JOTA manuscript No.
(will be inserted by the editor)

A Variational Approach to Lagrange Multipliers

Jonathan M. Borwein · Qiji J. Zhu

Received: date / Accepted: date

Abstract We discuss Lagrange multiplier rules from a variational perspective. This allows us to highlight many of the issues involved and also to illustrate how broadly an abstract version can be applied.

Keywords Lagrange multiplier · Variational method · Convex duality · Constrained optimization · Nonsmooth analysis

Mathematics Subject Classification (2000) 90C25 · 90C46 · 49N15

1 Introduction

The Lagrange multiplier method is fundamental in dealing with constrained optimization problems and is also related to many other important results. There are many different routes to reaching the fundamental result. The variational approach used in [1] provides a deep understanding of the nature of the Lagrange multiplier rule and is the focus of this survey.

David Gale's seminal paper [2] provides a penetrating explanation of the economic meaning of the Lagrange multiplier in the convex case. Consider maximizing the output of an economy with resource constraints. Then the optimal output is a function of the level of resources. It turns out the derivative of this function, if exists, is exactly the Lagrange multiplier for the constrained optimization problem. A Lagrange multiplier, then, reflects the *marginal gain* of the output function with respect to the vector of resource constraints. Following this observation, if we penalize the resource utilization with a (vector) Lagrange multiplier then the constrained optimization problem can be converted to an unconstrained one. One cannot emphasize enough the importance of this insight.

In general, however, an optimal value function for a constrained optimization problem is neither convex nor smooth. This explains why this view was not prevalent before the systematic

The reseach is partially supported by an Australian Research Council Discovery Projects grant.

Jonathan M. Borwein
Centre for Computer-assisted Research Mathematics and its Applications (CARMA), School of Mathematical and Physical Sciences, University of Newcastle, Callaghan, NSW 2308, Australia.
Tel.: +02-49-215535, Fax: +02-49-216898, E-mail: jonathan.borwein@newcastle.edu.au

Qiji J. Zhu, corresponding author
Department of Mathematics, Western Michigan University, Kalamazoo, MI 49008, USA.
Tel.: +01-269-3874535, Fax: +01-269-3874530, E-mail: zhu@wmich.edu

development of nonsmooth and variational analysis. This systematic development of convex and nonsmooth analysis during the 1950s through 1970s, respectively, provided suitable tools for the proper analysis of Lagrange multipliers. Gale himself provided a rigorous proof of the fact that for well behaved convex problems *the subdifferential of the optimal value function exactly characterizes the set of all Lagrange multipliers*. Subsequently, many researchers have derived versions of Lagrange multiplier theorems with different ranges of applicability using other generalized derivative concepts (see, e.g., [1]).

While rigorous justification of the variational view of the Lagrange multiplier only appeared in the 1970s, the basic idea can be traced back to early developments of the calculus of variations and is associated with the names of Euler, Hamilton, Lagrange, Legendre and many others (see [3–5]). Besides the explicit use of a Lagrange multiplier in calculus of variations problems involving isoperimetric or similar constraints, also very influential are the ideas of (i) imbedding an optimization problem in a related class of problems, of (ii) using optimal value functions and of (iii) *decoupling*. Using this team in combination with the so called *principle of least action*[1] in Newtonian or quantum mechanics is a very powerful classical approach (see, e.g., [6]).

Despite an extensive literature on various Lagrange multiplier rules, we now list several finer points which are, in our opinion, still worthy of further attention.

1. First, Lagrange multipliers are intrinsically related to the derivative or to derivative-like properties of the optimal value function. This is already well explained from the economic explanation of the Lagrange multiplier rule in David Gale's paper [2]. Gale's paper focuses on the convex case but the essential relationship extends to Lagrange multiplier rules that rely on other generalized derivatives.
2. Second, in a Lagrange multiplier rule a *complementary slackness* condition holds when the optimal solution exists. Nonetheless, without the a priori existence of an optimal solution, a Lagrange multiplier rule involving only the optimal value function still holds and is often useful (see, e.g., [7, 8] and below).
3. Third, the form of a Lagrange multiplier rule is dictated by the properties of the *optimal value function* and by the choice of generalized derivative. In many developments, sufficient conditions for ensuring the existence of such generalized derivatives are not always clearly disentangled from what was necessary to derive the Lagrange multiplier rule itself.
4. Finally, computing Lagrange multipliers often relies on *decoupling* information in terms of each individual constraint. Sufficient conditions are often needed for this purpose and they are also not always clearly limned.

The goal of this relatively short survey is to provide a reasonably concise discussion of the variational approach to Lagrange multiplier rules. Our goal is to illustrate the more subtle points alluded to above, rather than to be comprehensive. For this purpose, we shall showcase two versions of the Lagrange multiplier rule: *global* using the (convex) *subdifferential* and *local* using rather the *Fréchet subdifferential*. The (convex) subdifferential provides a complete characterization of all the Lagrange multipliers. It also naturally relates to convex duality theory and allows the luxury of studying the underlying constrained optimization problem by way of its companion dual problem. The limitation of this version of Lagrange multiplier rule is that the general sufficient condition (note that it is not necessary) to ensure its applicability is *convexity of both the cost and the constraint functions*—which is a relatively restrictive condition.

By contrast, the Fréchet subdifferential belongs to the class of viscosity subdifferentials [9] whose existence requires only quite minimal conditions. These subdifferentials are also known to

[1] Explicitly named only in the last century by Feynman and others, the principle states that the path taken in a mechanical system will be the one which is stationary with respect to the *action* (which of course must be specified) [3, 5].

approximate various other generalized derivative concepts [10]. Lagrange multiplier rules in terms of the Fréchet subdifferential provide very natural ways of capturing local solutions of constrained optimization problems (or at least of finding necessary conditions and critical points).

To focus adequately on the variational approach — and yet still to be brief — we have opted to leave out many other perspectives such as the geometric derivation often used in calculus textbooks. This views the constrained optimization problem as an optimization problem on the manifold defined by the constraints [11]. Another influential perspective is to treat the Lagrange multiplier rule as a contraposition of the sufficient condition for *open covering* [12]. One can also take the related view of *image space analysis* as discussed in detail in the recent monograph [13]. For detailed discussions on various different perspectives in finite dimensions we recommend Rockafellar's 1993 comprehensive survey [14]. In infinite dimensions we direct the reader also to the work of Tikhomirov and his collaborators [15–17].

The remainder of the paper is organized as follows. In Section 2.1 we discuss the general issue. Then in Section 2.2 we turn to the convex case. In Section 2.3 we likewise consider local multiplier rules, while in Section 3 we relate the convex theory to more general convex duality theory. Then in Section 4 we collect up some other less standard examples. We end in Section 5 with a few further observations.

2 Lagrange Multiplier Rules

To understand the variational approach to Lagrange multiplier rules heuristically, we take real functions f and $g := (g_1, g_2, \ldots g_N)$, and consider the simple constrained optimization problem $v(y) := \min\{f(x) : g(x) = y\}$. If x_y is a (local) solution to this problem. Then the function $x \mapsto f(x) - v(g(x))$ attains an (unconstrained) minimum at x_y. Assuming all functions involved are smooth, we then have

$$f'(x_y) - Dv(g(x_y))g'(x_y) = 0,$$

revealing $\lambda_y = -Dv(g(x_y))$ as a Lagrange multiplier. This gives us an idea as to why we might expect a Lagrange multiplier to exist. Much of our discussion below is about how to make this heuristic rigorous especially when v is not differentiable. Let us set the stage:

2.1 A General Overview

Let X, Y and Z be Banach spaces, and let $\leq_K$ be the linear partial order in Y induced by a closed, nonempty, convex cone K in Y: $y_1 \leq_K y_2$ iff $y_2 - y_1 \in K$. We denote the *polar cone* of K by $K^+ := \{y^* \in Y^* : \langle y^*, y\rangle \geq 0, \forall y \in K\}$. Consider the following class of constrained optimization problems, for $(y, z) \in Y \times Z$,

$$P(y,z): \quad \min f(x) \quad \text{s.t.} \quad g(x) \leq_K y, h(x) = z, x \in C, \tag{1}$$

where C is a closed subset of X, $f : X \to \mathbb{R}$ is lower semicontinuous, $g : X \to Y$ is lower semicontinuous with respect to $\leq_K$ and $h : X \to Z$ is continuous. We shall use
$v(y,z) := \inf\{f(x)\colon f(x) \leq_K y, h(x) = z, x \in C\}$ to represent the *optimal value* function, which may take values $\pm\infty$ (in infeasible or unbounded below cases), and $S(y, z)$ the (possibly empty) solution set of problem $P(y, z)$. In general, when not given explicitly our terminology is consistent with that in [1,18] and [19].

2.2 Use of the (Convex) Subdifferential

Recall the following definition.

Definition 2.1 (Subdifferential) The subdifferential of a lower semicontinuous function ϕ at $x \in \operatorname{dom} \phi$ is defined by

$$\partial\phi(x) := \{x^* \in X^* : \phi(y) - \phi(x) \geq \langle x^*, y - x\rangle, \forall y \in X\}.$$

This globally-defined subdifferential is introduced as a replacement for the possibly nonexistent derivative of a convex function. It has many applications. While it arises naturally for convex functions the definition works equally well, at least formally, for nonconvex functions. As we shall see from the two versions of Lagrange multiplier rules given below, the subdifferential of the optimal value function completely characterizes the set of Lagrange multipliers (denoted λ in these theorems).

