Set-Valued Analysis (2001)

PRIMAL-DUAL SOLUTION PERTURBATIONS IN CONVEX OPTIMIZATION

A. L. Dontchev[1] **and R. T. Rockafellar**[2 3]

Abstract. Solutions to optimization problems of convex type are typically characterized by saddle point conditions in which the primal vector is paired with a dual 'multiplier' vector. This paper investigates the behavior of such a primal-dual pair with respect to perturbations in parameters on which the problem depends. A necessary and sufficient condition in terms of certain matrices is developed for the mapping from parameter vectors to saddle points to be single-valued and Lipschitz continuous locally. It is shown that the saddle point mapping is then semi-differentiable, and that its semi-derivative at any point and in any direction can be calculated by determining the unique solutions to an auxiliary problem of extended linear-quadratic programming and its dual. A matrix characterization of calmness of the solution mapping is provided as well.

Keywords. Parametric convex optimization, saddle points, sensitivity analysis, stability, solution perturbations, semi-derivatives, proto-derivatives, variational analysis

April 2000

[1] Mathematical Reviews, Ann Arbor, Mi 4801-8604

[2] Department of Mathematics, University of Washington, Seattle, WA 98195-4350

[3] This research was undertaken under grant DMS–9803089 from the National Science Foundation. Support is gratefully acknowledged also from the FONDAP program in Applied Mathematics at the University of Chile during the second author's stay there, when a central part of the work was carried out.

1. Introduction

Saddle point expressions of optimality are a characteristic feature in convex optimization. They pair the solution to a given primal problem with an auxiliary vector which solves a dual problem. In the study of what happens when parameter values are perturbed, it is important to consider the effects on both of these vectors. This paper is devoted to the study of such effects by methods of variational analysis [1], especially through application of results that have recently been obtained for variational inequalities [2].

In the primal-dual framework we adopt, there are convex sets $X \subset I\!R^n$ and $Y \subset I\!R^m$ (nonempty) along with an open set $W \subset I\!R^d$ and a function $L : W \times I\!R^n \times I\!R^m \to I\!R$ such that $L(w,x,y)$ is concave with respect to $y \in Y$ for each $(w,x) \in W \times X$ and convex with respect to $x \in X$ for each $(w,y) \in W \times Y$. Regarding $w \in W$ as a parameter, we associate with these data elements a *primal* problem of optimization

$\mathcal{P}(w)$ minimize $f(w,x)$ over $x \in X$, where $f(w,x) := \sup_Y L(w,x,\cdot)$,

a *dual problem* of optimization

$\mathcal{D}(w)$ maximize $g(w,y)$ over $y \in Y$, where $g(w,y) := \inf_X L(w,\cdot,y)$,

and a *primal-dual* problem

$\mathcal{PD}(w)$ minimaximize $L(w,x,y)$ over $(x,y) \in X \times Y$,

a solution to this being by definition a pair (x,y) that is a *saddle point* of L with respect to minimizing over X and maximizing over Y, i.e., satisfies

$$x \in \operatorname{argmin}_X L(w,\cdot,y), \qquad y \in \operatorname{argmax}_Y L(w,x,\cdot). \tag{1.1}$$

Note that $f(w,x)$ is convex in $x \in X$ in $\mathcal{P}(w)$, while $g(w,y)$ is concave in $y \in Y$ in $\mathcal{D}(w)$.

The general relationships in such a triad of problems are elementary and well known. If a solution to $\mathcal{PD}(w)$ exists, the optimal values in $\mathcal{P}(w)$ and $\mathcal{D}(w)$ (the infimum and supremum, respectively) must be equal. On other hand, as long as these values are equal, a pair (x,y) solves $\mathcal{PD}(w)$ if and only if x solves $\mathcal{P}(w)$ and y solves $\mathcal{D}(w)$, in which case both optimal values equal $L(x,y)$. These facts are supplemented by criteria in minimax theory and duality theory for the existence of a solution to $\mathcal{PD}(w)$ or the equality of the optimal values in $\mathcal{P}(w)$ and $\mathcal{D}(w)$; cf. [1; Chap. 11] for a recent treatment.

Here, taking such facts as background, we go straight to the analysis of the (generally set-valued) *saddle point mapping*

$$S : W \rightrightarrows X \times Y \text{ with } S(w) := \big\{ \text{solutions } (x,y) \text{ to } \mathcal{PD}(w) \big\}. \tag{1.2}$$

Our goal is to shed light on the behavior of $S(w)$ with respect to changes in w, especially questions of local single-valuedness, Lipschitz continuity and semi-differentiability.

Example 1.1 (ordinary convex programming). *Suppose $X = I\!R^n$, $Y = I\!R^s_+ \times I\!R^{m-s}$ and, with $y = (y_1, \ldots, y_m)$, that*

$$L(w,x,y) = f_0(w,x) + y_1 f_1(w,x) + \cdots + y_m f_m(w,x) \tag{1.3}$$

for $\mathcal{C}^2$ functions f_i on $W \times I\!R^n$ such that, for each $w \in W$, $f_i(w,x)$ is convex in x for $i = 0, 1, \ldots, s$ but affine in x for $i = s+1, \ldots, m$. Then problem $\mathcal{P}(w)$ takes the form

$$\text{minimize } f_0(w,x) \text{ in } x \text{ subject to } f_i(w,x) \begin{cases} \leq 0 & \text{for } i = 1, \ldots, s, \\ = 0 & \text{for } i = s+1, \ldots, m, \end{cases}$$

and the saddle point condition (1.1) corresponds to the Karush-Kuhn-Tucker conditions.

Example 1.2 (extended convex programming). *Let X and Y be polyhedral sets and let*

$$L(w,x,y) = f_0(w,x) + y_1 f_1(w,x) + \cdots + y_m f_m(w,x) - k(w,y) \tag{1.4}$$

for $\mathcal{C}^2$ functions f_i on $W \times I\!R^n$ and k on $W \times I\!R^m$ with $k(w,y)$ concave in $y \in Y$. Suppose that for each $y \in Y$ the expression $f_0(w,x) + y_1 f_1(w,x) + \cdots + y_m f_m(w,x)$ is convex in $x \in X$. Then all assumptions are fulfilled, and problem $\mathcal{P}(w)$ comes out as

$$\begin{aligned} &\text{minimize } \; f_0(w,x) + \theta\big(w, f_1(w,x), \ldots, f_m(w,x)\big) \\ &\text{over } x \in X, \;\; \text{where } \theta(w,u) := \sup_{y \in Y} \big\{ \langle u, y \rangle - k(w,y) \big\}. \end{aligned}$$

This fits with the model of *extended nonlinear programming* proposed in [3] (see [4] also), where the θ term can take on ∞ and thereby represent additional constraints beyond $x \in X$. Ordinary nonlinear programming is obtained by taking X and Y as in Example 1.1 and setting $k(w,y) \equiv 0$. The motivation for this type of extension is explained at length in [3] and will not be repeated here. A special case will be central to our endeavor, however.

