_2$

$$N_{C_1 \cap C_2}(x) \supset N_{C_1}(x) + N_{C_2}(x)$$

with equality holding if, in addition, $0 \in \operatorname{core}(C_1 - C_2)$ or $C_1 \cap \operatorname{int} C_2 \neq \emptyset$.

Optimality Conditions. Another important consequence of these ideas is the Pshenichnyi-Rockafellar [71, 73] condition for optimality for nonsmooth constrained optimization.

Theorem 3.12 (Pshenichnyi-Rockafellar conditions)**.** *Let X be a Banach space, let $C \subset X$ be closed and convex, and let $f : X \to (-\infty, +\infty]$ be a convex function. Suppose that either* $\operatorname{int} C \cap \operatorname{dom} f \neq \emptyset$ *and f is bounded below on C, or* $C \cap \operatorname{cont} f \neq \emptyset$*. Then there is an affine function $\alpha \le f$ with $\inf_C f = \inf_C \alpha$. Moreover, $\overline{x}$ is a solution to*

$$(\mathcal{P}_0) \qquad \begin{array}{ll} \underset{x \in X}{\textit{minimize}} & f(x) \\ \textit{subject to} & x \in C \end{array}$$

if and only if

$$0 \in \partial f(\overline{x}) + N_C(\overline{x})$$

Guide. Apply the subdifferential sum rule to $f + \iota_C$ at $\overline{x}$. □

A slight generalization extends this to linear constraints

$$(\mathcal{P}_{lin}) \qquad \begin{array}{ll} \underset{x \in X}{\text{minimize}} & f(x) \\ \text{subject to} & Tx \in D \end{array}$$

Theorem 3.13 (first-order necessary and sufficient)**.** *Let X and Y be Banach spaces with $D \subset Y$ convex, and let $f : X \to (-\infty, +\infty]$ be convex and $T : X \to Y$ a bounded linear mapping. Suppose further that one of the following holds:*

$$0 \in \operatorname{core}(D - T \operatorname{dom} f), \; D \text{ is closed and } f \text{ is lsc}, \tag{3.7}$$

or

$$T \operatorname{dom} f \cap \operatorname{int}(D) \neq \emptyset. \tag{3.8}$$

Then the feasible set $C := \{x \in X \mid Tx \in D\}$ satisfies

$$\partial(f + \iota_C)(x) = \partial f(x) + T^*(N_D(Tx)), \tag{3.9}$$

and $\overline{x}$ is a solution to ($\mathcal{P}_{lin}$) if and only if

$$0 \in \partial f(\overline{x}) + T^*(N_D(T\overline{x})). \tag{3.10}$$

A point $y^* \in Y^*$ satisfying $T^* y^* \in -\partial f(\overline{x})$ in Theorem 3.13 is a *Lagrange multiplier.*

Lagrangian duality. We limit the setting to Euclidean space and consider the general convex program

$$(\mathcal{P}_{cvx}) \qquad \begin{array}{ll} \underset{x \in E}{\text{minimize}} & f_0(x) \\ \text{subject to} & f_j(x) \leq 0 \quad (j = 1, 2, \ldots, m) \end{array}$$

where the functions f_j for $j = 0, 1, 2, \ldots, m$ are convex and satisfy

$$\bigcap_{j=0}^{m} \operatorname{dom} f_j \neq \emptyset. \tag{3.11}$$

Define the *Lagrangian* $L : E \times \mathbb{R}^m{}_+ \to (-\infty, +\infty]$ by

$$L(x, \lambda) := f_0(x) + \lambda^T F(x)$$

where $F := (f_1, f_2, \ldots, f_m)^T$. A *Lagrange multiplier* in this context is a vector $\overline{\lambda} \in \mathbb{R}^m{}_+$ for a feasible solution $\overline{x}$ if $\overline{x}$ minimizes the function $L(\cdot, \overline{\lambda})$ over E and $\overline{\lambda}$ satisfies the so-called *complimentary slackness conditions* : $\overline{\lambda}_j = 0$ whenever $f_j(\overline{x}) < 0$. On the other hand, if $\overline{x}$ is feasible for the convex program ($\mathcal{P}_{cvx}$) and there is a Lagrange multiplier, then $\overline{x}$ is optimal. Existence of the Lagrange multiplier is guaranteed by the following *Slater constraint qualification* first introduced in the 1950s.

Assumption 3.14 (Slater constraint qualification)**.** *There exists an $\widehat{x} \in \operatorname{dom} f_0$ with $f_j(\widehat{x}) < 0$ for $j = 1, 2, \ldots, m$.*

Theorem 3.15 (Lagrangian necessary conditions)**.** *Suppose that $\overline{x} \in \operatorname{dom} f_0$ is optimal for the convex program ($\mathcal{P}_{cvx}$) and that Assumption 3.14 holds. Then there is a Lagrange multiplier vector for $\overline{x}$.*

Guide. See [15, Theorem 3.2.8]. □

Denote the optimal value of $(\mathcal{P}_{cvx})$ by p. Note that, since

$$\sup_{\lambda \in \mathbb{R}^m{}_+} L(x,\lambda) = \begin{cases} f(x) & \text{if } x \in \operatorname{dom} f \\ +\infty & \text{otherwise,} \end{cases}$$

then

$$p = \inf_{x \in E} \sup_{\lambda \in \mathbb{R}^m{}_+} L(x,\lambda). \tag{3.12}$$

It is natural, then to consider the problem

$$d = \sup_{\lambda \in \mathbb{R}^m{}_+} \inf_{x \in E} L(x,\lambda) \tag{3.13}$$

where d is the *dual value*. It follows immediately that $p \geq d$. The difference between d and p is called the *duality gap*. The interesting problem is to determine when the gap is zero, that is when $d = p$.