Theorem 2.1 (Lagrange Multiplier without Existence of Optimal Solution) *Let $v(y, z)$ be the optimal value function of the constrained optimization problem $P(y, z)$. Then $-\lambda \in \partial v(0, 0)$ if and only if*

(i) (non-negativity) $\lambda \in K^+ \times Z^*$; and
(ii) (unconstrained optimum) for any $x \in C$,

$$f(x) + \langle\lambda, (g(x), h(x))\rangle \geq v(0, 0).$$

Proof. (a) The "only if" part. Suppose that $-\lambda \in \partial v(0, 0)$. It is easy to see that $v(y, 0)$ is non-increasing with respect to the partial order $\leq_K$. Thus, for any $y \in K$,

$$0 \geq v(y, 0) - v(0, 0) \geq \langle -\lambda, (y, 0)\rangle$$

so that $\lambda \in K^+ \times Z^*$. Conclusion (ii) follows from, the fact that for all $x \in C$,

$$f(x) + \langle\lambda, (g(x), h(x))\rangle \geq v(g(x), h(x)) + \langle\lambda, (g(x), h(x))\rangle \geq v(0, 0). \tag{2}$$

(b) The "if" part. Suppose λ satisfies conditions (i) and (ii). Then we have, for any $x \in C$, $g(x) \leq_K y$ and $h(x) = z$,

$$f(x) + \langle\lambda, (y, z)\rangle \geq f(x) + \langle\lambda, (g(x), h(x))\rangle \geq v(0, 0). \tag{3}$$

Taking the infimum of the leftmost term under the constraints $x \in C$, $g(x) \leq_K y$ and $h(x) = z$, we arrive at

$$v(y, z) + \langle\lambda, (y, z)\rangle \geq v(0, 0). \tag{4}$$

Therefore, $-\lambda \in \partial v(0, 0)$. □

If we denote by $\Lambda(y, z)$ the multipliers satisfying (i) and (ii) of Theorem 2.1 then we may write the useful set equality

$$\Lambda(0, 0) = -\partial v(0, 0).$$

The next corollary is now immediate. It is often a useful variant since h may well be affine.

Corollary 2.1 (Lagrange Multiplier without Existence of Optimal Solution) *Let $v(y, z)$ be the optimal value function of the constrained optimization problem $P(y, z)$. Then $-\lambda \in \partial v(0, 0)$ if and only if*

(i) (non-negativity) $\lambda \in K^+ \times Z^*$; and

(*ii*) *(unconstrained optimum) for any $x \in C$, satisfying $g(x) \leq_K y$ and $h(x) = z$,*

$$f(x) + \langle \lambda, (y, z) \rangle \geq v(0, 0).$$

When an optimal solution for the problem $P(0, 0)$ exists, we can also derive a so called *complementary slackness* condition.

Theorem 2.2 (Lagrange Multiplier when Optimal Solution Exists) *Let $v(y, z)$ be the optimal value function of the constrained optimization problem $P(y, z)$. Then the pair $(\bar{x}, \lambda)$ satisfies $-\lambda \in \partial v(0, 0)$ and $\bar{x} \in S(0, 0)$ if and only if both the following hold:*

(*i*) *(non-negativity) $\lambda \in K^+ \times Z^*$;*
(*ii*) *(unconstrained optimum) the function*

$$x \mapsto f(x) + \langle \lambda, (g(x), h(x)) \rangle$$

attains its minimum over C at $\bar{x}$;
(*iii*) *(complementary slackness) $\langle \lambda, (g(\bar{x}), h(\bar{x})) \rangle = 0$.*

Proof. (a) The "only if" part. Suppose that $\bar{x} \in S(0, 0)$ and $-\lambda \in \partial v(0, 0)$. By Theorem 2.1 we have $\lambda \in K^+ \times Z^*$. By the definition of the subdifferential and the fact that $v(g(\bar{x}), h(\bar{x})) = v(0, 0)$, we have

$$0 = v(g(\bar{x}), h(\bar{x})) - v(0, 0) \geq \langle -\lambda, (g(\bar{x}), h(\bar{x})) \rangle \geq 0,$$

so that the complementary slackness condition $\langle \lambda, (g(\bar{x}), h(\bar{x})) \rangle = 0$ holds.

Observing that $v(0, 0) = f(\bar{x}) + \langle \lambda, (g(\bar{x}), h(\bar{x})) \rangle$, the strengthened unconstrained optimum condition follows directly from that of Theorem 2.1.

(b) The "if" part. Let $\lambda, \bar{x}$ satisfy conditions (i), (ii) and (iii). Then, for any $x \in C$ satisfying $g(x) \leq_K 0$ and $h(x) = 0$,

$$f(x) \geq f(x) + \langle \lambda, (g(x), h(x)) \rangle \geq f(\bar{x}) + \langle \lambda, (g(\bar{x}), h(\bar{x})) \rangle = f(\bar{x}). \tag{5}$$

That is to say $\bar{x} \in S(0, 0)$.

Moreover, for any $g(x) \leq_K y, h(x) = z$, $f(x) + \langle \lambda, (y, z) \rangle \geq f(x) + \langle \lambda, (g(x), h(x)) \rangle$. Since $v(0, 0) = f(\bar{x})$, by (5) we have

$$f(x) + \langle \lambda, (y, z) \rangle \geq f(\bar{x}) = v(0, 0). \tag{6}$$

Taking the infimum on the left-hand side of (6) yields

$$v(y, z) + \langle \lambda, (y, z) \rangle \geq v(0, 0),$$

which is to say, $-\lambda \in \partial v(0, 0)$. □

We can deduce from Theorems 2.1 and 2.2 that $\partial v(0, 0)$ completely characterizes the set of Lagrange multipliers. Thus, sufficient conditions to ensure the non-emptiness of $\partial v(0, 0)$ are important in analyzing Lagrange multipliers. When v is a lower semicontinuous convex function it is well known that in Banach space $(0, 0) \in \operatorname{core} \operatorname{dom}(v)$ ensures $\partial v(0, 0) \neq \emptyset$. By contrast, ensuring the convexity of v needs strong conditions. The following is a sufficient — if far from necessary — condition. Recall that a function $g \colon C \subseteq X \to Y$ is convex with respect to a convex cone $K \subseteq Y$ (K-*convex*) if $\{(x, y) \colon g(x) \leq_K y, x \in C\}$ is convex.

Lemma 2.1 (Convexity of the Value Function) *Assume that f is a convex function, C is a closed convex set, g is K-convex and h is affine. Then the optimal value function v is a convex extended real-valued function.*

Proof. We consider only the interesting case when $b_1 = (y_1, z_1), b_2 = (y_2, z_2) \in \operatorname{dom} v$ and $t_i \in [0, 1]$ with $t_1 + t_2 = 1$. For any $\varepsilon > 0$, we can find $x_1^\varepsilon, x_2^\varepsilon$ such that $g(x_i^\varepsilon) \le y_i, h(x_i^\varepsilon) = z_i, i = 1, 2,$ and

$$f(x_i^\varepsilon) \le v(b_i) + \varepsilon, i = 1, 2.$$

Since g is K-convex we have

$$g(t_1 x_1^\varepsilon + t_2 x_2^\varepsilon) \le t_1 g(x_1^\varepsilon) + t_2 g(x_2^\varepsilon) \le t_1 y_1 + t_2 y_2,$$

and

$$h(t_1 x_1^\varepsilon + t_2 x_2^\varepsilon) = t_1 h(x_1^\varepsilon) + t_2 h(x_2^\varepsilon) = t_1 z_1 + t_2 z_2,$$

Now using the convexity of f we have

$$v(t_1 b_1 + t_2 b_2) \le f(t_1 x_1^\varepsilon + t_2 x_2^\varepsilon) \le t_1 f(x_1^\varepsilon) + t_2 f(x_2^\varepsilon) \le t_1 v(b_1) + t_2 v(b_2) + \varepsilon.$$

Letting $\varepsilon \to 0$ we derive the convexity of v. □

Remark 2.1 While the subdifferential provides a clear economic interpretation for the Lagrange multiplier, it is not a convenient way of calculating the Lagrange multiplier. This is because finding the Lagrange multiplier this way requires us to solve the original optimization problem for parameters at least in a neighbourhood of $(0, 0)$, which is usually more difficult than the original task. ♢

Remark 2.2 In practice one usually uncovers a Lagrange multiplier by using its properties. To use this method we need to know that (1) the Lagrange multiplier exists, and (2) a convenient calculus for the subdifferential so that condition (ii) can be represented as an inclusion in terms of the subdifferentials of the individual constraint functions.

Requirement (1) amounts to determining that $\partial v(0, 0) \ne \emptyset$. When v is convex, sufficient conditions that ensure this are often called *constraint qualifications*. When the equality constraint $h(x) = 0$ is absent, a rather common but somewhat restrictive condition is the *Slater condition*: there exists $\hat{x} \in X$ such that $g(\hat{x}) <_K 0$ (i.e, the value lies in the topological interior of the cone $-K$.[2] Under this condition $\operatorname{dom} v$ contains an open neighborhood of 0. By a well-known theorem essentially due to Moreau, Rockafellar and Pshenichnyi this implies that $\partial v(0) \ne \emptyset$ (see e.g. [1, Theorem 4.2.8]).

Requirement (2) is also not automatic. For example, again assume the equality constraints are absent and consider the case when $g(x) = (g_1(x), \dots, g_N(x))$ and that all the components of g and f are convex. A well known condition is (see [1, Theorem 4.3.3])

$$\operatorname{dom} f \cap_{n=1}^N \operatorname{cont} g_n \ne \emptyset, \tag{7}$$

where cont g signifies the set of continuous point of function g. Under condition (7), item (ii) in Theorem 2.2 becomes

$$0 \in \partial f(\bar{x}) + \sum_{n=1}^N \lambda_n \partial g_n(\bar{x}). \tag{8}$$

We see that the nature of Lagrange multipliers is to ensure that we can use them as 'shadow price' to penalize the constraints so as to convert the original constrained optimization problem to an unconstrained one. Whether the unconstrained optimization problem leads to a first-order necessary condition in the form of (8), while important for using the Lagrange multiplier to assist us in solving the original problem (1), is a separate issue. ♢

[2] We are assuming that f is everywhere finite, if not we must also require that $f(\hat{x}) < +\infty$.

Remark 2.3 Adding convexity requirements on the constraints and the cost function will ensure that the optimal value function is also convex. This is convenient for establishing the existence of Lagrange multipliers; and allows decomposition of the first-order necessary conditions in a fashion that helps in calculating them. ♢

We again emphasize, however, that these convexity conditions are not intrinsically related to the existence of (local) Lagrange multipliers, as the following example illustrates.