Example 1.3 (extended linear-quadratic programming). *Again let X and Y be polyhedral sets, but take*

$$L(w,x,y) = \langle c(w), x \rangle + \tfrac{1}{2} \langle C(w)x, x \rangle + \langle b(w), y \rangle - \tfrac{1}{2} \langle B(w)y, y \rangle - \langle A(w)x, y \rangle,$$

where the matrices $C(w)$ and $B(w)$ are symmetric and positive semi*definite (at least in relation to X and Y; see below). Then the primal problem $\mathcal{P}(w)$ is to*

$$\begin{aligned} &\text{minimize } \; \langle c(w), x \rangle + \tfrac{1}{2} \langle C(w)x, x \rangle + \theta_{Y,B(w)}\big(b(w) - A(w)x\big) \\ &\text{over } x \in X, \;\; \text{where } \; \theta_{Y,B(w)}(u) := \sup_{y \in Y} \big\{ \langle u, y \rangle - \tfrac{1}{2} \langle B(w)y, y \rangle \big\}, \end{aligned}$$

whereas the dual problem $\mathcal{D}(w)$, involving the transpose matrix $A(w)^T$ is to

$$\text{maximize} \quad \langle b(w), y\rangle - \tfrac{1}{2}\langle B(w)y, y\rangle - \theta_{X,C(w)}\big(A(w)^T y - c(w)\big)$$
$$\text{over } y \in Y, \ \text{ where } \ \theta_{X,C(w)}(v) := \sup_{x\in X}\big\{\langle v, x\rangle - \tfrac{1}{2}\langle C(w)x, x\rangle\big\}.$$

Extended linear-quadratic programming goes back to [5] and [6] and has a role in approximating other problems of optimization, as well as in furnishing a problem model that is useful in itself. It will serve us below in formulas for the calculation of solution perturbations for our general problems $\mathcal{P}(w)$, $\mathcal{D}(w)$, and $\mathcal{PD}(w)$.

Because everything only depends on the values of L on $W \times X \times Y$, it is not really essential, in the context of extended linear-quadratic programming problems of convex type in Example 1.3, for the matrices $C(w)$ and $B(w)$ to be positive semidefinite in full. They merely have to yield quadratic forms $x \mapsto \langle C(w)x, x\rangle$ and $y \mapsto \langle C(w)y, y\rangle$ that give convex functions on X and Y, respectively, which is a slightly weaker requirement when X or Y might have empty interior. This is what was meant in Example 1.3 by $C(w)$ and $B(w)$ being *positive semidefinite in relation to X and Y*. (In the case of $C(w)$ and X, the weaker property is equivalent to the following: the matrix P giving the projection of $\mathbb{R}^n$ onto the subspace parallel to the affine hull of X should make $P^T C(w) P$ positive semidefinite.) This generalization will come into play in our perturbation formulas, where the convex sets associated with certain derived subproblems might well have empty interior.

Special duality results are available for extended linear-quadratic programming problems of convex type. No constraint qualification is needed in relating primal and dual solutions to saddle points; $S(w)$ is always the product set consisting of the pairs (x, y) such that x solves the primal problem and y solves the dual problem. Indeed, the infimum in the primal problem equals the supremum in the dual problem unless the former is ∞ and the latter is $-\infty$, i.e., neither problem has a feasible solution; see [5] and [1; 11.42, 11.43]. (The cited results are stated for positive semidefinite matrices but carry over to positive semidefiniteness in relation to X and Y in the manner just explained.)

Our analysis of the saddle point mapping S rests on interpreting the saddle point condition (1.1) as a special kind of variational inequality over $X \times Y$ in terms of the normal cones $N_X(x)$ and $N_Y(y)$ at points $x \in X$ and $y \in Y$ (in the sense of convex analysis). To assist technically in this, we make the following restriction.

Assumption. *It will be supposed throughout the paper that X and Y are polyhedral and that L is a $\mathcal{C}^2$ function.*

This will open the way for utilizing certain results from our paper [2] on smooth variational inequalities over polyhedral sets.

Proposition 1.4 (variational inequality format). *The saddle point mapping S, which is convex-valued, has $(x, y) \in S(w)$ if and only if*

$$-\nabla_x L(w, x, y) \in N_X(x), \qquad \nabla_y L(w, x, y) \in N_Y(y). \tag{1.5}$$

This corresponds to (x, y) solving a variational inequality over $X \times Y$, namely

$$\begin{aligned} &G(w, x, y) + N_{X \times Y}(x, y) \ni (0, 0), \;\; \textit{where} \\ &\qquad G(w, x, y) := \big(\nabla_x L(w, x, y),\, -\nabla_y L(w, x, y)\big). \end{aligned} \tag{1.6}$$

Moreover, this variational inequality is one in which the set $X \times Y$ is polyhedral, the mapping G is of class $\mathcal{C}^1$, and $G(w, \cdot, \cdot)$ is monotone on $X \times Y$:

$$\begin{aligned} &\big\langle G(w, x_1, y_1) - G(w, x_0, y_0),\, (x_1, y_1) - (x_0, y_0) \big\rangle \geq 0 \\ &\qquad \textit{for all } (x_0, y_0) \in X \times Y, \;\; (x_1, y_1) \in X \times Y. \end{aligned} \tag{1.7}$$

Proof. The convex-valuedness of S comes from $S(w)$ being (if not empty) the product of the optimal solution sets to $\mathcal{P}(w)$ and $\mathcal{D}(w)$. Those solution sets are convex because the problems belong to the category of convex optimization.

The first condition in (1.1) is known always to imply the first condition in (1.5). The converse implication would be immediate from standard convex analysis if $L(w, x, y)$ had been assumed convex over all $x \in I\!R^n$ instead of just relative to $x \in X$. The extension to the slightly more general case is easy to make. The basic convexity inequality $L(w, x', y) \geq L(w, x, y) + \langle \nabla_x L(w, x, y),\, x' - x \rangle$ holds for any two points x and x' in X, regardless of an possible lack of convexity of $L(w, \cdot, y)$ outside of X, because it comes out the convexity of $\varphi(t) = L(w, (1-t)x + tx', y)$ relative to $t \in [0, 1]$ and the consequent monotonicity $\varphi'(t)$ on that interval. The normal cone $N_X(x)$ consists of the vectors v such that $\langle v,\, x' - x \rangle \leq 0$ for all $x' \in X$. Therefore, the relation $-\nabla_x L(w, x, y) \in N_X(x)$ means that $\langle \nabla_x L(w, x, y),\, x' - x \rangle \geq 0$ for all $x' \in X$. This inequality, in combination with the basic convexity inequality above, yields $L(w, x', y) \geq L(w, x, y)$ for all $x' \in X$. Hence if $-\nabla_x L(w, x, y) \in N_X(x)$ we have $x \in \operatorname{argmin}_X L(w, \cdot, y)$.