Theorem 3.16 (dual attainment)**.** *If Assumption 3.14 holds for the convex programming problem $(\mathcal{P}_{cvx})$, then the primal and dual values are equal and the dual value is attained if finite.*

Guide. For a more detailed treatment of the theory of Lagrangian duality see [15, Section 4.3] □

3.4 Optimality and Lagrange Multipliers

In the previous sections we introduced duality theory via the Hahn-Banach/Fenchel duality circle of ideas to provide many entry points to the theory of convex and variational analysis. For our purposes, however, the real significance of duality lies with its power to illuminate duality in convex optimization, not only as a theoretical phenomenon, but as an algorithmic strategy.

In order to get to optimality criteria and the existence of solutions to convex optimization problems, we turn our focus to the approximation of minima, or more generally the *regularity* and *well-posedness* of convex optimization problems. Due to its reliance on the Slater constraint qualification 3.14, Theorem 3.16 is not adequate for problems with equality constraints:

$$(\mathcal{P}_{eq}) \qquad \begin{array}{ll} \underset{x \in S}{\text{minimize}} & f_0(x) \\ \text{subject to} & F(x) = 0 \end{array}$$

for $S \subset E$ closed and $F : E \to Y$ a Fréchet differentiable mapping between the Euclidean spaces E and Y.

More generally, we consider problems of the form

$$(\mathcal{P}_E) \qquad \begin{array}{ll} \underset{x \in S}{\text{minimize}} & f_0(x) \\ \text{subject to} & F(x) \in D \end{array} \tag{3.14}$$

for E and Y Euclidean spaces, and $S \subset E$ and $D \subset Y$, are convex but not necessarily with nonempty interior.

Example 3.17 (simple Karush Kuhn-Tucker)**.** For linear optimization problems, relatively elementary linear algebra is all that is needed to assure the existence of Lagrange multipliers. Consider

$$(\mathcal{P}_E) \qquad \begin{array}{ll} \underset{x \in S}{\text{minimize}} & f_0(x) \\ \text{subject to} & f_j(x) \in D_j, \quad j = 1, 2, \ldots, m \end{array}$$

for $f_j : \mathbb{R}^n \to \mathbb{R}$ $(j = 0, 1, 2, \ldots, s)$ continuously differentiable, $f_j : \mathbb{R}^n \to \mathbb{R}$ $(j = s+1, \ldots, m)$ linear. Suppose $S \subset E$ is closed and convex, while $D_i := (-\infty, 0]$ for $j = 1, 2, \ldots, s$ and $D_j := \{0\}$ for $j = s+1, \ldots, m$.

Theorem 3.18. *Denote by $f'_{\mathbb{J}}(x)$ the submatrix of the Jacobian of $(f_1, \ldots, f_s)^T$ (assuming this is defined at x) consisting only of those f'_j for which $f_j(x) = 0$. In other words, $f'_{\mathbb{J}}(x)$ is the Jacobian of the* active inequality constraints *at x. Let $\overline{x}$ be a local minimizer for $(\mathcal{P}_E)$ at which f_j are continuously differentiable $(j = 0, 1, \ldots, s)$ and the matrix*

$$\begin{pmatrix} f'_{\mathbb{J}}(\overline{x}) \\ A \end{pmatrix} \tag{3.15}$$

is full-rank where $A := (\nabla f_{s+1}, \ldots, \nabla f_m)^T$. Then there are $\overline{\lambda} \in \mathbb{R}^s$ and $\overline{\mu} \in \mathbb{R}^m$ satisfying

$$\overline{\lambda} \geq 0 \tag{3.16a}$$

$$(f_1(\overline{x}), \ldots, f_s(\overline{x}))\overline{\lambda} = 0 \tag{3.16b}$$

$$f'_0(\overline{x}) + \sum_{j=1}^{s} \overline{\lambda}_j f'_j(\overline{x}) + \overline{\mu}^T A = 0 \tag{3.16c}$$

Guide. An elegant and elementary proof is given in [21]. □

For more general constraint structure, *regularity* of the feasible region is essential for the normal cone calculus which plays a key role in the requisite optimality criteria. More specifically, we consider the following constraint qualification.

Assumption 3.19 (basic constraint qualification)**.**

$$y = (0, \ldots, 0) \textit{ is the only solution in } N_D(F(\overline{x})) \textit{ to } 0 \in \nabla F^T(\overline{x})y + N_S(\overline{x})$$

Theorem 3.20 (optimality on sets with constraint structure)**.** *Let*

$$C = \{x \in S \mid F(x) \in D\}$$

for $F = (f_1, f_2, \ldots, f_m) : E \to \mathbb{R}^m$ with f_j continuously differentiable $(j = 1, 2, \ldots, m)$, $S \subset E$ closed, and for $D = D_1 \times D_2 \times \ldots D_m \subset \mathbb{R}^m$ with D_j closed intervals $(j = 1, 2, \ldots, m)$. Then for any $\overline{x} \in C$ at which Assumption 3.19 is satisfied one has

$$N_C(\overline{x}) = \nabla F^T(\overline{x}) N_D(F(\overline{x})) + N_S(\overline{x}). \tag{3.17}$$

If, in addition, f_0 is continuously differentiable and $\overline{x}$ is a locally optimal solution to $(\mathcal{P}_E)$ then there is a vector $\overline{y} \in N_D(F(\overline{x}))$, called a Lagrange multiplier *such that $0 \in \nabla f_0(\overline{x}) + \nabla F^T(\overline{x})\overline{y} + N_S(\overline{x})$.*

Guide. See [77, Theorems 6.14–6.15]. □

3.5 Variational principles

The Slater condition (3.14) is an *interiority* condition on the solutions to optimization problems. Interiority is just one type of *regularity* required of the solutions, wherein one is concerned with the behavior of solutions under perturbations. The next classical result lays the foundation for many modern notions of regularity of solutions.