Example 2.1 (Global Lagrange Multiplier for a Nonconvex Problem) Consider

$$v(z) := \min |x| + |\sin(\pi x)| \text{ s.t. } x = z. \tag{9}$$

Clearly, $v(z) = |z| + |\sin(\pi z)|$ so that $\partial v(0) = [-1, 1]$ and $S(0) = \{0\}$. Although the problem is not convex, every $\lambda \in [-1, 1]$ is a Lagrange multiplier globally: for all $x \in R$,

$$|x| + |\sin(\pi x)| + \lambda x \geq 0.$$

Thus, convexity is not essential. ♢

It is possible to give useful conditions on the data in nonlinear settings that still imply v is convex. Indeed, we only need the argument of v to reside in a linear space [20].

Remark 2.4 (The Linear Conic Problem) The following linear conic problem can be considered the simplest form of problem (1): $C = X$, $f(x) = \langle c, x\rangle$, $g(x) = -x$, $h(x) = Ax - b$ where A is a linear operator from X to Z. For this problem $P(0, 0)$ is

$$\min\langle c, x\rangle \text{ s.t. } Ax = b, x \in K. \tag{10}$$

When K is polyhedral and $b \in Z$ with $\dim Z < +\infty$ this problem is indeed simple (see Examples 3.4 and 3.5 below). However, even for this simplest form, it turns out that characterizing the existence of a Lagrange multiplier in general is nontrivial [21, 22]. ♢

2.3 Use of the Fréchet Subdifferential

Lagrange multipliers as discussed in Section 2.2 help to convert a constrained optimization problem to a *globally* unconstrained problem. Success, of course, is desirable but rare. Most of the time for nonlinear nonconvex problems one can only hope for a local solution. For these problems we need subdifferentials that are more suitable for capturing local behavior. The development of nonsmooth analysis in the past several decades has led to many such subdifferentials (see [18, 23, 24]).

Each such subdifferential is accompanied by its own version(s) of Lagrange multiplier rules with different strengths for different problems. An exhaustive survey is neither possible nor is it our goal. Rather, we illustrate the derivation of suitable Lagrange multiplier rules by using the very importsnt Fréchet subdifferential as an example. It reflects the key features of many results of this kind.

First, we recall the definition of a *Fréchet subdifferential* and the corresponding concept of a normal cone.

Definition 2.2 (Fréchet Subdifferential and Normal Cone) The Fréchet subdifferential of a function ϕ at $x \in \mathrm{dom}\ \phi \subset X$ is defined by

$$\partial_F \phi(x) := \{x^* \in X^* : \liminf_{\|y\| \to 0} \frac{\phi(x+y) - \phi(x) - \langle x^*, y \rangle}{\|y\|} \geq 0\}.$$

For a set $C \subset X$ and $x \in C$, we define the Fréchet normal cone of C at x by

$$N_F(C; x) := \partial \iota_C(x),$$

where ι_C is the *indicator function* of set C that is equalsto 0 when $x \in C$ and is $+\infty$ otherwise.

Remark 2.5 (Viscosity Subdifferentials) When the Banach space X has an equivalent Fréchet smooth norm [1], as holds in all reflexive spaces and in many others, something lovely happens: $x^* \in \partial_F \phi(x)$ if and only if there exists a concave function $\eta \in C^1$ with $\eta'(x) = 0$ such that

$$y \mapsto \phi(y) - \langle x^*, y \rangle + \eta(y) \tag{11}$$

attains a local minimum at $y = x$ (for details see [25,26]).

Thus, any subdifferential of f at x is actually the derivative of a minorizing concave function g which *osculates* with f at x as in Figure 1. We call such an object a *viscosity subdifferential* [1,25,26]. The use of viscosity subdifferentials frequently allows us to use smooth techniques in nonsmooth analysis.

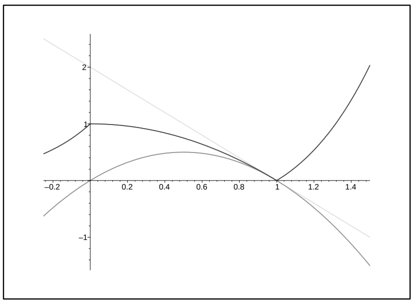

Fig. 1 A Viscosity Subdifferential

Theorem 2.3 (Local Lagrange Multiplier Rule) *Let X be a Banach space and let Y, Z be Banach spaces with equivalent Fréchet smooth norms. Assume that in problem $P(y,z)$, f, g are lower semicontinuous and h is continuous on X. Let $v(y,z)$ be the optimal value function of the constrained local optimization problem $P(y,z)$ and let $\bar{x} \in S(0,0)$. Suppose $-\lambda \in \partial_F v(0,0)$. Then*

(*i*) *(non-negativity) $\lambda \in K^+ \times Z^*$;*
(*ii*) *(unconstrained optimum) there exists $\eta \in C^1$ such that the function*

$$x \mapsto f(x) + \langle \lambda, (g(x), h(x)) \rangle + \eta(g(x), h(x))$$

attains a local minimum at $\bar{x}$ over C;
(*iii*) *(complementary slackness) $\langle \lambda, (g(\bar{x}), h(\bar{x})) \rangle = 0$.*

Proof. Suppose that $-\lambda \in \partial_F v(0,0)$. It is easy to see that $v(y,0)$ is non-increasing with respect to the partial order $\leq_K$. Thus, for any $y \in K$,

$$\langle \lambda, (y/\|y\|, 0)\rangle \geq \liminf_{t\downarrow 0} \frac{v(y,0) - v(0,0) + \langle \lambda, (ty,0)\rangle}{t\|(y,0)\|} \geq 0, \tag{12}$$

so that $\lambda \in K^+ \times Z^*$ which verifies (i).

The complementary slackness condition holds trivially if $g(\bar{x}) = 0$. Assuming, thus, that $g(\bar{x}) \neq 0$, then for any $t \in (0,1)$, $g(\bar{x}) \leq_K tg(\bar{x}) \leq_K 0$ and $h(\bar{x}) = th(\bar{x}) = 0$. Since

$$v(0,0) \geq v(g(\bar{x}), h(\bar{x})) \geq v(tg(\bar{x}), th(\bar{x})) \geq v(0,0),$$

the terms above are all equal. It follows that

$$0 \geq \left\langle \lambda, \left(\frac{g(\bar{x})}{\|g(\bar{x})\|}, 0\right)\right\rangle = \liminf_{t\downarrow 0} \frac{v(tg(\bar{x}), th(\bar{x})) - v(0,0) + \langle \lambda, (tg(\bar{x}), th(\bar{x}))\rangle}{t\|(tg(\bar{x}), th(\bar{x}))\|} \geq 0. \tag{13}$$

Thus, $\langle \lambda, (g(\bar{x}), 0)\rangle = \langle \lambda, (g(\bar{x}), h(\bar{x}))\rangle = 0$. This establishes (iii).

Finally, to prove (ii) we first observe that

$$\partial_F v(0,0) \subset \partial_F v(g(\bar{x}), h(\bar{x})). \tag{14}$$

This is because if $(y^*, z^*) \in \partial_F v(0,0)$ then

$$\begin{aligned} &\liminf_{\|(y,z)\|\to 0} \frac{v(g(\bar{x}) + y, h(\bar{x}) + z) - v(g(\bar{x}), h(\bar{x})) - \langle (y^*, z^*), (y,z)\rangle}{\|(y,z)\|} \\ &\geq \liminf_{\|(y,z)\|\to 0} \frac{v((y,z)) - v(0,0) - \langle (y^*, z^*), (y,z)\rangle}{\|(y,z)\|} \geq 0, \end{aligned}$$

that is, $(y^*, z^*) \in \partial_F v(g(\bar{x}), h(\bar{x}))$.

Since $-\lambda \in \partial_F v(0,0) \subset \partial_F v(g(\bar{x}), h(\bar{x}))$ by Remark 2.5 we have the existence of a C^1 function η with $\eta'(g(\bar{x}), h(\bar{x})) = 0$ such that, for any $x \in C$ close enough to $\bar{x}$,

$$\begin{aligned} f(x) &+ \langle \lambda, (g(x), h(x))\rangle + \eta(g(x), h(x)) \\ &\geq v(g(x), h(x)) + \langle \lambda, (g(x), h(x))\rangle + \eta(g(x), h(x)) \\ &\geq v(g(\bar{x}), h(\bar{x})) + \langle \lambda, (g(\bar{x}), h(\bar{x}))\rangle + \eta(g(\bar{x}), h(\bar{x})), \end{aligned}$$

as required. We are done. □

Remark 2.6 Unlike the situation in Theorem 2.2, the converse of Theorem 2.3 may not be true because the inclusion $\partial_F v(0,0) \subset \partial_F v(g(\bar{x}), h(\bar{x}))$ is typically proper. Also, analysis of the proof we gave shows that the existence of the local solution $\bar{x}$ is prerequisite to give an anchor to the local behavior and, therefore, an analog of Theorem 2.1 is not to be expected. ♢

Remark 2.7 As with Theorem 2.2, it is important in practice to have a convenient calculus for the Fréchet subdifferential so that (ii) can be represented in terms of the subdifferentials of the individual constraints — the data.