Similar observations confirm the equivalence of the second conditions in (1.1) and (1.5) as well as the monotonicity inequality in (1.7). The equivalence between (1.5) and the variational inequality in (1.6) is immediate from $N_{X \times Y}(x, y) = N_X(x) \times N_Y(y)$. The Assumption above makes $X \times Y$ polyhedral and G of class $\mathcal{C}^1$. $\square$

2. Graphical Derivatives and Their Calculation

The *graph* of the saddle point mapping S is the set $\operatorname{gph} S \subset \mathbb{R}^d \times \mathbb{R}^n \times \mathbb{R}^m$ consisting of all (w, x, y) such that $w \in W$ and $(x, y) \in S(w)$. The local properties of $\operatorname{gph} S$ around one of its points (w_*, x_*, y_*) will be our central concern. Eventually we will identify the circumstances in which S is single-valued on a neighborhood of w_* with (x_*, y_*) as the sole element of $S(w_*)$, and then the analysis of S can have a familiar tone, although differentiability must often be replaced by "semi-differentiability." But much can be said in the absence of such single-valuedness, and even about generalized derivatives, despite a serious lack of continuity often in the behavior of S.

The key to this is a fundamental fact about the geometry of $\operatorname{gph} S$ around one of its points (w_*, x_*, y_*), which holds under the following condition of *ample parameterization*:

$$\operatorname{rank}\left[\nabla^2_{wx} L(w_*, x_*, y_*) \,\middle|\, \nabla^2_{wy} L(w_*, x_*, y_*)\right] = n + m. \tag{2.1}$$

This condition on the $d \times (n+m)$ matrix $\left[\nabla^2_{wx} L(w_*, x_*, y_*) \,\middle|\, \nabla^2_{wy} L(w_*, x_*, y_*)\right]$ is more a normalization of the model than a serious assumption. The vector w could always be augmented by vectors $v \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$ in the pattern of the parameterized saddle point problem

$$\begin{aligned} &\text{minimaximize } L^+(u, v, w, x, y) \text{ over } (x, y) \in X \times Y, \\ &\text{where } L^+(u, v, w, x, y) := L(w, x, y) + \langle v, x\rangle + \langle u, y\rangle. \end{aligned} \tag{2.2}$$

Such an augmented parameterization by $w^+ = (u, v, w) \in W^+ = \mathbb{R}^m \times \mathbb{R}^n \times W$ surely satisfies (2.1). In working with (2.1), instead of explicitly with auxiliary vectors u and v giving the so-called *canonical primal and dual* perturbations as would be the traditional pattern, we are able to formulate our results more succinctly as well more generally.

The concept of ample parameterization was introduced in [2] in a broader framework of variational inclusions. We employ it now in stating some basic but powerful facts about the saddle point mapping S which specialize certain of our results in [2].

Theorem 2.1 (Lipschitzian graphical geometry)**.** *Let $(x_*, y_*) \in S(w_*)$. If the ample parameterization condition (2.1) is satisfied at (w_*, x_*, y_*), then S is graphically Lipschitzian of dimension d at w_* for (x_*, y_*), in the sense that there is a change of coordinates around $(w_*, x_*, y_*) \in \mathbb{R}^{d+n+m}$ under which $\operatorname{gph} S$ can be identified locally with the graph of a single-valued Lipschitz continuous mapping from $\mathbb{R}^d$ to $\mathbb{R}^{n+m}$.*

Proof. We apply [2; Thm. 7.1] to the variational inequality format in Proposition 1.4. □

In the graphical context of Theorem 2.1, there is a geometric kind of generalized differentiation that can be used even though S may only be set-valued. One says that S

is *proto-differentiable at* w_* *for* (x_*, y_*) when $(x_*, y_*) \in S(w_*)$ and the difference quotient mappings

$$\Delta_\tau S(w_* \,|\, x_*, y_*) : w' \;\mapsto\; \tau^{-1}[S(w_* + \tau w') - (x_*, y_*)], \quad \text{where } \tau > 0,$$

converge graphically as $\tau \searrow 0$; in other words, there is a mapping $T : I\!R^d \rightrightarrows I\!R^n \times I\!R^m$ such that $\operatorname{gph} \Delta_\tau S(w_* \,|\, x_*, y_*)$ converges to $\operatorname{gph} T$ as $\tau \searrow 0$. When this is true, $\operatorname{gph} T$ must be the tangent cone to $\operatorname{gph} S$ at (w_*, x_*, y_*); the symbol for the mapping T with this tangent cone as its graph is $DS(w_* \,|\, x_*, y_*)$. In fact, proto-differentiability of S at w_* for (x_*, y_*) is equivalent to $\operatorname{gph} S$ being *geometrically derivable* at (w_*, x_*, y_*) in the sense that the outer set limit defining the tangent cone coincides with the inner set limit. For more about the meaning proto-differentiability and its many features, see [1].

Another concept of importance to us now is that of a mapping D being *piecewise polyhedral*, which means that $\operatorname{gph} D$ is the union of a finite collection of polyhedral (convex) sets. Piecewise linear mappings are known to be the piecewise polyhedral mappings that are single-valued [1; 2.48, 9.57].

Theorem 2.2 (saddle point proto-differentiability). *Let* $(x_*, y_*) \in S(w_*)$. *If the ample parameterization condition (2.1) is satisfied at* (w_*, x_*, y_*), *then* S *is proto-differentiable at* w_* *for* (x_*, y_*) *and the proto-derivative mapping* $DS(w_* \,|\, x_*, y_*)$ *is piecewise polyhedral.*

Proof. Again we simply apply [2; Theorem 7.1] to the variational inequality format in Proposition 1.4. □

The particular nature of the variational inequality in Proposition 1.4 as a representation of saddle points makes a difference now, as we specialize further from [2].

Theorem 2.3 (perturbation formula). *In the framework of Theorem 2.2, the mapping* $DS(w_* \,|\, x_*, y_*)$ *has the following description:* $(x', y') \in DS(w_* \,|\, x_*, y_*)(w')$ *if and only if* (x', y') *is a solution to the saddle point problem*

$$\mathcal{PD}_*(w') \qquad\qquad \textit{minimaximize } L_*(w', x', y') \textit{ over } (x', y') \in X_* \times Y_*,$$

that corresponds to the data elements

$$\begin{gathered} L_*(w', x', y') = \langle E_* w', x' \rangle + \langle D_* w', y' \rangle + \tfrac{1}{2}\langle C_* x', x' \rangle - \tfrac{1}{2}\langle B_* y', y' \rangle - \langle A_* x', y' \rangle, \\ X_* = T_X(x_*) \cap \nabla_x L(w_*, x_*, y_*)^\perp, \qquad Y_* = T_Y(y_*) \cap \nabla_y L(w_*, x_*, y_*)^\perp, \end{gathered} \tag{2.3}$$

where $T_X(x_*)$ *and* $T_Y(y_*)$ *are the tangent cones to* X *and* Y *at the points* x_* *and* y_*, $\nabla_x L(w_*, x_*, y_*)^\perp$ *and* $\nabla_y L(w_*, x_*, y_*)^\perp$ *denote the subspaces orthogonal to the gradients* *in question, and the matrices in the expression of L_* are specified by*

$$\begin{aligned} &E_* = \nabla^2_{xw} L(w_*, x_*, y_*), \quad D_* = \nabla^2_{yw} L(w_*, x_*, y_*), \\ &C_* = \nabla^2_{xx} L(w_*, x_*, y_*), \quad B_* = -\nabla^2_{yy} L(w_*, x_*, y_*), \quad A_* = -\nabla^2_{yx} L(w_*, x_*, y_*). \end{aligned} \tag{2.4}$$