Theorem 3.21 (Ekeland's variational principle)**.** *Let (X, d) be a complete metric space and let $f : X \to (-\infty, +\infty]$ be a lsc function bounded from below. Suppose that $\epsilon > 0$ and $z \in X$ satisfy*

$$f(z) < \inf_X f + \epsilon.$$

For a given fixed $\lambda > 0$, there exists $y \in X$ such that

(i) $d(z, y) \leq \lambda$,

(ii) $f(y) + \frac{\epsilon}{\lambda} d(z, y) \leq f(z)$, *and*

(iii) $f(x) + \frac{\epsilon}{\lambda} d(x, y) > f(y)$, *for all $x \in X \setminus \{y\}$.*

Guide. For a proof see [42]. □

An important application of Ekeland's variational principle is to the theory of subdifferentials. Given a function $f : X \to (-\infty, +\infty]$, a point $x_0 \in \operatorname{dom} f$ and $\epsilon \geq 0$, the *ϵ-subdifferential* of f at x_0 is defined by

$$\partial_\epsilon f(x_0) = \{\phi \in X^* \mid \langle \phi, x - x_0 \rangle \leq f(x) - f(x_0) + \epsilon, \ \forall x \in X\}.$$

If $x_0 \notin \operatorname{dom} f$ then by convention $\partial_\epsilon f(x_0) := \emptyset$. When $\epsilon = 0$ we have $\partial_\epsilon f(x) = \partial f(x)$. For $\epsilon > 0$ the domain of the ϵ-subdifferential coincides with $\operatorname{dom} f$ when f is a proper convex lsc function.

Theorem 3.22 (Brønsted-Rockafellar)**.** *Suppose f is a proper lsc convex function on a Banach space X. Then given any $x_0 \in \operatorname{dom} f$, $\epsilon > 0$, $\lambda > 0$ and $w_0 \in \partial_\epsilon f(x_0)$ there exist $x \in \operatorname{dom} f$ and $w \in X^*$ such that*

$$w \in \partial f(x), \quad \|x - x_0\| \leq \epsilon/\lambda \quad \text{and} \quad \|w - w_0\| \leq \lambda.$$

In particular, the domain of ∂f is dense in $\operatorname{dom} f$.

Guide. Define $g(x) := f(x) - \langle w_0, x \rangle$ on X, a proper lsc convex function with the same domain as f. Then $g(x_0) \leq \inf_X g(x) + \epsilon$. Apply Theorem 3.21 to yield a nearby point y that is the minimum of a slightly perturbed function, $g(x) + \lambda\|x - y\|$. Define the new function $h(x) := \lambda\|x - y\| - g(y)$ so that $h(x) \leq g(x)$ for all X. The sandwich theorem 2.17 establishes the existence of an affine separator $\alpha + \phi$ which is used to construct the desired element of $\partial f(x)$. □

A nice application of Ekeland's variational principle provides an elegant proof of Klee's problem in Euclidean spaces [52]: is every Čebyčev set C convex? Here a *Čebyčev set* is one with the property that every point in the space has a unique best approximation in C. A famous result is:

Theorem 3.23. *Every Čebyčev set in a Euclidean space is closed and convex.*

Guide. Since, for every finite dimensional Banach space with smooth norm, approximately convex sets are convex, it suffices to show that C is approximately convex, that is, that for every closed ball disjoint from C there is another closed ball disjoint from C of arbitrarily large radius containing the first. This follows from the mean value theorem 2.15 and Theorem 3.21 . See [19, Theorem 3.5.2]. It is not known whether the same holds for Hilbert space. □

3.6 Fixed point theory and monotone operators

Another application of Theorem 3.21 is Banach's fixed point theorem.

Theorem 3.24. *Let (X, d) be a complete metric space and let $\phi : X \to X$. Suppose there is a $\gamma \in (0,1)$ such that $d(\phi(x), \phi(y)) \le \gamma d(x,y)$ for all $x, y \in X$. Then there is a unique* fixed point *$\overline{x} \in X$ such that $\phi(\overline{x}) = \overline{x}$.*

Guide. Define $f(x) := d(x, \phi(x))$. Apply Theorem 3.21 to f with $\lambda = 1$ and $\epsilon = 1 - \gamma$. The fixed point $\overline{x}$ satisfies $f(x) + \epsilon d(x, \overline{x}) \ge f(\overline{x})$ for all $x \in X$. □

The next theorem is a celebrated result in convex analysis concerning the *maximality* of lsc proper convex functions. A monotone operator T on X is *maximal* if $\operatorname{gph} T$ cannot be enlarged in $X \times X$ without destroying the monotonicity of T.

Theorem 3.25 (maximal monotonicity of subdifferentials)**.** *Let $f : X \to (-\infty, +\infty]$ be a lsc proper convex function on a Banach space. Then ∂f is maximal monotone.*

Guide. The result was first shown by Moreau for Hilbert spaces [66, Proposition 12.b.], and shortly thereafter extended to Banach spaces by Rockafellar [72, 74]. For a modern infinite dimensional proof see [1, 19]. This result fails badly in incomplete normed spaces [19]. □

Maximal monotonicity of subdifferentials of convex functions lies at the heart of the success of algorithms as this is equivalent to *firm nonexpansiveness* of the *resolvent* of the subdifferential $(I + \partial f)^{-1}$ [63]. An operator T is *firmly nonexpansive* on a closed convex subset $C \subset X$ when

$$\|Tx - Ty\|^2 \le \langle x - y, Tx - Ty\rangle \quad \text{for all} \quad x, y \in X; \tag{3.18}$$

T is just *nonexpansive* on the closed convex subset $C \subset X$ if

$$\|Tx - Ty\| \le \|x - y\| \quad \text{for all} \quad x, y \in C. \tag{3.19}$$

Clearly, all firmly nonexpansive operators are nonexpansive. One of the most longstanding questions in geometric fixed point theory is whether a nonexpansive self-map T of a closed bounded convex subset C of a reflexive space X must have a fixed point. This is known to hold in Hilbert space.