For instance, let us suppose $Y = \mathbb{R}^N, Z = \mathbb{R}^M$, g and h are both C^1. Then we can derive from (ii) — using a derivative sum rule [1] — that

$$f'(\bar{x}) + \sum_{n=1}^{N} \lambda_n g_n'(\bar{x}) + \sum_{m=1}^{M} \lambda_{N+m} h_m'(\bar{x}) = 0.$$

This is the Lagrange multiplier rule one most usually sees (except that the constraint qualification condition is given in the form of $\partial_F v(0,0) \neq \emptyset$). Note that, in deriving this decoupled form we used both sum and chain rules for the Fréchet subdifferential. ♢

Without smoothness assumptions on the data the situation is more challenging. Any decoupled form has to be in approximate or "fuzzy" form [1]. It turns out there are more direct ways of deriving such results, as first discussed in [25] and then in [27,28] with [27] requiring the weakest conditions. To illustrate the idea, we again use $Y = \mathbb{R}^N$ and $Z = \mathbb{R}^M$. Using *indicator functions* of the level sets to penalize infeasible points we may write problem $P(0,0)$ as an unconstrained problem:

$$\min \ f(x) + \sum_{n=1}^{N} \iota_{[g_n \le 0]}(x) + \sum_{m=1}^{M} \iota_{[h_m = 0]}(x) + \iota_C(x). \tag{15}$$

Suppose that $\bar{x} \in S(0,0)$. Denote the closed ball around a point x with radius r by $B_r(x)$. Then applying an approximate sum rule in [1,25] we have, for any $\varepsilon > 0$ and weak* neighborhood V of 0, that there exist $y'_n \in B_\varepsilon(\bar{x})$, $n = 0, 1, \ldots N$, $z'_m \in B_\varepsilon(\bar{x}), m = 1, \ldots, M$, $x_0, x' \in B_\varepsilon(\bar{x})$ such that

$$0 \in \partial_F f(x_0) + \sum_{n=1}^{N} \partial_F \iota_{[g_n \le 0]}(y'_n) + \sum_{m=1}^{M} \partial_F \iota_{[h_m = 0]}(z'_m) + N_F(x', C) + V. \tag{16}$$

The key is then to represent $\partial_F \iota_{[g_n \le 0]}(y'_n)$ and $\partial_F \iota_{[h_m = 0]}(z'_m)$ in terms of the Fréchet subdifferential of the related functions. Such representations were first discussed in [25] with the additional condition that that Fréchet subdifferentials of these functions were bounded away from zero near the points of concern. These conditions have been weakened somewhat in [28] and were eventually completely avoided in [27].

Lemma 2.2 (Subdifferential Representation of Normal Vectors to a Level Set [27]) *Let V be a weak* neighborhood of $0 \in X^*$ and let f be a lower semicontinuous (resp. continuous) function around $x \in [f \le 0]$ $(x \in [f = 0])$.*

Then, for any $\eta > 0$ and $x^* \in N_F(x, [f \le 0])$ $(x^* \in N_F(x, [f = 0]))$, there exist $u \in B_\eta(x)$, $u^* \in \partial_F f(u)$ $(u^* \in \partial_F f(u) \cup \partial(-f)(u))$ and $\xi > 0$ such that

$$x^* \in \xi u^* + V, \text{ and } \xi |f(u)| < \eta.$$

Combining (16) and Lemma 2.2 we arrive at a powerful result:

Theorem 2.4 (Decoupled Local Lagrange Multiplier Rule) *Assume that in problem $P(y,z)$, f, g are lower semicontinuous, h is continuous and X has an equivalent Fréchet smooth norm. Assume that $Y = \mathbb{R}^N$, $Z = \mathbb{R}^M$ and that $\bar{x} \in S(0,0)$.*

Then, for any $\varepsilon > 0$ and weak neighborhood V of $0 \in X^*$, there exist $y_n \in B_\varepsilon(\bar{x})$, $n = 0, 1, \ldots, N$, $z_m \in B_\varepsilon(\bar{x}), m = 1, \ldots, M$, $x_0, \hat{x} \in B_\varepsilon(\bar{x})$, $\lambda \in \mathbb{R}_+^{N+M+1}$ with $\sum_{n=0}^{N+M} \lambda_n = 1$ such that*

$$|f(x_0) - f(\bar{x})| < \varepsilon, \lambda_n |g_n(y_n)| < \varepsilon, \lambda_{m+N} |h_m(z_m)| < \varepsilon,$$

and

$$0 \in \lambda_0 \partial_F f(x_0) + \sum_{n=1}^{N} \lambda_n \partial_F g_n(y_n) + \sum_{m=1}^{M} \lambda_{m+N} (\partial_F h_m(z_m) \cup \partial_F(-h_m)(z_m)) + N_F(\hat{x}, C) + V.$$

Proof. First using Lemma 2.2 we can replace all the normal vectors for the sublevel and level sets in (16) by the corresponding subdifferentials of the components of g and h at nearby points, respectively. Finally, we rescale the resulting inclusion using $\lambda_0 = 1/(1 + \sum_{n=1}^{N+M} \lambda_n)$. □

We call this an *approximate decoupling.* In finite dimensions where the neighbourhoods are bounded, or under stronger compactness conditions on f, g, h, we may often move to a limiting form as $x_0, \hat{x}, y_n, z_m$ approach $\bar{x}$. If in this process λ_0 can be arbitrarily close to 0 then the multipliers are detached from the cost function f, and as a result the limit does not provide much useful information. Such multipliers are usually referred to as *degenerate* or *singular*. To avoid degenerate multipliers additional *constraint qualification* conditions are needed. The following is one such condition [1], generalizing the classical Mangasarian-Fromovitz constraint qualification condition [29] to nonsmooth problems. Many new developments regarding various types of constraint qualification conditions can be found in [30].

(CQ) There exist constants $\varepsilon, c > 0$ such that, for any $(x_0, f(x_0)) \in B_\varepsilon((\bar{x}, f(\bar{x})))$, $z_m \in B_\varepsilon(\bar{x})$, $m = 1, \ldots, M$, $(y_n, g_n(y_n)) \in B_\varepsilon((\bar{x}, g_n(\bar{x}))), n = 1, \ldots, N$, and $\hat{x} \in B_\varepsilon(\bar{x})$ and for any $\lambda_n \geq 0, n = 1, \ldots, N+M$ with $\sum_{n=1}^{N+M} \lambda_n = 1$,

$$d\left(0, \sum_{n=1}^{N} \lambda_n \partial_F g_n(y_n) + \sum_{m=1}^{M} \lambda_{m+N}(\partial_F h_m(z_m) \cup \partial_F(-h_m)(z_m)) + N_F(\hat{x}, C)\right) \geq c.$$

Here $d(y, S) := \inf\{\|y - s\| : s \in S\}$ represents the distance from point y to set S.

With this constraint qualification condition we have the following generalization of the classical Karush-Kuhn-Tucker necessary condition for constrained optimization problems.

Theorem 2.5 (Nonsmooth Karush-Kuhn-Tucker Necessary Condition) *Assume that in problem $P(y, z)$, f is locally Lipschitz, g is lower semicontinuous, h is continuous and X is finite dimensional. Assume that $Y = \mathbb{R}^N$, $Z = \mathbb{R}^M$, $\bar{x} \in S(0, 0)$ and that the constraint qualification condition CQ holds.*

Then, for any $\varepsilon > 0$, there exist $y_n \in B_\varepsilon(\bar{x}), n = 0, 1, \ldots, N$, $z_m \in B_\varepsilon(\bar{x}), m = 1, \ldots, M$, $x_0, \hat{x} \in B_\varepsilon(\bar{x})$, $\lambda \in \mathbb{R}_+^{N+M+1}$ and a positive constant K such that

$$|f(x_0) - f(\bar{x})| < \varepsilon, \lambda_n |g(y_n)| < \varepsilon, \lambda_{m+N}|h(z_m)| < \varepsilon,$$

and

$$0 \in \partial_F f(x_0) + \sum_{n=1}^{N} \lambda_n \partial_F g_n(y_n) + \sum_{m=1}^{M} \lambda_{m+N}(\partial_F h_m(z_m) \cup \partial_F(-h_m)(z_m)) + N_F(\hat{x}, C) + \varepsilon B_{X^*},$$

where $\lambda_n \in [0, K], n = 1, \ldots, N+M$.

Proof. Let L be the Lipschitz constant of f near $\bar{x}$. Then for x_0 sufficiently close to $\bar{x}$ we have $d(0, \partial_F f(x_0)) \leq L$. Without loss of generality we may assume that $\varepsilon < c$. Applying Theorem 2.4 we have $\hat{x} \in B_\varepsilon(\bar{x})$, $(x_0, f(x_0)) \in B_\varepsilon((\bar{x}, f(\bar{x})))$, $(y_n, g_n(y_n)) \in B_\varepsilon((\bar{x}, g_n(\bar{x})))$ and $(z_m, h_m(z_m)) \in B_\varepsilon((\bar{x}, h_m(\bar{x})))$ for $n = 1, \ldots, N, m = 1, \ldots, M$ and $\lambda' \in \mathbb{R}_+^{N+M+1}$ such that

$$\lambda'_n |g(y_n)| < \varepsilon, \lambda'_{m+N}|h(z_m)| < \varepsilon,$$

and

$$0 \in \lambda'_0 \partial_F f(x_0) + \sum_{n=1}^{N} \lambda'_n \partial_F g_n(y_n) \qquad (17)$$
$$+ \sum_{m=1}^{M} \lambda'_{m+N}(\partial_F h_m(z_m) \cup \partial_F(-h_m)(z_m)) + N_F(\hat{x}, C) + (\varepsilon/2) B_{X^*},$$

where $\sum_{n=0}^{N+M} \lambda'_n = 1$. If $\lambda'_0 = 1$, the theorem holds with any $K > 0$. Otherwise, using CQ we have

$$\lambda'_0 L \geq \lambda'_0 d(0, \partial_F f(x_0)) = d(0, \lambda'_0 \partial_F f(x_0)) \geq (1 - \lambda'_0)c - \varepsilon/2 \geq (1 - \lambda'_0)c - c/2.$$

Thus, $\lambda'_0 \geq c/2(c + L)$. It remains to multiply (17) by $1/\lambda'_0$ and set $\lambda_n := \lambda'_n/\lambda'_0$ and $K = 2(c + L)/c$ to complete the proof. □

3 Convex Duality Theory

Obtaining Lagrange multipliers by using convex subdifferentials is irrevocably related to convex duality theory. Let us first recall that for a lower semi-continuous extended valued function f, its *convex conjugate* is defined by

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

Note that f^* is always convex even though f itself may not be. We also consider the conjugate of f^* denoted f^{**} defined on $X^{**} \supset X$. It is easy to show that f^{**}, when restricted to X, is the largest lower semi-continuous convex function satisfying $f^{**} \leq f$. The following *Fenchel-Young inequality* follows directly from the definition:

$$f^*(x^*) + f(x) \geq \langle x^*, x\rangle \tag{18}$$

with equality holding if and only if $x^* \in \partial f(x)$.

Using the Fenchel-Young inequality for each constrained optimization problem we can write its companion *dual problem*. There are several different but equivalent perspectives.

3.1 Rockafellar Duality

We start with the Rockafellar formulation of bi-conjugate. It is very general and — as we shall see — other perspectives can easily be written as special cases.