Here X_* and Y_* are polyhedral (convex) cones, and $L_*(w', x', y')$ is convex in $x' \in X_*$ for each $(w', y') \in \mathbb{R}^d \times Y_*$, but concave in $y' \in Y_*$ for each $(w', x') \in \mathbb{R}^d \times X_*$.

Proof. Because X and Y are polyhedral, the tangent cones $T_X(x_*)$ and $T_Y(y_*)$ are polyhedral and the cones X_* and Y_* are polyhedral. The convexity-concavity of L_* on $X_* \times Y_*$ is immediate from that of L on $X \times Y$: the convexity of $L(w, x, y)$ with respect to $x \in X$ implies the positive semidefiniteness of C_* relative to the subspace parallel to the affine hull of X, which of course includes X_*; likewise, B_* is positive semidefinite relative to Y_*.

The derivative formula will comes out of [2; Theorem 7.1], but we need to work a bit to obtain the details of the specialization. The general formula in the cited theorem, in preliminary translation to the format of the variational inequality in Proposition 1.4, is

$$\begin{aligned} DS(w_* \,|\, x_*, y_*)(w') &= \big\{ (x', y') \,\big|\, g(w', x', y') + N_{K_*}(x', y') \ni 0 \big\}, \text{ where} \\ g(w', x', y') &= \nabla_w G(w_*, x_*, y_*) w' + \nabla_x G(w_*, x_*, y_*) x' + \nabla_y G(w_*, x_*, y_*) y', \\ K_* &= T_{X \times Y}(x_*, y_*) \cap G(w_*, x_*, y_*)^{\perp}. \end{aligned} \tag{2.5}$$

Since $G(w, x, y) = (\nabla_x L(w, x, y), -\nabla_y L(w, x, y))$ and $T_{X \times Y}(x_*, y_*) = T_X(x_*) \times T_Y(y_*)$ we evidently have $K_* = X_* \times Y_*$ and, in the notation (2.4),

$$\begin{aligned} \nabla_w G(w_*, x_*, y_*) w' &= (E_* w', -D_* w'), \\ \nabla_x G(w_*, x_*, y_*) x' &= (C_* x', A_* x'), \\ \nabla_y G(w_*, x_*, y_*) y' &= (-A_*^T y', B_* y'). \end{aligned}$$

According to (2.5), therefore, a pair (x', y') belongs to $DS(w_* \,|\, x_*, y_*)(w')$ if and only if

$$-E_* w' - C_* x' + A_*^T y' \in N_{X_*}(x'), \qquad -D_* w' + A_* x' + B_* y' \in N_{Y_*}(y').$$

These conditions have the form

$$-\nabla_{x'} L_*(w', x', y') \in N_{X_*}(x'), \qquad \nabla_{y'} L_*(w', x', y') \in N_{Y_*}(y').$$

In view of the convexity properties of L_*, X_* and Y_*, this means by Proposition 1.4 (as applied to these elements) that $L_*(w', \cdot, \cdot))$ has a saddle point over $X_* \times Y_*$ at (x', y'). $\square$

Note that the extended linear-quadratic programming problems referred to in Theorem 2.3 fit the pattern in Example 1.3 as extended to positive semidefiniteness only in relation to the sets X and Y rather than all of $\mathbb{R}^n$ and $\mathbb{R}^m$.

Corollary 2.4 (auxiliary problems)**.** *In the framework of Theorem 2.2, the saddle point problem $\mathcal{PD}_*(w')$ corresponds to a primal-dual pair of extended linear-quadratic programming problems of convex type. The elements of $DS(w_* \,|\, x_*, y_*)(w')$ are the pairs (x', y') such that x' solves the primal problem*

$$\mathcal{P}_*(w') \qquad \begin{array}{l} \text{minimize } \ \langle E_* w', x'\rangle + \frac{1}{2}\langle C_* x', x'\rangle + \theta_{Y_*,B_*}\big(D_* w' - A_* x'\big) \\ \text{over } x' \in X_*, \ \text{ where } \ \theta_{Y_*,B_*}(u') := \sup_{y' \in Y_*} \big\{\langle u', y'\rangle - \frac{1}{2}\langle B_* y', y'\rangle\big\} \end{array}$$

while y' solves the corresponding dual problem

$$\mathcal{D}_*(w') \qquad \begin{array}{l} \text{maximize } \ \langle D_* w', y'\rangle - \frac{1}{2}\langle B_* y', y'\rangle - \theta_{X_*,C_*}\big(A_*^T y' - E_* w'\big) \\ \text{over } y' \in Y_*, \ \text{ where } \ \theta_{X_*,C_*}(v') := \sup_{x' \in X_*} \big\{\langle v', x'\rangle - \frac{1}{2}\langle C_* x', x'\rangle\big\}. \end{array}$$

In particular, the set $DS(w_* \,|\, x_*, y_*)(w')$ is always convex.

Proof. This is immediate from the convexity properties at the end of Theorem 2.3 and the facts about duality in extended linear-quadratic programming that were reviewed after Example 1.3. □

Theorem 2.3 yields additional information also about the special geometric nature of the graph of $DS(w_* \,|\, x_*, y_*)$.

Corollary 2.5 (proto-derivative geometry)**.** *In the framework of Theorem 2.2, the graph of $DS(w_* \,|\, x_*, y_*)$ is a piecewise linear manifold of dimension d in the sense of being a Lipschitzian manifold formed as the union of a finite collection of d-dimensional polyhedral subsets of $\mathbb{R}^d \times \mathbb{R}^n \times \mathbb{R}^m$.*

Proof. Theorem 2.3 reveals that $DS(w_* \,|\, x_*, y_*)$ is a mapping of the kind to which Theorem 2.1 is applicable. For that reason, $\operatorname{gph} DS(w_* \,|\, x_*, y_*)$ is a d-dimensional Lipschitzian manifold, in fact "globally" because this graph is a cone and therefore determined by its properties around the origin. On the other hand, $DS(w_* \,|\, x_*, y_*)$ is piecewise polyhedral by Theorem 2.2. That supplies the piecewise linearity; in expressing the graph as the union of a finite collection of polyhedral sets, it can be arranged that none of these sets is included in any of the others, and they must then all be of dimension d. □

3. Lipschitzian Single-Valuedness in Perturbations

Especially of interest in applications is the possibility of S being single-valued and Lipschitz continuous around w_*. A definitive characterization of that case can be given, and it furnishes a stronger property of differentiation. Here we denote by $o(|z|)$ a term with the property that $o(|t|)/t \to 0$ as $t \to 0$, $t > 0$.