4 Case Studies

One can now collect the dividends from the analysis outlined above for problems of the form

$$\text{(4.1)}\qquad \begin{array}{ll} \underset{x\in C\subset X}{\text{minimize}} & I_\varphi(x) \\ \text{subject to} & Ax \in D \end{array}$$

where X and Y are real Banach spaces with continuous duals X^* and Y^*, C and D are closed and convex, $A : X \to Y$ is a continuous linear operator, and the integral functional $I_\varphi(x) := \int_T \varphi(x(t))\mu(dt)$ is defined on some vector subspace $L_p(T,\mu)$ of X.

4.1 Linear inverse problems with convex constraints

Suppose X is a Hilbert space, $D = \{b\} \in \mathbb{R}^m$ and $\varphi(x) := \frac{1}{2}\|x\|^2$. To apply Fenchel duality, we rewrite (1.12) using the indicator function

$$\text{(4.2)}\qquad \begin{array}{ll} \underset{x\in X}{\text{minimize}} & \frac{1}{2}\|x\|^2 + \iota_C(x) \\ \text{subject to} & Ax = b. \end{array}$$

Note that the problem is posed on an infinite dimensional space, while the constraints (the measurements) are finite dimensional. Here we use of Fenchel duality to transform an infinite dimensional problem into a finite dimensional problem. Let $F := \{x \in C \subset E \mid Ax = b\}$ and let G denote the extensible set in E consisting of all measurement vectors b for which F is nonempty. Potter and Arun show that the existence of $\overline{y} \in \mathbb{R}^m$ such that $b = AP_CA^*\overline{y}$ is guaranteed by the constraint qualification $b \in \operatorname{ri} G$ where ri denotes the *relative interior* [70, Corollary 2]. This is a special case of Assumption 2.16, which here reduces to $b \in \int A(C)$. Though at first glance the latter condition is more restrictive, it is no real loss of generality since, if it fails, we restrict ourselves to range(A) which is closed. Then it turns out that $b \in A\operatorname{qri} C$, the image of the *quasi-relative interior* of C [15, Exercise 4.1.20]. Assuming this holds, Fenchel duality, Theorem 3.10, yields the dual problem

$$\text{(4.3)}\qquad \sup_{y\in\mathbb{R}^m} \langle b,\ y\rangle - \left(\tfrac{1}{2}\|\cdot\|^2 + \iota_C\right)^* (A^*y)$$

whose value is equivalent to the value of the primal problem. This is a finite dimensional unconstrained convex optimization problem whose solution is characterized by the inclusion (Theorem 2.13)

$$\text{(4.4)}\qquad 0 \in \partial\left(\tfrac{1}{2}\|\cdot\|^2 + \iota_C\right)^* (A^*y) - b.$$

Now from Lemma 3.7, Example 3.1(ii)-(iii) and (3.2),

$$\left(\tfrac{1}{2}\|\cdot\|^2 + \iota_C\right)^* (x) = (\sigma_C \Box \tfrac{1}{2}\|\cdot\|)(x) = \inf_{z\in X} \sigma_C(z) + \tfrac{1}{2}\|x - z\|^2.$$

The argmin of the Yosida approximation above (see Theorem 3.8) is the proximal operator (3.2). Applying the sum rule for differentials, Theorem 2.19 and Proposition 3.5 yields

$$\text{(4.5)}\qquad \operatorname{prox}_{1,\sigma_C}(x) = \operatorname{argmin}_{z\in X}\left\{\sigma_C(z) + \tfrac{1}{2}\|z - x\|^2\right\} = x - P_C(x)$$

where P_C is the orthogonal projection onto the set C. This together with (4.4) yields the optimal solution $\overline{y}$ to (4.3):

$$b = AP_C(A^*\overline{y}). \tag{4.6}$$

Note that the existence of a solution to (4.6) is guaranteed by Assumption 2.16. This yields the solution to the primal problem as $\overline{x} = P_C(A^*\overline{y})$.

With the help of (4.5), the iteration proposed in [70] can be seen as a subgradient descent algorithm for solving

$$\inf_{y\in\mathbb{R}^m} h(y) := \sigma_C(A^*y - P_C(A^*y)) + \tfrac{1}{2}\|P_C(A^*y)\|^2 - \langle b,\ y\rangle\,.$$

The proposed algorithm is, given $y_0 \in \mathbb{R}^m$, generate the sequence $\{y_n\}_{n=0}^{\infty}$ by

$$y_{n+1} = y_n - \lambda\partial h(y_n) = y_n + \lambda\left(b - AP_CA^*y_n\right).$$

For convergence results of this algorithm in a much larger context see [34].

4.2 Imaging with missing data

This application is formally simpler than the previous example since there is no abstract constraint set. As discussed in subsection 1.2 we consider relaxations to the conventional problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & I_{\varphi^*_{\epsilon,L}}(x) \\ \text{subject to} & Ax = b. \end{array} \tag{4.7}$$

where

$$\varphi^*_{\epsilon,L}(x) = \frac{\epsilon}{\ln(2)}\ln\left(4^{xL/\epsilon} + 1\right) - xL - \epsilon. \tag{4.8}$$

Using Fenchel duality, the dual to this problem is the concave optimization problem

$$\sup_{y\in\mathbb{R}^m} y^T b - I_{\varphi_{\epsilon,L}}(A^*y)$$

where

$$\begin{aligned} \varphi_{\epsilon,L}(x) \quad := \quad & \epsilon\left(\frac{(L+x)\ln(L+x) + (L-x)\ln(L-x)}{2L\ln(2)} - \frac{\ln(L)}{\ln(2)}\right) \\ & L, \epsilon > 0\, x \in [-L, L]. \end{aligned}$$

If there exists a point $\overline{y}$ satisfying $b = AA^*\overline{y}$, then the optimal value in the dual problem is attained and the primal solution is given by $A^*\overline{y}$. The objective in the dual problem is smooth and convex, so we could apply any number of efficient unconstrained optimization algorithms. Also, for this relaxation, the same numerical techniques can be used for all $L \to 0$.