Consider a two variable function $F(x, y)$. Treating y as a parameter, consider the parameterized optimization problem

$$v(y) = \inf_x F(x, y). \tag{19}$$

Our associated *primal* optimization problem[3] is

$$p = v(0) = \inf_{x\in X} F(x, 0) \tag{20}$$

and the *dual* problem is

$$d = v^{**}(0) = \sup_{y^*\in Y^*} -F^*(0, -y^*). \tag{21}$$

Since v dominates v^{**} as the Fenchel-Young inequality establishes, we have

$$v(0) = p \geq d = v^{**}(0).$$

This is called *weak duality* and the non-negative number $p-d = v(0)-v^{**}(0)$ is called the *duality gap* — which we aspire to be small or zero.

Let $F(x, (y, z)) := f(x)+\iota_{\text{epi}(g)}(x, y)+\iota_{\text{graph}(h)}(x, z)$. Then problem $P(y, z)$ becomes problem (19) with parameters (y, z). On the other hand, we can rewrite (19) as

$$v(y) = \inf_x \{F(x, u) : u = y\}$$

which is problem $P(0, y)$ with $x = (x, u)$, $C = X \times Y$, $f(x, u) = F(x, u)$, $h(x, u) = u$ and $g(x, u) = 0$. So where we start is a matter of taste and predisposition.

[3] The use of the term ‘primal’ is much more recent than the term ‘dual’ and was suggested by George Dantzig’s father Tobias when linear programming was being developed in the 1940’s.

Theorem 3.1 (Duality and Lagrange Multipliers) *The following are equivalent:*

(i) *the primal problem has a Lagrange multiplier* λ*;*
(ii) *there is no duality gap, i.e.,* $d = p$ *is finite and the dual problem has solution* $-\lambda$*.*

Proof. If the primal problem has a Lagrange multiplier λ then $-\lambda \in \partial v(0)$. By the Fenchel-Young equality

$$v(0) + v^*(-\lambda) = \langle -\lambda, 0\rangle = 0.$$

Direct calculation yields

$$v^*(-\lambda) = \sup_y\{\langle -\lambda, y\rangle - v(y)\} = \sup_{y,x}\{\langle -\lambda, y\rangle - F(x,y)\} = F^*(0,-\lambda).$$

Since

$$-F^*(0,-\lambda) \le v^{**}(0) \le v(0) = -v^*(-\lambda) = -F^*(0,-\lambda), \tag{22}$$

λ is a solution to the dual problem and $p = v(0) = v^{**}(0) = d$.

On the other hand, if $v^{**}(0) = v(0)$ and λ is a solution to the dual problem then all the quantities in (22) are equal. In particular,

$$v(0) + v^*(-\lambda) = 0.$$

This implies that $-\lambda \in \partial v(0)$ so that λ is a Lagrange multiplier of the primal problem. □

Example 3.1 (Finite duality gap) Consider

$$v(y) = \inf\{|x_2 - 1| : \sqrt{x_1^2 + x_2^2} - x_1 \le y\}.$$

We can easily calculate

$$v(y) = \begin{cases} 0, & y > 0, \\ 1, & y = 0, \\ +\infty, & y < 0, \end{cases}$$

and $v^{**}(0) = 0$, i.e. there is a finite duality gap $v(0) - v^{**}(0) = 1$.

In this example neither the primal nor the dual problem has a Lagrange multiplier yet both have solutions. Hence, even in two dimensions, existence of a Lagrange multiplier is only a sufficient condition for the dual to attain a solution and is far from necessary. ♢

3.2 Fenchel Duality

Let us specify $F(x,y) := f(x) + g(Ax + y)$, where $A : X \to Y$ is a linear operator. We thence get the Fenchel formulation of duality [1]. Now the primal problem is

$$p = v(0) = \inf_x\{f(x) + g(Ax)\}. \tag{23}$$

To derive the dual problem we calculate

$$F^*(0,-y^*) = \sup_y\{\langle -y^*, y\rangle - f(x) - g(Ax + y)\}.$$

Letting $u = Ax + y$ we have

$$F^*(0, -y^*) = \sup_{x,u}\{\langle -y^*, u - Ax\rangle - f(x) - g(u)\}$$
$$= \sup_x\{\langle y^*, Ax\rangle - f(x)\} + \sup_u\{\langle -y^*, u\rangle - g(u)\} = f^*(A^*y^*) + g^*(-y^*).$$

Thus, the dual problem is

$$d = v^{**}(0) = \sup_{y^*}\{-f^*(A^*y^*) - g^*(-y^*)\}. \tag{24}$$

If both f and g are convex functions it is easy to see that so is

$$v(y) = \inf_x\{f(x) + g(Ax + y)\}.$$

Now, in the Euclidean setting, a sufficient condition for the existence of Lagrange multipliers[4] for the primal problem, i.e., $\partial v(0) \neq \emptyset$, is

$$0 \in \text{ri core dom}\, v = \text{ri core}[\text{dom}\, g - A\ \text{dom}\, f]. \tag{25}$$

Figure 2 illustrate the Fenchel duality theorem for $f(x) := x^2/2+1$ and $g(x) = (x-1)^2/2+1/2$. The upper function is f and the lower one is $-g$. The minimum gap occurs at 1/2 and is 7/4.

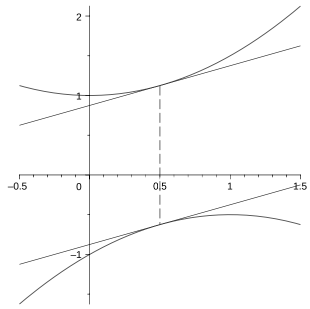

Fig. 2 The Fenchel Duality Sandwich

In infinite dimensions more care is necessary, but in Banach space when the functions are lower semicontinuous and the operator is continuous (or merely closed)

$$0 \in \text{core dom}\, v = \text{core}[\text{dom}\, g - A\ \text{dom}\, f]. \tag{26}$$

suffices, as does the extension to the relative core in the closed subspace generated by dom v [31].

A condition of the form (25) or (26) is often referred to as a *constraint qualification* or a *transversality* condition. Enforcing such constraint qualification conditions we can write Theorem 3.1 in the following form:

Theorem 3.2 (Duality and Constraint Qualification) *If the convex functions f, g and the linear operator A satisfy the constraint qualification conditions* (25) *or* (26) *then there is a zero duality gap between the primal and dual problems,* (23) *and* (24)*, and the dual problem has a solution.*

[4] In infinite dimensions we also assume that f, g are lsc and that A is continuous.

A really illustrative example is the application to entropy optimization.

Example 3.2 (Entropy Optimization Problem) *Entropy maximization*[5] refers to

$$\min\ f(x) \text{ s.t. } Ax = b \in \mathbb{R}^N, \tag{27}$$

with the lower semicontinuous convex function f defined on a Banach space of signals, emulating the negative of an entropy and A emulating a finite number of continuous linear constraints representing conditions on some given *moments*. A wide variety of applications can be covered by this model due to its physical relevance.

Applying Theorem 3.2 with $g = \iota_{\{b\}}$ as in [1] we have if $b \in \text{core}\,(A\ \text{dom}\, f)$ then

$$\inf_{x\in X}\{f(x) : Ax = b\} = \max_{\phi\in\mathbb{R}^N}\{\langle\phi, b\rangle - f^*(A^*\phi)\}. \tag{28}$$

If $N < \dim X$ (often infinite) the dual problem is often much easier to solve then the primal. $\diamondsuit$

Example 3.3 (Boltzmann–Shannon entropy in Euclidean space) Let

$$f(x) := \sum_{n=1}^{N} p(x_n), \tag{29}$$

where

$$p(t) := \begin{cases} t\ln t - t, & \text{if } t > 0,\\ 0. & \text{if } t = 0,\\ +\infty, & \text{if } t < 0.\end{cases}$$

The functions p and f defined above are (negatives of) Boltzmann–Shannon entropy functions on $\mathbb{R}$ and $\mathbb{R}^N$, respectively. For $c \in \mathbb{R}^N$, $b \in \mathbb{R}^M$ and linear mapping $A : \mathbb{R}^N \to \mathbb{R}^M$ consider the entropy optimization problem

$$\min\ f(x) + \langle c, x\rangle \text{ s.t. } Ax = b. \tag{30}$$

Example 3.2 can help us conveniently derive an explicit formula for solutions of (30) in terms of the solution to its dual problem.

First we note that the sublevel sets of the objective function are compact, thus ensuring the existence of solutions to problem (30). We can also see by direct calculation that the directional derivative of the cost function is $-\infty$ on any boundary point x of $\text{dom}\, f = \mathbb{R}^N_+$, the domain of the cost function, in the direction of $z - x$. Thus, any solution of (30) must be in the interior of $\mathbb{R}^N_+$. Since the cost function is strictly convex on int $(\mathbb{R}^N_+)$, the solution is unique.

Let us denote this solution of (30) by $\bar{x}$. The duality result in Example 3.2 implies that

$$f(\bar{x}) + \langle c, \bar{x}\rangle = \inf_{x\in\mathbb{R}^N}\{f(x) + \langle c, x\rangle : Ax = b\} = \max_{\phi\in\mathbb{R}^M}\{\langle\phi, b\rangle - (f + c)^*(A^\top\phi)\}.$$

Now let $\bar{\phi}$ be a solution to the dual problem, i.e., a Lagrange multiplier for the constrained minimization problem (30).

We have

$$f(\bar{x}) + \langle c, \bar{x}\rangle + (f + c)^*(A^\top\bar{\phi}) = \langle\bar{\phi}, b\rangle = \langle\bar{\phi}, A\bar{x}\rangle = \langle A^\top\bar{\phi}, \bar{x}\rangle.$$

[5] Actually this term arises because in the Boltzmann-Shannon case one is minimizing the negative of the entropy.