Theorem 3.1 (single-valuedness and semi-differentiability). *Let $(x_*, y_*) \in S(w_*)$ and suppose the ample parameterization condition (2.1) is satisfied at (w_*, x_*, y_*). Then the following properties are equivalent:*

(a) *S is single-valued and Lipschitz continuous on some neighborhood of w_*,*

(b) *$DS(w_* \,|\, x_*, y_*)$ is single-valued everywhere.*

Moreover, in this case the mapping S is semi-differentiable at w_* for (x_*, y_*) in the sense that

$$S(w_* + w') = S(w_*) + DS(w_* \,|\, x_*, y_*)(w') + o(|w'|), \tag{3.1}$$

and the positively homogeneous mapping $DS(w_* \,|\, x_*, y_*)$ is not only Lipschitz continuous but piecewise linear.

Proof. This is immediate from specializing [2; Thm. 7.4] to the variational inequality in Proposition 1.4; that Proposition also provides the convex-valuedness required by the cited theorem. □

Note that for the single-valuedness of $DS(w_* \,|\, x_*, y_*)$ to hold globally in (b) of Theorem 3.1 it would suffice to have it hold on a neighborhood of the origin. That follows from the positive homogeneity of the mapping $DS(w_* \,|\, x_*, y_*)$.

It will now be demonstrated that the important properties in Theorems 3.1 hold if and only if the matrices A_*, B_* and C_* in (2.4) exhibit a kind of nonsingularity relative to certain linear subspaces of $I\!R^n$ and $I\!R^m$ which are naturally associated with the convex cones X_* and Y_* in (2.3).

Theorem 3.2 (nonsingularity criterion). *Let $(x_*, y_*) \in S(w_*)$ and suppose that the ample parameterization condition (2.1) is satisfied at (w_*, x_*, y_*). Then for S to be single-valued and Lipschitz continuous on a neighborhood of w_* it is necessary and sufficient that the following conditions hold for the matrices in (2.4) in terms of the convex cones in (2.3):*

$$\begin{cases} x' \in X_* - X_*, \quad C_* x' = 0, \quad A_* x' \in [Y_* \cap -Y_*]^\perp \quad \Longrightarrow \quad x' = 0, \\ y' \in Y_* - Y_*, \quad B_* y' = 0, \quad A_*^T y' \in [X_* \cap -X_*]^\perp \quad \Longrightarrow \quad y' = 0. \end{cases} \tag{3.2}$$

(Here $X_* - X_*$ and $Y_* - Y_*$ are the smallest subspaces that include X_* and Y_*, respectively, whereas $X_* \cap -X_*$ and $Y_* \cap -Y_*$ are the largest subspaces included within X_* and Y_*.)

Proof. In view of Theorem 3.1, we need only demonstrate that (3.2) is necessary and sufficient for the single-valuedness of the mapping $T = DS(w_* \,|\, x_*, y_*)$ relative to its domain. We know that T is piecewise polyhedral (Theorem 2.2) and convex-valued (Corollary 2.4). For such T and a point $(w'_*, x'_*, y'_*) \in \operatorname{gph} T$, saying that (x'_*, y'_*) is the unique element of $T(w'_*)$ is equivalent to saying that the tangent cone to $\operatorname{gph} T$ at (w'_*, x'_*, y'_*), contains no vector of the form $(0, x''_*, y''_*)$, or in other words, in terms of graphical derivatives, that

$$DT(w'_* \,|\, x'_*, y'_*)(0) = \{(0,0)\}. \tag{3.3}$$

Our task therefore is to demonstrate that (3.2) is necessary and sufficient for (3.3) to hold for all choices of (w'_*, x'_*, y'_*) with $(x'_*, y'_*) \in T(w'_*)$.

By Theorem 2.3, T is a mapping of the same saddle point form as S, except that L, X and Y have been replaced by L_*, X_* and Y_*. Hence we can apply Theorem 2.3 to T in place of S and get the interpretation that the elements (x'', y'') of $DT(w'_* \,|\, x'_*, y'_*)(w'')$ are the saddle points of a certain function L_{**} on a product of polyhedral cones X_{**} and Y_{**}, namely

$$\begin{aligned} X_{**} &= T_{X_*}(x'_*) \cap \nabla_{x'} L_*(w'_*, x'_*, y'_*)^\perp, \\ Y_{**} &= T_{Y_*}(y'_*) \cap \nabla_{y'} L_*(w'_*, x'_*, y'_*)^\perp, \end{aligned} \tag{3.4}$$

where the gradients are given by

$$\begin{aligned} \nabla_{x'} L_*(w'_*, x'_*, y'_*) &= E_* w'_* + C_* x'_* - A_*^T y'_*, \\ \nabla_{y'} L_*(w'_*, x'_*, y'_*) &= D_* w'_* - B_* y'_* - A_* x'_*. \end{aligned} \tag{3.5}$$

Although these cones depend on (w'_*, x'_*, y'_*), the function L_{**} does not; we simply have $L_{**} = L_*$, inasmuch as L_* is quadratic. The saddle point condition therefore translates, by way of Corollary 2.4, to an assertion about solutions to a pair of extended linear-quadratic programming problems in which X_{**} and Y_{**} substitute for X_* and Y_*. Specifically, in the case of $w'' = 0$ that is our focus, we have $(x'', y'') \in DT(w'_* \,|\, x'_*, y'_*)(0)$ if and only if x'' solves the primal problem

$$\mathcal{P}_{**} \qquad \begin{aligned} &\text{minimize } \ \tfrac{1}{2}\langle C_* x'', x''\rangle + \theta_{Y_{**},B_*}(-A_* x'') \ \text{ over } x'' \in X_{**}, \\ &\text{where } \ \theta_{Y_{**},B_*}(u'') := \sup_{y'' \in Y_{**}} \big\{\langle u'', y''\rangle - \tfrac{1}{2}\langle B_* y'', y''\rangle\big\}, \end{aligned}$$

while y'' solves the corresponding dual problem

$$\mathcal{D}_{**} \qquad \begin{aligned} &\text{maximize } \ -\tfrac{1}{2}\langle B_* y'', y''\rangle - \theta_{X_{**},C_*}(A_*^T y'') \ \text{ over } y'' \in Y_{**}, \\ &\text{where } \ \theta_{X_{**},C_*}(v'') := \sup_{x'' \in X_{**}} \big\{\langle v'', x''\rangle - \tfrac{1}{2}\langle C_* x'', x''\rangle\big\}. \end{aligned}$$

The problems $\mathcal{P}_{**}$ and $\mathcal{D}_{**}$, like the cones X_{**} and Y_{**}, depend on (w'_*, x'_*, y'_*) through (3.4). We have to show that the case in which these problems have only the solutions $x'' = 0$ and $y'' = 0$, no matter how (w'_*, x'_*, y'_*) is chosen from $\operatorname{gph} T$, is the case where (3.2) holds.