4.3 Inverse scattering

Theorem 4.1. *Let X, Y be reflexive Banach spaces with duals X^* and Y^*. Let $F: Y^* \to Y$ and $G: X \to Y$ be bounded linear operators with $F = GSG^*$ for $S: X^* \to X$ a bounded linear operator satisfying the coercivity condition*

$$|\langle \varphi,\ S\varphi\rangle| \geq c\|\varphi\|^2_{X^*} \quad \text{for some } c > 0 \text{ and all } \varphi \in \operatorname{range}(G^*) \subset X^*.$$

Define $h(\psi): Y^* \to (-\infty, +\infty] := |\langle \psi, F\psi\rangle|$, and let h^* denote the Fenchel conjugate of h. Then $\operatorname{range}(G) = \operatorname{dom} h^*$.

Proof. Following [51, Theorem 1.16], we show that $h^*(\phi) = \infty$ for $\phi \notin \operatorname{range}(G)$. To do this we work with a dense subset of range G: $G^*(C)$ for $C := \{\psi \in Y^* \mid \langle \psi,\ \phi\rangle = 0\}$. It was shown in [51, Theorem 1.16] that $G^*(C)$ is dense in range(G).

Now by the Hahn-Banach theorem 2.10 there is a $\widehat{\phi} \in Y^*$ such that $\left\langle \widehat{\phi},\ \phi \right\rangle = 1$. Since $G^*(C)$ is dense in range(G^*) there is a sequence $\{\psi_n\}_{n=1}^{\infty} \subset C$ with

$$G^*\psi_n \to -G^*\widehat{\phi}, \quad n \to \infty.$$

Now set $\psi_n := \widehat{\psi}_n + \widehat{\phi}$. Then $\langle \phi,\ \alpha\psi_n\rangle = \alpha$ and $G^*(\alpha\psi_n) = \alpha G^*\psi_n \to 0$ for any $\alpha \in \mathbb{R}$. Using the factorization of F we have

$$|\langle \psi_n,\ F\psi_n\rangle| = |\langle G^*\psi_n,\ SG^*\psi_n\rangle| \leq \|S\|\|G^*\psi_n\|^2_{X^*}$$

hence $\alpha^2 \langle \psi_n,\ F\psi_n\rangle \to 0$ as $n \to \infty$ for all α, but $\langle \phi,\ \alpha\psi_n\rangle = \alpha$, that is, $\langle \phi, \alpha\psi_n\rangle - h(\alpha\psi_n) \to \alpha$ and $h^*(\phi) = \infty$. □

In the scattering application, we have a scatterer supported on a domain $D \subset \mathbb{R}^m$ ($m = 2$ or 3) that is illuminated by an incident field. The Helmholtz equation models the behavior of the fields on the exterior of the domain and the boundary data belongs to $X = H^{1/2}(\Gamma)$. On the sphere at infinity the leading-order behavior of the fields, the so-called *far field pattern*, lies in $Y = L^2(\mathbb{S})$. The operator mapping the boundary condition to the far field pattern – the *data-to-pattern operator* – is $G: H^{1/2}(\Gamma) \to L^2(\mathbb{S})$. Assume that the *far field operator* $F: L^2(\mathbb{S}) \to L^2(\mathbb{S})$ has the factorization $F = GS^*G^*$, where $S: H^{-1/2}(\Gamma) \to H^{1/2}(\Gamma)$ is a *single layer boundary operator* defined by

$$(S\varphi)(x) := \int_\Gamma \Phi(x,y)\varphi(y)ds(y),\ x \in \Gamma,$$

for $\Phi(x,y)$ the fundamental solution to the Helmholtz equation. With a few results about the denseness of G and the coercivity of S, which, though standard, we will not go into here, we have the following application to inverse scattering.

Corollary 4.2 (Application to Inverse Scattering)**.** *Let $D \subset \mathbb{R}^m$ ($m = 2$ or 3) be an open bounded domain with connected exterior and boundary Γ. Let $G: H^{1/2}(\Gamma) \to L^2(\mathbb{S})$, be the* data-to-pattern *operator, $S: H^{-1/2}(\Gamma) \to H^{1/2}(\Gamma)$, the* single layer boundary operator *and let the* far field pattern *$F: L^2(\mathbb{S}) \to L^2(\mathbb{S})$ have the factorization $F = GS^*G^*$. Assume k^2 is not a Dirichlet eigenvalue of $-\triangle$ on D. Then* range $G = \operatorname{dom} h^*$ *where $h(\psi): L^2(\mathbb{S}) \to (-\infty, +\infty] := |\langle \psi, F\psi\rangle|$.*

4.4 Fredholm integral equations

We showed in the introduction the failure of Fenchel duality for Fredholm integral equations. Here we briefly sketch a result on regularizations, or relaxations, that recovers duality relationships. The result will show that by introducing a relaxation, we can recover the solution to ill-posed integral equations as the norm limit of solutions computable from a dual problem of maximum entropy type.