It follows from the Fenchel-Young equality [33] that $A^\top\bar\phi \in \partial(f+c)(\bar x)$. Since $\bar x \in \operatorname{int}(\mathbb{R}^N_+)$ where f is differentiable, we have $A^\top\bar\phi = f'(\bar x)+c$. Explicit computation shows $\bar x = (\bar x_1, \dots, \bar x_N)$ is determined by

$$\bar x_n = \exp(A^\top\bar\phi - c)_n, n = 1, \dots, N. \tag{31}$$

Indeed, we can use the existence of the dual solution to prove that the primal problem has the given solution without direct appeal to compactness — we deduce the existence of the primal directly from convex duality theory [1,32]. ♢

We say a function f is polyhedral if its epigraph is polyhedral, i.e., a finite intersection of closed half-spaces. When both f and g are polyhedral constraint qualification condition (25) simplifies to (see [33, Section 5.1] for the hard work required here)

$$\operatorname{dom} g \cap A \ \operatorname{dom} f \neq \emptyset. \tag{32}$$

This is very useful in dealing with polyhedral cone programming and, in particular, linear programming problems. One can also similarly handle a subset of polyhedral constraints [1].

Example 3.4 (Abstract linear programming) Consider the cone programming problem (10) and its dual is

$$\sup\{\langle b, \phi\rangle \colon A^*\phi + c \in K^+\}. \tag{33}$$

Assuming that K is polyhedral, if $b \in AK$ (i.e., the primary problem is feasible) implies that there is no duality gap and the dual optimal value is attained when finite. Symmetrically, if $-c \in A^*X^* - K^+$ (i.e., the dual problem is feasible) then there is no duality gap and the primary optimal value is attained if finite.

3.3 Lagrange Duality Reobtained

For problem (1) define the *Lagrangian*

$$L(\lambda, x; (y, z)) = f(x) + \langle \lambda, (g(x) - y, h(x) - z)\rangle.$$

Then

$$\sup_{\lambda \in Y^*_+ \times Z^*} L(\lambda, x; (y, z)) = \begin{cases} f(x), & \text{if } g(x) \leq_K y, h(x) = z, \\ +\infty, & \text{otherwise.} \end{cases}$$

Then problem (1) can be written as

$$p = v(0) = \inf_{x \in C} \sup_{\lambda \in Y^*_+ \times Z^*} L(\lambda, x; 0). \tag{34}$$

We can calculate

$$\begin{aligned} v^*(-\lambda) &= \sup_{y,z}\{\langle -\lambda, (y, z)\rangle - v(y, z)\} \\ &= \sup_{y,z}\{\langle -\lambda, (y, z)\rangle - \inf_{x \in C}[f(x) : g(x) \leq_K y, h(x) = z]\} \\ &= \sup_{x \in C, y, z}\{\langle -\lambda, (y, z)\rangle - f(x) : g(x) \leq_K y, h(x) = z\}. \end{aligned}$$

Letting $\xi = y - g(x) \in K$ we can rewrite the expression above as

$$\begin{aligned} v^*(-\lambda) &= \sup_{x\in C,\xi\in K} \{\langle -\lambda, (g(x), h(x))\rangle - f(x) + \langle -\lambda, (\xi, 0)\rangle\} \\ &= -\inf_{x\in C,\xi\in K} \{L(\lambda, x; 0) - \langle \lambda, (\xi, 0)\rangle\} = \begin{cases} -\inf_{x\in C} L(\lambda, x; 0), & \text{if } \lambda \in Y_+^* \times Z^*, \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned}$$

Thus, the dual problem is

$$d = v^{**}(0) = \sup_{\lambda} -v^*(-\lambda) = \sup_{\lambda\in Y_+^*\times Z^*} \inf_{x\in C} L(\lambda, x; 0). \tag{35}$$

We can see that the weak duality inequality $v(0) \geq v^{**}(0)$ is simply the familiar fact that

$$\inf\sup \geq \sup\inf .$$

Example 3.5 (Classical Linear Programming Duality) Consider a linear programming problem

$$\min\langle c, x\rangle \text{ s.t. } Ax \leq b, \tag{36}$$

where $x \in \mathbb{R}^N$, $b \in \mathbb{R}^M$, A is a $M \times N$ matrix and $\leq$ is the partial order generated by the cone $\mathbb{R}^M_+$. Then by Lagrange duality the dual problem is

$$\max\langle -b, \phi\rangle \text{ s.t. } A^*\phi = -c, \phi \geq 0. \tag{37}$$

Clearly, all the functions involved herein are polyhedral. Applying the polyhedral cone duality results in Example 3.4, we can conclude that if either the primary problem or the dual problem is feasible then there is no duality gap. Moreover, when the common optimal value is finite then both problems have optimal solutions. ♢

The hard work in Example 3.5 was hidden in establishing that the constraint qualification (32) is sufficient, but unlike many applied developments we have *rigorously* recaptured linear programming duality within our framework.

4 Further Examples and Applications

The basic Lagrange multiplier rule is related to many other important results, and its value goes much beyond t merely facilitating computation when looking for solutions to constrained optimization problems. The following are a few substantial examples.

4.1 Subdifferential of a Maximum Function

The *maximum function* (or max function) given by

$$m(x) := \max\{f_1(x), \ldots, f_N(x)\}. \tag{38}$$

is very useful yet intrinsically nonsmooth. Denote $I(\bar{x}) := \{n \in \{1, 2, \ldots, N\} : f_n(\bar{x}) = m(\bar{x})\}$. This is the subset of *active* component functions. Characterizing the (generalized) derivative of m is rather important.

For simplicity, we assume throughout that we work in a Euclidean space. We consider the convex case first.

Theorem 4.1 (Convex Max Function) *When all the functions $f_n, n = 1, \dots, N$ in (38) are convex and lower semicontinuous, then so is $m(x)$. Suppose further that $x^* \in \partial m(\bar{x})$.*

Then, there exists nonnegative numbers $\lambda \in R^N_+$ with $\sum_{n=1}^N \lambda_n = 1$ satisfying the complementary slackness condition $\langle \lambda, f(\bar{x}) - m(\bar{x})\mathbf{1}\rangle = \mathbf{0}$, where $f = (f_1, \dots, f_N)$ such that

$$x^* \in \partial \sum_{n \in I} \lambda_n f_n(\bar{x}).$$

Proof. Consider the following form of a constrained minimization problem

$$v(y, z) := \min\{r - \langle x^*, u\rangle : f(u) - r\mathbf{1} \le \mathbf{y}, \mathbf{u} - \bar{\mathbf{x}} = \mathbf{z}\}. \tag{39}$$

Clearly, $(\bar{x}, m(\bar{x}))$ is a solution to problem $v(0, 0)$. Let ∂_z represent the subdifferential of v with respect to variable z. We can see that $v(0, z) = m(\bar{x} + z) - \langle x^*, \bar{x} + z\rangle$ so that $0 \in \partial_z v(0, 0)$.

Furthermore, $v(y, 0) = \max(f_1(\bar{x}) + y_1, \dots, f_N(\bar{x}) + y_N) - \langle x^*, x\rangle$ is a convex function of y with domain R^N and, thus, $\partial_y v(0, 0) \neq \emptyset$. It follows that problem (39) has a Lagrange multiplier of the form $(\lambda, 0)$ with $\lambda \in \mathbb{R}^N_+$ such that

$$(r, u) \mapsto r - \langle x^*, u\rangle + \sum_{n=1}^N \lambda_n (f_n(u) - r)$$

attains minimum at $(\bar{x}, m(\bar{x}))$ and

$$\sum_{n=1}^N \lambda_n (f_n(\bar{x}) - m(\bar{x})) = 0. \tag{40}$$

Using the definition of the subdifferential, and decoupling information regarding variables u, r, we have $\sum_{n=1}^N \lambda_n = 1$ and $x^* \in \partial \sum_{n=1}^N \lambda_n f_n(\bar{x})$. The complementary slackness condition (40) implies the more precise inclusion

$$x^* \in \partial \sum_{n \in I(\bar{x})} \lambda_n f_n(\bar{x}).$$

This completes the proof. □

Remark 4.1 We note that Theorem 4.1 is a special case of much more general formulas for the generalized derivatives of the supremum of (possibly) infinitely many functions (see, e.g., [34]). Nevertheless, despite the different levels of generality, in all such representation theorems for generalized derivatives of the maximum function, it is the Lagrange multipliers that plays a key role. ♢

4.2 Separation and Sandwich Theorems

We start with a proof of the standard Hahn-Banach separation theorem using the Lagrange multiplier rule.

Theorem 4.2 (Hahn-Banach Separation Theorem) *Let X be a Banach space and let C_1 and C_2 be convex subsets of X. Suppose that*

$$C_2 \cap \operatorname{int} C_1 = \emptyset. \tag{41}$$

Then there exists $\lambda \in X^* \backslash \{0\}$ such that, for all $x \in C_1$ and $y \in C_2$,

$$\langle \lambda, y\rangle \ge \langle \lambda, x\rangle. \tag{42}$$

Proof. Without loss of generality we assume that $0 \in \operatorname{int} C_1$. Then the convex *gauge function* of C_1,

$$\gamma_{C_1}(x) := \inf\{t \colon x \in tC_1\}$$

is defined for all $x \in X$ and has the property that $\gamma_{C_1}(x) < 1$ if and only if $x \in \operatorname{int} C_1$.

Consider the constrained convex optimization problem

$$p = \min\{\gamma_{C_1}(x) - 1 \colon y - x = 0, y \in \text{cl } C_2\}. \tag{43}$$

Then the separation condition (41) and the propertyies of the gauge function together imply that $p \geq 0$.

Setting $f(x) = \gamma_{C_1}(x) - 1$ and $g(x) = \iota_{\text{cl } C_2}(x)$, clearly (43) is equivalent to the Fenchel primal optimization problem (23) with A being the identity operator. Since $\operatorname{dom} f = X$ the constraint qualification condition (25) holds. Thus, problem (43) has a Lagrange multiplier $\lambda \in X^* \backslash \{0\}$[6] such that, by Theorem 2.1, for all $x \in X$ and $y \in \operatorname{cl} C_2$,

$$\gamma_{C_1}(x) - 1 + \langle \lambda, y - x \rangle \geq p \geq 0. \tag{44}$$

Rewrite (44) as

$$\langle \lambda, y \rangle \geq \langle \lambda, x \rangle + 1 - \gamma_{C_1}(x).$$

Noting that $x \in C_1$ implies that $1 - \gamma_{C_1}(x) \geq 0$ we derive (42). □

Note that the Lagrange multiplier in this example plays the role of the separating hyperplane. The proof given above also works for more general convex functions f and g in composition with a linear mapping. This yields what gets called a sandwich theorem.