Let $X_*^+ = X_* - X_*$ and $Y_*^+ = Y_* - Y_*$, and on the other hand $X_*^- = X_* \cap -X_*$ and $Y_*^- = Y_* \cap -Y_*$. These subspaces not only satisfy $X_*^- \subset X_* \subset X_*^+$ and $Y_*^- \subset Y_* \subset Y_*^+$, but on the basis of the definitions in (3.4) also

$$X_*^- \subset X_{**} \subset X_*^+, \qquad\qquad Y_*^- \subset Y_{**} \subset Y_*^+, \tag{3.6}$$

and accordingly they furnish the function bounds

$$\theta_{Y_*^-,B_*} \le \theta_{Y_{**},B_*} \le \theta_{Y_*^+,B_*}, \qquad\qquad \theta_{X_*^-,C_*} \le \theta_{X_{**},C_*} \le \theta_{X_*^+,C_*}. \tag{3.7}$$

Note also that X_*^+ is the affine hull of X_*, whereas Y_*^+ is the affine hull of Y_*.

Consider now in this setting the special problem

$$\text{minimize } \tfrac{1}{2}\langle C_* x'', x''\rangle + \theta_{Y_*^-,B_*}(-A_* x'') \text{ over } x'' \in X_*^+, \tag{3.8}$$

and alongside of it (but not dual to it) the special problem

$$\text{maximize } \; -\tfrac{1}{2}\langle B_* y'', y''\rangle - \theta_{X_*^-,C_*}(A_*^T y'') \text{ over } y'' \in Y_*^+. \tag{3.9}$$

These problems are independent of the choice of (w'_*, x'_*, y'_*) in $\operatorname{gph} T$. Clearly, it they have only the solutions $x'' = 0$ and $y'' = 0$, then the same must be true for all instances of $\mathcal{P}_{**}$ and $\mathcal{D}_{**}$. We claim, however, that the converse holds as well, so that our solution analysis in fact can be reduced to just these two problems.

To establish the converse, it suffices to show that a choice of (w'_*, x'_*, y'_*) can actually be made for which $\mathcal{P}_{**}$ comes out as the problem in (3.8), and on the other hand a choice can be made for which $\mathcal{D}_{**}$ comes out as the problem in (3.9). For the case of $\mathcal{P}_{**}$, we take x'_* to belong to the relative interior of X_*, so that $T_{X_*}(x'_*) = X_*^+$. Take $y'_* = 0$. Then $T_{Y_*}(y'_*) = Y_*$. Let z be any vector in the relative interior of the cone polar to Y_*, so that $Y_* \cap z^\perp = Y_*^-$. The condition (2.1) of ample parameterization that we have assumed, in which the matrices $\nabla^2_{wx} L(w_*, x_*, y_*)$ and $\nabla^2_{wy} L(w_*, x_*, y_*)$ are the transposes of E_* and D_*, says that the linear transformation $w' \mapsto (E_* w', D_* w')$ maps $\mathbb{R}^d$ onto $\mathbb{R}^n \times \mathbb{R}^m$. This ensures the existence of a vector w'_* such that

$$E_* w'_* = -C_* x'_* + A_*^T y'_*, \qquad D_* w'_* = z + B_* y'_* + A_* x'_*,$$

or in other words by (3.6), $\nabla_{x'}L_*(w'_*, x'_*, y'_*) = 0$ and $\nabla_{y'}L_*(w'_*, x'_*, y'_*) = z$. We then have in (3.4) that $X_{**} = X_*^+ \cap 0^\perp = X_*^+$ while $Y_{**} = Y_* \cap z^\perp = Y_*^-$, as desired, so that $\mathcal{P}_{**}$ is (3.8). The case of $\mathcal{D}_{**}$ is analogous, and indeed we can appeal to symmetry.

The final stage of the proof has been reached: we demonstrate now that the first line of (3.2) is necessary and sufficient for problem (3.8) to have only $x'' = 0$ as a solution (the notations x'' and x' are here interchangeable), whereas the second line of (3.2) is necessary and sufficient for problem (3.9) to have only $y'' = 0$ as a solution.

Problem (3.8) is the primal problem associated with the Lagrangian triple $L_*(0,\cdot,\cdot)$, X_*^+, Y_*^-. Hence, x'' is optimal if and only if $x'' \in X_*^+$ and there exists $y'' \in Y_*^-$ such that

$$-\nabla_{x'}L_*(0, x'', y'') \in N_{X_*^+}(x''), \qquad \nabla_{y'}L_*(0, x'', y'') \in N_{Y_*^-}(x'').$$

Here $\nabla_{x'}L_*(0, x'', y'') = C_*x'' - A_*^T y''$ and $\nabla_{y'}L_*(0, x'', y'') = -B_*y'' - A_*x''$, whereas (because X_*^+ and Y_*^- are subspaces) we have $N_{X_*^+}(x'') = (X_*^+)^\perp$ and $N_{Y_*^-}(y'') = (Y_*^-)^\perp$. Thus, to say that (3.9) has only 0 as a solution is to say that the first implication in (3.2) is correct.

Likewise, problem (3.9) is the dual problem associated with the Lagrangian triple $L_*(0,\cdot,\cdot)$, X_*^-, Y_*^+. By a parallel argument characterizing its solutions by the corresponding saddle points, one sees that for (3.9) to have only 0 as a solution it is necessary and sufficient for the second implication in (3.2) to be correct. □

The subspaces in Theorem 3.2 are especially easy to describe in the case where X and Y are boxes, as we explain next.