Theorem 4.3 (Theorem 3.1 of [12])**.** *Let $X = L_1(T, \mu)$ on a complete measure finite measure space and let $(Y, \|\cdot\|)$ be a normed space. The infimum $\inf_{x\in X} \{I_\varphi(x) \mid Ax = b\}$ is attained when finite. In the case where it is finite, consider the relaxed problem for $\epsilon > 0$*

$$(\mathcal{P}^{\epsilon}_{MEP}) \qquad \begin{array}{ll} \underset{x\in X}{\textit{minimize}} & I_\varphi(x) \\ \textit{subject to} & \|Ax - b\| \leq \epsilon. \end{array}$$

Let p_ϵ denote the value of $(\mathcal{P}^{\epsilon}_{MEP})$. The value of p_ϵ equals d_ϵ, the value of the dual problem

$$(\mathcal{P}^{\epsilon}_{DEP}) \qquad \underset{y^*\in Y^*}{\textit{maximize}}\ \langle b,\ y^*\rangle - \epsilon\|y^*\|_* - I_{\varphi^*}(A^*y^*),$$

and the unique optimal solution of $(\mathcal{P}^{\epsilon}_{MEP})$ is given by

$$\overline{x}_{\varphi,\epsilon} := \frac{\partial\varphi^*}{\partial r}(A^*y^*_\epsilon)$$

where y^*_ϵ is any solution to $(\mathcal{P}^{\epsilon}_{DEP})$. Moreover, as $\epsilon \to 0^+$, $\overline{x}_{\varphi,\epsilon}$ converges in mean to the unique solution of $(\mathcal{P}^{0}_{MEP})$ and $p_\epsilon \to p_0$.

Guide. Attainment of the infimum in $\inf_{x\in X} \{I_\varphi(x) \mid Ax = b\}$ follows from *strong convexity* of I_φ [14, 82]: strictly convex with weakly-compact lower level sets and with the *Kadec property*, i.e. that weak convergence together with convergence of the function values implies norm convergence. Let $g(y) := \iota_S(y)$ for $S = \{y \in Y \mid b \in y + \epsilon B_Y\}$ and rewrite $(\mathcal{P}^{\epsilon}_{MEP})$ as $\inf\{I_\varphi(x) + g(Ax) \mid x \in X\}$. An elementary calculation shows that the Fenchel dual to $(\mathcal{P}^{\epsilon}_{MEP})$ is $(\mathcal{P}^{\epsilon}_{DEP})$. The relaxed problem $(\mathcal{P}^{\epsilon}_{MEP})$ has a constraint for which a Slater-type constraint qualification holds at any feasible point for the unrelaxed problem. The value d_ϵ is thus attained and equal to p_ϵ. Subgradient arguments following [13] show that $\overline{x}_{\varphi,\epsilon}$ is feasible for $(\mathcal{P}^{\epsilon}_{MEP})$ and is the unique solution to $(\mathcal{P}^{\epsilon}_{MEP})$. Convergence follows from weak compactness of the lower level set $L(p_0) := \{x \mid I_\varphi(x) \leq p_0\}$, which contains the sequence $(\overline{x}_{\varphi,\epsilon})_{\epsilon>0}$. Weak convergence of $\overline{x}_{\varphi,\epsilon}$ to the unique solution to the unrelaxed problem follows from strict convexity of I_φ. Convergence of the function values and strong convexity of I_φ then yields norm convergence. □

Notice that the dual in Theorem 4.3 is unconstrained and easier to compute with, especially when there are finitely many constraints. This theorem remains valid for objectives of the form $I_\varphi(x) + \langle x*,\ x\rangle$ for $x*$ in $L_\infty(T)$. This enables one to apply them to many *Bregman distances* – that is, integrands of the form $\phi(x) - \phi(x_0) - \langle\phi'(x_0),\ x - x_0\rangle$ where ϕ, is closed and convex on $\mathbb{R}$.

5 Open Questions

Regrettably, due to space constraints, we have omitted fixed point theory and many facts about monotone operators that are useful in proving convergence of algorithms. However, it is worthwhile noting two long-standing problems that impinge on fixed point and monotone operator theory.

(i) Klee's problem: is every Čebyčev set C in a Hilbert space convex?

(ii) Must a nonexpansive self-map T of a closed bounded convex subset C of a reflexive space X have a fixed point?

6 Conclusion

Duality and convex programming provides powerful techniques for solving a wide range of imaging problems. While frequently a means toward computational ends, the dual perspective can also yield new insight into image processing problems and the information content of data implicit in certain models. Five main applications illustrate the convex analytical approach to problem solving and the use of duality: linear inverse problems with convex constraints, compressive imaging, image denoising and deconvolution, nonlinear inverse scattering, and finally Fredholm integral equations. These are certainly not exhaustive, but serve as good templates. The Hahn-Banach/Fenchel duality cycle of ideas developed here not only provides a variety of entry points into convex and variational analysis, but also underscores duality in convex optimization as both a theoretical phenomenon and an algorithmic strategy.

As readers of this volume will recognize, not all problems of interest are convex. But just as nonlinear problems are approached numerically by sequences of linear approximations, nonconvex problems can be approached by sequences of convex approximations. Convexity is the central organizing principle and has tremendous algorithmic implications, including not only computable guarantees about solutions, but efficient means towards that end. In particular, convexity implies the existence of implementable, polynomial-time, algorithms. This chapter is meant to be a foundation for more sophisticated methodologies applied to more complicated problems.

7 Cross References

Readers of the present chapter will find the following chapters of particular interest: Compressive Sensing, Inverse Scattering, Iterative Solution Methods, Numerical Methods for Variational Approach in Image Analysis, Regularization Methods for Ill-Posed Problems, Total Variation in Imaging, Variational Approach in Image Analysis, and Variational methods and Shape Spaces.

8 Recommended Reading

References

[1] M. Alves and B. F. Svaiter. A new proof for maximal monotonicity of subdifferential operators. *J. Convex Analysis*, 15(2):345–348, 2008.

[2] G. Aubert and P. Kornprost. *Mathematical Problems Image Processing in*, volume 147 of *Applied Mathematical Sciences*. Springer, New York, 2002.

[3] A. Auslender and M. Teboulle. *Asymptotic Cones and Functions in Optimization and Variational Inequalities*. Springer, New York, 2003.

[4] H. H. Bauschke and J. M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Rev.*, 38(3):367–426, 1996.

[5] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer-Verlag, New York, to appear.

[6] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Phase retrieval, error reduction algorithm and Fienup variants: a view from convex feasibility. *J. Opt. Soc. Amer. A.*, 19(7):1334–45, 2002.