Theorem 4.3 (Sandwich Theorem) *Let X and Y be Banach spaces, let f and g be convex lower semicontinuous extended valued functions, and let $A : X \to Y$ be a bounded linear mapping. Suppose that $f \geq -g \circ A$ and*

$$0 \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} f). \tag{45}$$

Then there exists an affine function $\alpha : X \to R$ of the form

$$\alpha(x) = \langle A^* y^*, x \rangle + c$$

satisfying

$$f \geq \alpha \geq -g \circ A.$$

Proof. Consider the corresponding constrained minimization problem

$$v(y, z) = \min\{f(x) + g(Ax + y) - z\} = \min\{f(x) + r : u - Ax = y, g(u) - r \leq z\}. \tag{46}$$

We can see that $v(0, 0) \geq 0$ because $f \geq -g \circ A$. Since v is linear in z it follows that when the constraint qualification condition (45) holds $\partial v(0, 0) \neq \emptyset$.

Thus, by Theorem 2.3, problem (46) has a Lagrange multiplier of the form $(y^*, \mu) \in Y^* \times R_+$ such that, for $x \in X$, $u \in Y$,

$$f(x) + r + \langle y^*, u - Ax \rangle + \mu(g(u) - r) \geq v(0, 0) \geq 0. \tag{47}$$

Letting $u = Ax'$ and $r = g(Ax')$ in (47) we have, for $x, x' \in X$,

$$f(x) - \langle A^* y^*, x \rangle \geq -g(Ax') - \langle A^* y^*, x' \rangle.$$

[6] Since $\gamma_{C_1}(x) - 1 < 0$ for $x \in \text{int}C_1$, $\lambda \neq 0$.

Thus,

$$a := \inf_x \{f(x) - \langle A^*y^*, x\rangle\} \geq b := \sup_{x'}\{-g(Ax') - \langle A^*y^*, x'\rangle\}.$$

Picking any $c \in [a, b]$, the affine function $\alpha(x) = \langle A^*y^*, x\rangle + c$ separates f and $-g \circ A$ as was to be shown. □

We observe that in both of the previous proofs the existence of an optimal primal solution is neither important nor expected. Also, in deriving the separation and the sandwich theorems the (generalized) derivative information regarding the functions involved is not exploited.

An attractive feature of Theorem 4.3 is that it often reduces the study of a linearly constrained convex program to linear program — without any assumption of primal solvability.

4.3 Generalized Linear Complementary Problems

Let X be a reflexive Banach space, X^* its dual and S a closed convex cone in X. Let $T : X \mapsto X^*$ be a closed linear operator and suppose $q \in X^*$. The *generalized linear complementary problem* wants to find $x \in X$ solving

$$\langle Tx - q, x\rangle = 0, \qquad \text{while } x \in S \text{ and } Tx - q \in S^+. \tag{48}$$

As discussed in [8] we can break down this problem into two parts: first we show that for every $\varepsilon > 0$, the problem has an ε-solution ($\langle Tx - q, x\rangle \leq \varepsilon$) and then we add coercivity conditions to ensure that a sequence of $1/n$-solutions will converges to a true solution as n goes to infinity.

Herein, we will only discuss the first part which is closely related to existence of Lagrange multipliers. Following [8] when T is monotone, continuous and linear, we can convert this problem to a convex quadratic programming problem:

$$\min f(x) := \langle Tx - q, x\rangle + \iota_S = \left\langle \frac{T + T^\top}{2} x - q, x \right\rangle + \iota_S \text{ s.t. } q - Tx \leq_{S^+} 0. \tag{49}$$

and consider the perturbed problem of

$$v(y) = \inf\{f(x) : q - Tx \leq_{S^+} y\}.$$

If a Lagrange multiplier $\lambda \in S^{++} = S$ exists then for any $x \in S$ we have

$$\langle Tx - q, x\rangle + \langle \lambda, q - Tx\rangle \geq v(0).$$

In particular, $x = \lambda$ is feasible and we derive

$$0 \geq \langle T\lambda - q, \lambda\rangle + \langle \lambda, q - T\lambda\rangle \geq v(0) \geq 0,$$

that is $v(0) = 0$ or, equivalently, problem (48) always has a ε-solution.

Thus, the existence of an ε-solution to problem (48) is entirely determined by a constraint qualification condition that ensures $\partial v(0) \neq \emptyset$. Several such conditions are elaborated in Section 2 of [8] which we do not repeat here. A point to emphasize here is that in establishing the ε-solution of this problem the simpler version of the Lagrange multiplier rule Theorem 2.1, more precisely Corollary 2.1, suffices.

4.4 Characterization of Quasiconvex Vector-valued Functions

A *quasiconvex* function is a real-valued function whose lower level sets are all convex. This notion can easily be extended to *quasiconvex* vector-valued functions using a partial order induced by a cone: we require that $\{x\colon f(x) \le_K y\}$ is convex for all $y \in Y$. We next provide a characterization of cone-quasiconvexity in terms of scalarization using the Lagrange multiplier rule. Our proof follows [7].

Consider a partially ordered Banach space $(Y, \le_K)$ where K is a nonempty closed convex cone with polar cone $K^+ \subset Y^*$. We use $\operatorname{extd} K^+$ to denote the *extreme directions*: those $y^* \in K^+$ such that any $y_1, y_2 \in K^+$ with $y^* = y_1 + y_2$ must belong to the ray $\{ry^* : r \ge 0\}$. We say that $(Y, \le_K)$ is *directed* if for all $y_1, y_2 \in Y$ there exists $z \in Y$ such that $y_1 \le_K z$ and $y_2 \le_K z$. It is easy to check that $(Y, \le_K)$ is directed if and only if $Y = K - K$ — that is K generates Y.

The key is the following lemma which we shall prove by using the Lagrange multiplier rule.

Lemma 4.1 (Lattice-like Behavior) *Assume that $(Y, \le_K)$ is directed. Let $y^* \in \operatorname{extd} K^+$ and let $y_1, y_2 \in Y$ be such that $\langle y^*, y_i\rangle \le 0, i = 1, 2$. Then for every $\varepsilon > 0$ there exists $z_\varepsilon \in Y$ such that $y_i \le_K z_\varepsilon, (i = 1, 2)$ and $\langle y^*, z_\varepsilon\rangle \le \varepsilon$.*

Proof. Consider the optimization problem

$$v(u_1, u_2) = \inf\{\langle y^*, z\rangle : y_i - z \le_K u_i, i = 1, 2\} = \inf\{\langle y^*, z\rangle : y_i - u_i \le_K z, i = 1, 2\}. \tag{50}$$

We need only to show that the optimal value $v(0, 0) \le 0$. Since $(Y, \le_K)$ is directed, for any $y_i, u_i, i = 1, 2$, we can find $z \in Y$ such that $y_i - u_i \le_K z, (i = 1, 2)$, so that

$$\max(\langle y^*, y_1 - u_1\rangle, \langle y^*, y_2 - u_2\rangle) \le v(u_1, u_2) \le \langle y^*, z\rangle.$$

Thus, $\operatorname{dom} v = Y \times Y$ and a Lagrange multiplier exists for problem (50) when $(u_1, u_2) = (0, 0)$ (using the core constraint qualification of (26) above).

Let $(\lambda_1, \lambda_2) \in K^+ \times K^+$ be the Lagrange multiplier for (50) when $(u_1, u_2) = (0, 0)$. Then we have, for all $z \in Y$,

$$v(0, 0) \le \langle y^*, z\rangle + \langle \lambda_1, y_1 - z\rangle + \langle \lambda_2, y_1 - z\rangle. \tag{51}$$

Since $z \in Y$ in (51) is free we must have

$$y^* = \lambda_1 + \lambda_2. \tag{52}$$

Using the fact that $y^* \in \operatorname{extd} K^+$, we conclude from (52) that $\lambda_1 = t_1 y^*, \lambda_2 = t_2 y^*$ where $t_1, t_2 \ge 0$ and (51) becomes

$$v(0, 0) \le t_1\langle y^*, y_1\rangle + t_2\langle y^*, y_2\rangle \le 0,$$

as asserted. □

Remark 4.2 Note that $v(0, 0)$ is not necessarily attained and we do not need it to be to obtain the desired conclusion. ♢

Now we are ready to prove the following equivalence.

Theorem 4.4 (Scalarization Characterization of Cone-Quasiconvexity) *Assume that $(Y, \le_K)$ is directed and K^+ is the weak-star closed convex hull of $\operatorname{extd} K^+$. Then f is K-quasiconvex if and only if $y^* \circ f$ is quasiconvex for all $y^* \in \operatorname{extd} K^+$.*

Proof. (a) The "if" part. Let $x_i \in \{x\colon f(x) \le_K y\}$, for $i = 1, 2$ and let $t \in [0, 1]$. Then for all $y^* \in \text{extd } K^+$, we have $\langle y^*, f(x_i)\rangle \le \langle y^*, y\rangle$, for $i = 1, 2$. Since $y^* \circ f$ is quasiconvex, we have

$$\langle y^*, f(tx_1 + (1-t)x_2) - y\rangle \le 0. \tag{53}$$

Since (53) holds for all $y^* \in \text{extd } K^+$ and K^+ is the weak-star closed convex hull of extd K^+, we conclude that (53) holds for all $y^* \in K^+$. Thus, $f(tx_1 + (1-t)x_2) - y \in K^{++} = K$ or

$$tx_1 + (1-t)x_2 \in \{x\colon f(x) \le_K y\}.$$

(b) The "only if" part. Fix any $y^* \in \text{extd } K^+$ and $r \in \mathbb{R}$. Let $\varepsilon > 0$ be arbitrary and positive and select $x_i \in \{x\colon y^* \circ f(x) \le r\}$, for $i = 1, 2$ and let $t \in [0, 1]$. Select $y \in Y$ such that $\langle y^*, y\rangle = r$. Applying Lemma 4.1 to $y_i = f(x_i) - y, (i = 1, 2)$, there exists $z_\varepsilon \in Y$ such that $\langle y^*, z_\varepsilon\rangle \le \varepsilon$ and $y_i \le_K z_\varepsilon, (i = 1, 2)$. Since f is K-quasiconvex, $f(tx_1 + (1-t)x_2) - y \le_K z_\varepsilon$ which implies that

$$y^* \circ f(tx_1 + (1-t)x_2) - r \le \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we have

$$y^* \circ f(tx_1 + (1-t)x_2) \le r$$

or $tx_1 + (1-t)x_2 \in \{x\colon y^* \circ f(x) \le r\}$, and we are done. □

We remark that when K has non-empty norm-interior, all the hypotheses of Theorem 4.4 certainly hold. Moreover, when K is the positive orthant in $\mathbb{R}^N$ the equivalence is with quasiconvexity of the coordinate functions.