Proposition 3.3 (box case). *Suppose that X and Y are boxes with nonempty interior: $X = \Pi_{j=1}^n X_j$ and $Y = \Pi_{i=1}^m Y_i$ for closed intervals X_j and Y_i that are not singletons (but need not be bounded). Let $(x_*, y_*) \in S(w_*)$ and define the index sets*

$$\begin{aligned} J_1 &= \big\{ j \in \{1,\ldots,n\} \,\big|\, (\partial L/\partial x_j)(w_*, x_*, y_*) = 0 \big\}, \\ J_2 &= \big\{ j \in \{1,\ldots,n\} \,\big|\, x_{*j} \in \operatorname{int} X_j \big\}, \\ I_1 &= \big\{ i \in \{1,\ldots,m\} \,\big|\, (\partial L/\partial y_i)(w_*, x_*, y_*) = 0 \big\}, \\ I_2 &= \big\{ i \in \{1,\ldots,m\} \,\big|\, y_{*i} \in \operatorname{int} Y_i \big\}. \end{aligned} \tag{3.10}$$

Then $J_1 \supset J_2$ and $I_1 \supset I_2$, and the subspaces in condition (3.2) of Theorem 3.2 have the following descriptions:

$$\begin{aligned} x' \in X_* - X_* \quad &\Longleftrightarrow \quad \text{for all } j \notin J_1, \text{ the } j\text{th coordinate of } x' \text{ is } 0, \\ y' \in Y_* - Y_* \quad &\Longleftrightarrow \quad \text{for all } i \notin I_1, \text{ the } i\text{th coordinate of } y' \text{ is } 0, \\ v' \in [X_* \cap -X_*]^\perp \quad &\Longleftrightarrow \quad \text{for all } j \in J_2, \text{ the } j\text{th coordinate of } v' \text{ is } 0, \\ u' \in [Y_* \cap -Y_*]^\perp \quad &\Longleftrightarrow \quad \text{for all } i \in I_2, \text{ the } i\text{th coordinate of } u' \text{ is } 0. \end{aligned} \tag{3.11}$$

Proof. By its definition in (2.3), the critical cone X_* consists of all $x' \in T_X(x_*)$ such that $x' \perp \nabla_x L(w_*, x_*, y_*)$. In this situation, where $(x_*, y_*) \in S(w_*)$, $\nabla_x L(w_*, x_*, y_*)$ is already known from Proposition 1.4 to belong to the normal cone $N_X(x_*)$, which is polar to $T_X(x_*)$. Therefore,

$$X_* = \Pi_{j=1}^n X_{*j}, \text{ where } X_{*j} = \begin{cases} T_{X_j}(x_{*j}) & \text{if } (\partial L/\partial x_j)(w_*, x_*, y_*) = 0, \\ \{0\} & \text{if } (\partial L/\partial x_j)(w_*, x_*, y_*) \neq 0, \end{cases} \tag{3.12}$$

where moreover

$$T_{X_j}(x_{*j}) = \begin{cases} (-\infty, \infty) & \text{if } x_{*j} \in \operatorname{int} X_j, \\ [0, \infty) & \text{if } X_j \text{ has } x_{*j} \text{ as left endpoint}, \\ (-\infty, 0] & \text{if } X_j \text{ has } x_{*j} \text{ as right endpoint}. \end{cases} \tag{3.13}$$

Note that necessarily $(\partial L/\partial x_j)(w_*, x_*, y_*) = 0$ when $x_{j*} \in \operatorname{int} X_j$, so $J_1 \supset J_2$, as claimed.

Because $X_* - X_*$ is the smallest subspace $\supset X_*$, whereas $X_* \cap -X_*$ is the largest subspace $\subset X_*$, the product form of X_* gives us

$$\begin{aligned} X_* - X_* &= \Pi_{j=1}^n [X_{j*} - X_{j*}], \\ X_* \cap -X_* &= \Pi_{j=1}^n [X_{j*} \cap -X_{j*}], \\ [X_* \cap -X_*]^\perp &= \Pi_{j=1}^n [X_{j*} \cap -X_{j*}]^\perp. \end{aligned} \tag{3.14}$$

According to (3.13) we have $T_{X_j}(x_{*j}) - T_{X_j}(x_{*j}) = (-\infty, \infty)$ in all cases, whereas we have $T_{X_j}(x_{*j}) \cap -T_{X_j}(x_{*j}) = (-\infty, \infty)$ when $x_{*j} \in \operatorname{int} X_j$, but $T_{X_j}(x_{*j}) \cap -T_{X_j}(x_{*j}) = \{0\}$ otherwise. Therefore through (3.12) we have

$$\begin{aligned} X_{j*} - X_{j*} &= \begin{cases} (-\infty, \infty) & \text{when } j \in J_1, \\ \{0\} & \text{when } j \notin J_1, \end{cases} \\ X_{j*} \cap -X_{j*} &= \begin{cases} (-\infty, \infty) & \text{when } j \in J_2, \\ \{0\} & \text{when } j \notin J_2, \end{cases} \\ [X_{j*} \cap -X_{j*}]^\perp &= \begin{cases} \{0\} & \text{when } j \in J_2, \\ (-\infty, \infty) & \text{when } j \notin J_2. \end{cases} \end{aligned}$$

In applying these expressions in (3.14), we obtain the descriptions claimed for $X_* - X_*$ and $[X_* \cap -X_*]^\perp$. The argument for $Y_* - Y_*$ and $[Y_* \cap -Y_*]^\perp$ is completely parallel. $\square$

Example 3.4 (extended convex programming). *In the context of Example 1.2 with the sets X and Y taken to be boxes, the definition of index sets J_1, J_2, I_1, I_2 in Proposition 3.3 is realized in terms of*

$$\begin{aligned} \frac{\partial L}{\partial x_j}(w_*, x_*, y_*) &= \frac{\partial f_0}{\partial x_j}(w_*, x_*) + \sum\nolimits_{i=1}^m y_{*i} \frac{\partial f_i}{\partial x_j}(w_*, x_*), \\ \frac{\partial L}{\partial y_i}(w_*, x_*, y_*) &= f_i(w_*, x_*) - \frac{\partial k}{\partial y_i}(w_*, y_*). \end{aligned}$$

The criterion in (3.2) of Theorem 3.2 comes out then as follows:

$$\left.\begin{aligned} x'_j = 0 \;\; \text{for } j \notin J_1 \\ x' \perp \nabla_x f_i(w_*, x_*) \;\; \text{for } i \in I_2 \\ \nabla^2_{xx} L(w_*, x_*, y_*) x' = 0 \end{aligned}\right\} \implies x' = 0,$$

$$\left.\begin{aligned} y'_i = 0 \;\; \text{for } i \notin I_1 \\ \sum\nolimits_{i \in I_1} y'_i \frac{\partial f_i}{\partial x_j}(w_*, x_*) = 0 \;\; \text{for } j \in J_2 \\ \nabla^2_{yy} k(w_*, y_*) y' = 0 \end{aligned}\right\} \implies y' = 0. \tag{3.15}$$

In other words, $\nabla^2_{xx} L(w_*, x_*, y_*)$ must be positive definite relative to the subspace

$$\Big\{ x' \in \mathbb{R}^n \,\Big|\, x'_j = 0 \;\; \text{for } j \notin J_1 \;\; \text{and } x' \perp \nabla_x f_i(w_*, x_*) \;\; \text{for } i \in I_2 \Big\}, \tag{3.16}$$

whereas $\nabla^2_{yy} k(w_*, y_*)$ must be positive definite relative to the subspace

$$\Big\{ y' \in \mathbb{R}^m \,\Big|\, y'_i = 0 \;\; \text{for } i \notin I_1 \;\; \text{and } \sum\nolimits_{i \in I_1} y'_i \frac{\partial f_i}{\partial x_j}(w_*, x_*) = 0 \;\; \text{for } j \in J_2 \Big\}. \tag{3.17}$$