[7] H. H. Bauschke, P. L. Combettes, and D. R. Luke. A hybrid projection reflection method for phase retrieval. *J. Opt. Soc. Amer. A.*, 20(6):1025–34, 2003.

[8] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Finding best approximation pairs relative to two closed convex sets in Hilbert spaces. *J. Approx. Theory*, 127:178–92, 2004.

[9] J. Bect, L. Blanc-Féraud G. Aubert, and A. Chambolle. A ℓ_1-unified variational framework for image restoration. In T. Pajdla and J. Matas, editors, *Proc. Eighth Europ. Conf. Comput. Vision, Prague, 2004*, volume 3024 of *Lecture Notes in Comput. Sci.*, pages 1–13, New York, 2004. Springer-Verlag.

[10] A. Ben-Tal, J. M. Borwein, and M. Teboulle. A dual approach to multidimensional l_p spectral estimation problems. *SIAM J. Control Optim.*, 26:985–996, 1988.

[11] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal. *Numerical Optimization*. Springer, New York, 2nd edition, 2006.

[12] J. M. Borwein. On the failure of maximum entropy reconstruction for Fredholm equations and other infinite systems. *Math. Program.*, 61:251–261, 1993.

[13] J. M. Borwein and A. S. Lewis. Duality relationships for entropy-like minimization problems. *SIAM J. Contr. Optim.*, 29:325–338, 1990.

[14] J. M. Borwein and A. S. Lewis. Convergence of best entropy estimates. *SIAM J. Optim.*, 1:191–205, 1991.

[15] J. M. Borwein and A. S. Lewis. *Convex analysis and nonlinear optimization : theory and examples*. Springer Verlag, New York, 2nd edition, 2006.

[16] J. M. Borwein and Q. J. Zhu. *Techniques of Variational Analysis*. CMS Books in Mathematics. Springer, New York, 2005.

[17] J.M. Borwein, A.S. Lewis, M. N. Limber, and D. Noll. Maximum entropy spectral analysis using first order information. part 2: A numerical algorithm for fisher information duality. *Numerische Mathematik*, 69:243–256, 1995.

[18] J.M. Borwein, A.S. Lewis, and D. Noll. Maximum entropy spectral analysis using first order information. part 1: Fisher information and convex duality. *Mathematics of Operations Research*, 21:442–468, 1996.

[19] J.M. Borwein and Jon Vanderwerff. *Convex Functions: Constructions, Characterizations and Counterexamples*, volume 109 of *Encyclopedias in Mathematics*. Cambridge University Press, New York, 2009.

[20] S. Boyd and L. Vandenberghe. *Convex Optimization*. Oxford University Press, New York, 2003.

[21] O. A. Brezhneva, A. A. Tret'yakov, and S. E. Wright. A simple and elementary proof of the KarushKuhnTucker theorem for inequality-constrained optimization. *Optim Lett.*, 3:7–10, 2009.

[22] J. P. Burg. Maximum entropy spectral analysis. paper presented at The 37th Meeting of the Society of Exploration Geophysicists, Oklahoma City, 1967, 1967.

[23] J. V. Burke and D. R. Luke. Variational analysis applied to the problem of optical phase retrieval. *SIAM J. Contr. Optim.*, 42(2):576–595, 2003.

[24] C. L. Byrne. *Signal Processing: a mathematical approach*. A K Peters Ltd, Natick, MA, 2005.

[25] E. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, December 2006.

[26] Y. Censor and S. A. Zenios. *Parallel Optimization: Theory Algorithms and Applications*. Oxford University Press, 1997.

[27] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging and Vision*, 20:89–97, 2004.

[28] A. Chambolle and P. L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76:167–188, 1997.

[29] T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comput.*, 20(6):1964–1977, 1999.

[30] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1999.

[31] F. H. Clarke. *Optimization and Nonsmooth Analysis*, volume 5 of *Classics in Applied Mathematics*. SIAM, 1990.

[32] F. H. Clarke, R. J. Stern, Yu. S. Ledyaev, and P. R. Wolenski. *Nonsmooth Analysis and Control Theory*. Springer Verlag, 1998.

[33] P. L. Combettes. The convex feasibility problem in image recovery. In P. W. Hawkes, editor, *Advances in Imaging and Electron Physics*, volume 95, pages 155–270. Academic Press, New York, 1996.

[34] P. L. Combettes, D. Dũng, and B. C. Vũ. Dualization of signal recovery problems. Technical report, arXiv:0907.0436v2 [math.OC], 2009.

[35] P. L. Combettes and H. J. Trussell. Method of successive projections for finding a common point of sets in metric spaces. *J. Opt. Theory and Appl.*, 67(3):487–507, 1990.

[36] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM J. Multiscale Model. Simul.*, 4(4):1168–1200, 2005.

[37] D. Dacunha-Castelle and F. Gamboa. Maximum d'entropie et probléme des moments. *l'Institut Henri Poincaré*, 26:567–596, 1990.

[38] P. Destuynder, M. Jaoua, and H. Sellami. A dual algorithm for denoising and preserving edges in image processing. *J. Inverse Ill-Posed Probl.*, 15:149–165, 2007.

[39] F. Deutsch. *Best Approximation in Inner Product Spaces*. CMS Books in Mathematics. Springer, New York, 2001.

[40] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[41] P. P. B. Eggermont. Maximum entropy regularization for Fredholm integral equations of the first kind. *SIAM J. Math. Anal.*, 24(6):1557–1576, 1993.

[42] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*. Elsevier, New York, 1976.

[43] W. Fenchel. On conjugate convex functions. *Canadian J. Math.*, 1:7377, 1949.

[44] R. K. Goodrich and A. Steinhardt. L2 spectral estimation. *SIAM J. Appl. Math.*, 46:417–428, 1986.

[45] C. W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Integral Equations of the First Kind*. Pitman, Bostan, 1984.