4.5 Minimax Theorem

Our final example, following [35], deduces the Von Neumann-Fan minimax theorem from the Lagrange multiplier rule. This illustrates the variety of circumstances in which an abstract Lagrange multiplier can be shown to exist and its structure subsequently exploited.

Theorem 4.5 (von Neumann-Fan Minimax Theorem) *Let X and Y be Banach spaces. Let $C \subset X$ be nonempty and convex, and let $D \subset Y$ be nonempty, weakly compact and convex. Let $g : X \times Y \to \mathbb{R}$ be convex with respect to $x \in C$ and concave and (weakly) continuous with respect to $y \in D$. Then*

$$p := \inf_{x\in C} \max_{y\in D} g(x, y) = \max_{y\in D} \inf_{x\in C} g(x, y) =: d \tag{54}$$

Proof. We first note that by weak duality always $p \ge d$ and proceed to show $d \ge p$.

Define a vector function $G : X \times \mathbb{R} \to C(D)$, the Banach space of continuous functions on D with the sup norm, by $G(x, r)(y) := g(x, y) - r$. This is legitimate because g is continuous in the y variable. Let $K \subset C(D)$ be the non-negative continuous functions on D. Since g is convex in x for each $y \in D$ we can check that G is convex with respect to $\le_K$.

Consider the abstract convex program

$$p = \inf_{x\in C}\{r : g(x, y) \le r, \forall y \in D, r \in R\} = \inf_{x\in C, r\in R}\{r : G(x, r) \le_K 0\}. \tag{55}$$

Fix $\varepsilon \in (0, 1)$. Then there is some $\hat{x} \in C$ with $g(\hat{x}, y) \le p + \varepsilon$ for all $y \in D$. We have

$$G(\hat{x}, p + 2) \le_K -1 <_K 0,$$

where $-1, 0$ are constant functions in $C(D)$. Thence, Slater's condition holds for the abstract convex program (55). Consequently, problem (55) has a Lagrange multiplier $\lambda \in C(D)^*_+$. By the Riesz representation theorem (see, e.g., [36,37]) we can treat λ as a measure on D. So we have, for all $x \in C, r \in \mathbb{R}$,

$$r + \langle \lambda, G(x,r) \rangle = r + \int_D (g(x,y) - r)\lambda(dy) \geq p.$$

Since r is arbitrary we must have $\int_D \lambda(dy) = 1$, that is to say that λ is a probability measure. Using *Jensen's inequality* and noticing $\int_D y\lambda(dy) \in D$ we have, for all $x \in C$,

$$g(x, \int_D y\lambda(dy)) \geq \int_D g(x,y)\lambda(dy) = r + \int_D (g(x,y) - r)\lambda(dy) \geq p$$

Taking the infimum of the left hand over $x \in C$ yields $d \geq p$, which completes the proof. □

Note that, even if both C and D lie in finite dimensional space, we still need the abstract form of the Lagrange multiplier in this proof. On the other hand, sometimes in dynamical problems one can use a Lagrange multiplier rule for the finite dimensional constraint and 'propagate' the resulting multipliers along the dynamics. The fully convex problem of calculus of variations discussed in [38] provides such an example.

5 Conclusions

We hope we have persuaded the reader that there is much to be gained by yet again revisiting the subject of Lagrange multipliers. What we forget of past mathematics we frequently have to rediscover—and we may not do as good a job as our predecessors did.

We also emphasise that there is much to be learned from knowing different approaches to establishing multiplier rules. No one approach is uniformly the best and each problem has its own idiosyncrasies. Even when the hypotheses may not apply to the given problem, rough computations may lead one efficiently to the right solution method.

Acknowledgements We thank several of our colleagues for helpful input during the preparation of this work. We especially thank the referee whose incisive and detailed comments have substantially improved the final version of our manuscript.

References

1. J.M. Borwein and Q.J. Zhu, *Techniques of Variational Analysis*, Springer-Verlag, 2005.
2. D. Gale, A geometric duality theorem with economic applications, *The Review of Economic Studies*, **34** (1967), 19–24.
3. H.H. Goldstine, *A History of the Calculus of Variations from the 17th through 19th Century*, Springer-Verlag, 1980.
4. J. Gregory and C. Lin, *Constrained Optimization in the Calculus of Variations and Optimal Control Theory*, Chapman and Hall, 1992.
5. J. Hanca and E.F. Taylor, From conservation of energy to the principle of least action: A story line. *American Journal of Physics*, **72**:4 (2004), 514-521.
6. R. Weinstock, *Calculus of Variations with Applications to Physics and Engineering*, Dover, 1974.
7. J. Benoist, J.M. Borwein and N. Popovici, A characterization of quasiconvex vector-valued functions. *Proc. AMS*, **131** (2002), 1109–1113.
8. J.M. Borwein, Generalized linear complementary problems treated without fixed point theorem, *JOTA* **43** (1984), 343–356.
9. J.M. Borwein and Q.J. Zhu, Viscosity solutions and viscosity subderivatives in smooth Banach spaces with applications to metric regularity, *SIAM Journal on Control and Optimization*, **34** (1996), 1568–1591.

10. J.M. Borwein and A. D. Ioffe, Proximal analysis in smooth Banach spaces, *Set-valued Analysis*, **4** (1996), 1–24.
11. M. Spivak, *Calculus on Manifolds: a modern approach to classical theorem of advanced calculus*, fourth edition, Publish or Perish, 2008.
12. J. Warga, *Optimal Control of Differential and Functional Equations*, Academic Press, 1973.
13. F. Giannessi, *Constrained Optimization and Image Space Analysis* I, Springer-Verlag, 2005.
14. R.T. Rockafellar, Lagrange multipliers and optimality, SIAM Rev. **35** (1993) 183-238.
15. E. R. Avakov, G. G. Magaril-Il'yaev and V. M. Tikhomirov, Lagrange's principle in extremum problems with constraints, Russian Math. Surveys **68** (2013), 401-433.
16. A. D. Ioffe and V. M. Tikhomirov, Theory of Extremal Problems, North-Holland Publishing Co., Amsterdam, New York, 1979.
17. G. G. Magaril-Il'yaev and V. M. Tikhomirov, Newton's method, differential equations, and the Lagrangian principle for necessary extremum conditions, Proc. Steklov Inst. Math. **262** (2008), 149-169.
18. B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation* I, II, Springer-Verlag, 2006.
19. J.-P. Penot, *Calculus without Derivatives*, Springer-Verlag, 2014.
20. J.M. Borwein and D. Zhuang, On Fan's minimax theorem, *Math. Programming*, **34** (1986), 232–234.
21. A. Shapiro and A. Nemirovski, Duality of linear conic problems, Preprint, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 2003. (www.optimization-online.org/DB_HTML/2003/12/793.html).
22. C. Zălinescu, On duality gaps in linear conic problems, Preprint, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 2010. (www.optimization-online.org/DB_HTML/2010/09/2737.html).
23. F. H. Clarke, *Optimization and Nonsmooth Analysis*, Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley & Sons Inc., New York, 1983.
24. V.F. Demyanov and L.V. Vasilev, *Nondifferentiable Optimization* (Translations Series in Mathematics and Engineering), Springer-Verlag, New York, 1986.
25. J.M. Borwein, J.S. Treiman and Q.J. Zhu, Necessary conditions for constrained optimization problem with semicontinuous and continuous data, *Trans. AMS*, **350** (1998), 2409–2429.
26. J.M. Borwein and Q.J. Zhu, A survey of subdifferential calculus with applications, *Nonlinear Analysis, TMA*, **38** (1999), 687–773.
27. T.T.A. Nghia, A nondegenerate fuzzy optimality condition for constrained optimization problems without qualification conditions, *Nonlinear Analysis*, **75** (2012), 6379–6390.
28. Q.J. Zhu, Necessary conditions for constrained optimization problem in smooth Banach spaces with applications, *SIAM J. Optimization*, **12** (2002), 1032–1047.
29. O.L. Mangasarian and S. Fromovitz, The Fritz John necessary conditions in the presence of equality and inequality constraints, *J. Math. Anal. Appl.* **17** (1967), 37-47.
30. A. Y. Kruger, L. Minchenko, J. V. Outrata, On relaxing the Mangasarian-Fromovitz constraint qualification, *Positivity*, **18** (2014), 171-189.
31. J.M. Borwein, Convex relations in optimization and analysis, pp. 335–377, in: *Generalized Convexity in Optimization and Economics,* Academic Press, (1981).
32. J.M. Borwein, Maximum entropy and feasibility methods for convex and non-convex inverse problems. *Optimization*, **61** (2012), 1–33. DOI: http://link.springer.com/article/10.1080/02331934.2011.632502.
33. J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization*, Springer-Verlag, 2000. Second edition, 2005.
34. A.D. Ioffe and V.L. Levin, Subdifferentials of convex functions. *Trudy Moskov. Mat. Obšč.*, **26** (1972), 3–73.
35. J.M. Borwein, A very complicated proof of the minimax theorem, *Minimax Theory and its Applications.* **1**:1 (2015), xx-xx.
36. G.J.O. Jameson, *Topology and Normed Spaces*, Chapman and Hall, 1974.
37. D.G. Luenberger, *Optimization by Vector Space Methods*, John Wiley and Sons, 1972.
38. R.T. Rockafellar, R. T. and P.R. Wolenski, Convexity in Hamilton-Jacobi Theory 1: Dynamics and Duality, *SIAM J. Control and Optimization*, **40** (2001), 1323–1350.