[46] C. W. Groetsch. *Stable Approximate Evaluation of Unbounded Operators*, volume 1894 of *Lecture Notes in Mathematics*. Springer-Verlag, New York, 2007.

[47] M. Hintermüller and G. Stadler. An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration. *SIAM J. Sci. Comput.*, 28:1–23, 2006.

[48] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms, I & II*, volume 305–306 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, New York, 1993.

[49] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, New York, 2001.

[50] A. N. Iusem and M. Teboulle. A regularized dual-based iterative method for a class of image reconstruction problems. *Inverse Problems*, 9:679–696, 1993.

[51] A. Kirsch and N. Grinberg. *The Factorization Method for Inverse Problems*. Number 36 in Oxford Lecture Series in Mathematics and its Applications. Oxford University Press, New York, 2008.

[52] V. Klee. Convexity of Cebysev sets. *Math. Annalen*, 142:291–304, 1961.

[53] R. Kress. *Linear Integral Equations*, volume 82 of *Applied Mathematical Sciences*. Springer Verlag, New York, 2 edition, 1999.

[54] L. Levi. Fitting a bandlimited signal to given points. *IEEE Trans. Inform. Theory*, 11:372–376, 1965.

[55] A. S. Lewis, D. R. Luke, and J. Malick. Local linear convergence of alternating and averaged projections. *Found. Comput. Math.*, 9(4):485–513, 2009.

[56] A. S. Lewis and J. Malick. Alternating projections on manifolds. *Math. Oper. Res.*, 33:216–234, 2008.

[57] R. Lucchetti. *Convexity and Well-Posed Problems*, volume 22 of *CMS Books in Mathematics*. Springer-Verlag, New York, 2006.

[58] D. G. Luenberger and Y. Ye. *Linear and Nonlinear Programming*. Springer, New York, 3rd edition, 2008.

[59] D. R. Luke. Relaxed averaged alternating reflections for diffraction imaging. *Inverse Problems*, 21:37–50, 2005.

[60] D. R. Luke. Finding best approximation pairs relative to a convex and a prox-regular set in Hilbert space. *SIAM J. Optim.*, 19(2):714–739, 2008.

[61] D. R. Luke, J. V. Burke, and R. G. Lyon. Optical wavefront reconstruction: Theory and numerical methods. *SIAM Rev.*, 44:169–224, 2002.

[62] P. Maréchal and A. Lannes. Unification of some deterministic and probabilistic methods for the solution of inverse problems via the principle of maximum entropy on the mean. *Inverse Problems*, 13:135–151, 1997.

[63] G. J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke. Math. J.*, 29(3):341–346, 1962.

[64] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation, I: Basic Theory; II: Applications.* Grundlehren der mathematischen Wissenschaften. Springer-Verlag, New York, 2006.

[65] J. J. Moreau. Fonctions convexes duales et points proximaux dans un espace Hilbertien. *Comptes Rendus de l'Académie des Sciences de Paris*, 255:2897–2899, 1962.

[66] J. J. Moreau. Proximité et dualité dans un espace Hilbertian. *Bull. de la Soc. math. de France*, 93(3):273–299, 1965.

[67] Y. E. Nesterov and A. S. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming.* SIAM Publications, Philadelphia, 1994.

[68] J. Nocedal and S. Wright. *Numerical Optimization.* Springer Verlag, New York, 2000.

[69] R. R. Phelps. *Convex Functions, Monotone Operators and Differentiability*, volume 1364 of *Lecture Notes in Mathematics.* Springer -Verlag, New York, 2nd edition, 1993.

[70] L. C. Potter and K. S. Arun. A dual approach to linear inverse problems with convex constraints. *SIAM J. Contr. Opt.*, 31(4):1080–1092, 1993.

[71] B. N. Pshenichnyi. *Necessary Conditions for an Extremum*, volume 4 of *Pure and Applied Mathematics.* Marcel Dekker Inc., New York, 1971. Translated from Russian by Karol Makowski. Translation edited by Lucien W. Neustadt.

[72] R. T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific J. Math.*, 17:497–510, 1966.

[73] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, 1970.

[74] R. T. Rockafellar. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.*, 33:209–216, 1970.

[75] R. T. Rockafellar. Integrals which are convex functionals, II. *Pacific J. Math.*, 39:439–469, 1971.

[76] R. T. Rockafellar. *Conjugate Duality and Optimization.* SIAM, Philadelphia, 1974.

[77] R. T. Rockafellar and R. J. Wets. *Variational Analysis.* Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin, 1998.

[78] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[79] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences.* Springer, New York, 2009.

[80] S. Simons. *From Hahn-Banach to Monotonicity*, volume 1693 of *Lecture Notes in Mathematics.* Springer-Verlag, New York, 2008.

[81] I. Singer. *Duality for Nonconvex Approximation and Optimization.* Springer, 2006.

[82] M. Teboulle and I. Vajda. Convergence of best φ-entropy estimates. *IEEE Trans. Inform. Process.*, 39:279–301, 1993.

[83] A. N. Tihonov. On the regularization of ill-posed problems. (russian). *Dokl. Akad. Nauk SSSR*, 153:49–52, 1963.

[84] P. Weiss, G. Aubert, and L. Blanc-Féraud. Efficient schemes for total variation minimization under constraints in image processing. *SIAM J. Sci. Comput.*, 31:2047–2080, 2009.

[85] S. J. Wright. *Primal-Dual Interior-Point Methods*. SIAM, Philadelphia, 1997.

[86] E. H. Zarantonello. Projections on convex sets in Hilbert space and spectral theory. In E. H. Zarantonello, editor, *Contributions to Nonlinear Functional Analysis*, pages 237–424. Academic Press, New York, 1971.

[87] C. Zălinescu. *Convex Analysis in General Vector Spaces*. World Scientific, New Jersey, 2002.