Detail. In these circumstances we have

$$\begin{aligned} -A_* &= \left[\frac{\partial f_i}{\partial x_j}(w_*, x_*) \right]_{i=1,\, j=1}^{m,\, n} \\ B_* &= \nabla^2_{yy} k(w_*, y_*) \\ C_* &= \nabla^2_{xx} L(w_*, x_*, y_*) = \nabla^2_{xx} f_0(w_*, x_*) + \sum\nolimits_{i=1}^m y_{*i} \nabla^2_{xx} f_i(w_*, x_*), \end{aligned} \tag{3.19}$$

and the claims are then immediate from the subspace descriptions in Proposition 3.3. The matrix $\nabla^2_{xx} L(w_*, x_*, y_*)$ is known to be positive semidefinite, and therefore the condition $\nabla^2_{xx} L(w_*, x_*, y_*) x' = 0$ is equivalent to $\langle x', \nabla^2_{xx} L(w_*, x_*, y_*) x' \rangle = 0$. The first implication in (3.15) is thus identical to the positive definiteness of this matrix relative to the subspace in (3.16). Likewise, because the matrix $\nabla^2_{yy} k(w_*, y_*)$ is positive semidefinite, the condition $\nabla^2_{yy} k(w_*, y_*) y' = 0$ is equivalent to $\langle y', \nabla^2_{yy} k(w_*, y_*) y' \rangle = 0$, and the second implication in (3.15) means the positive definiteness of this matrix relative to the subspace in (3.17). □

Example 3.5 (ordinary convex programming)**.** *In the context of Example 1.1, where $X = \mathbb{R}^n$ and $Y = \mathbb{R}^s_+ \times \mathbb{R}^{m-s}$, the index sets J_1, J_2, I_1, I_2 of Proposition 3.3 reduce to*

$$\begin{aligned} &J_1 = J_2 = \{1, \ldots, n\}, \\ &I_1 = \{1, \ldots, m\} \setminus \big\{ i \in [1, s] \,\big|\, f_i(w_*, x_*) < 0 \big\}, \\ &I_2 = I_1 \setminus \big\{ i \in [1, s] \,\big|\, f_i(w_*, x_*) = 0,\; y_{*i} = 0 \big\}. \end{aligned}$$

The criterion in (3.2) of Theorem 3.2 comes out then as the linear independence of the gradients $\{\nabla_x f_i(w_*, x_*) \,|\, i \in I_1\}$ *plus the positive definiteness of* $\nabla^2_{xx} L(w_*, x_*, y_*)$ *relative to the subspace* $\{\nabla_x f_i(w_*, x_*) \,|\, i \in I_2\}^{\perp}$.

Detail. This specializes Example 3.4 to $k \equiv 0$ and the X and Y in question. □

The case of Theorem 3.2 obtained in Example 3.5 is the one that is known not only in convex programming but nonlinear programming more generally. See the discussion around Theorem 6 of [11].

4. Calmness in Perturbations

To conclude, we look at a more primitive form of Lipschitz-type behavior under perturbations and characterize it too by way of the proto-derivative descriptions in Theorem 2.3 and Corollary 2.4. There result will provide additional perspective on the nonsingularity condition in Theorem 3.2.

The saddle point mapping S is said to be *calm at* w_* *for isolated* (x_*, y_*) if there is a neighborhood O of (x_*, y_*) along with a constant κ such that

$$(x, y) \in S(w) \cap O \quad \Longrightarrow \quad |(x, y) - (x_*, y_*)| \leq \kappa |w - w_*| \quad \text{for } w \text{ near } w_* \ . \tag{4.1}$$

This condition entails (x_*, y_*) being an isolated point of $S(w_*)$. A broader definition of calmness covers situations where (x_*, y_*) need not be isolated, cf. [1], but we will not be concerned with that more general concept here.

Theorem 4.1 (calmness)**.** *Let* $(x_*, y_*) \in S(w_*)$ *and suppose that the ample parameterization condition (2.1) is satisfied at* (w_*, x_*, y_*). *Then* S *is calm at* w_* *for isolated* (x_*, y_*) *if and only if*

$$\begin{cases} x' \in X_*, \ C_* x' = 0, \ -A_* x' \in Y_*^* & \Longrightarrow \quad x' = 0, \\ y' \in Y_*, \ B_* y' = 0, \ \ A_*^T y' \in X_*^* & \Longrightarrow \quad y' = 0, \end{cases} \tag{4.2}$$

where X_*^* *and* Y_*^* *are the polyhedral convex cones polar to* X_* *and* Y_*.

Proof. On the basis of the calmness criterion furnished in [7; Thm. 4.1], the desired property of S holds if and only if $DS(w_* \,|\, x_*, y_*)(0) = \{0, 0)\}$. By invoking the description of $DS(w_* \,|\, x_*, y_*)$ in Corollary 2.4 we see that this corresponds to the problems $\mathcal{P}_*(w')$ and $\mathcal{D}_*(w')$ for $w' = 0$ having $x' = 0$ and $y' = 0$ as their only solutions. The objective in $\mathcal{P}_*(0)$ has the value 0 at $x' = 0$, and likewise the objective in $\mathcal{D}_*(0)$ has the value 0 at 0. Hence $\min \mathcal{P}_*(0) \leq 0 \leq \max \mathcal{D}_*(0)$. It follows that $\min \mathcal{P}_*(0) = 0 = \max \mathcal{D}_*(0)$ and therefore that the uniqueness of $x' = 0$ and $y' = 0$ as optimal solutions can be expressed as

$$\begin{aligned} &\tfrac{1}{2}\langle C_* x', x' \rangle + \theta_{Y_*, B_*}(-A_* x') > 0 \ \text{ for all nonzero } x' \in X_*, \\ &\tfrac{1}{2}\langle B_* y', y' \rangle \, + \, \theta_{X_*, C_*}(A_*^T y') \, > 0 \ \text{ for all nonzero } y' \in Y_*. \end{aligned} \tag{4.3}$$

Here we naturally have $\langle C_* x', x' \rangle \geq 0$ for all $x' \in X_*$, with strict inequality holding unless $C_* x' = 0$. On the other hand, the function θ_{Y_*, B_*} in Corollary 2.4 is by its definition the convex function conjugate to

$$\psi_{Y_*, B_*}(y') = \begin{cases} \frac{1}{2} \langle B_* y', y' \rangle & \text{if } y \in Y_*, \\ \infty & \text{if } y \notin Y_*. \end{cases}$$

Therefore, by the rule of convex analysis, its minimizing set is given

$$\operatorname{argmin} \theta_{Y_*, B_*} = \partial \psi_{Y_*, B_*}(0) = N_{Y_*}(0) = Y_*^*.$$

Since $0 \in Y_*^*$ we have $\min \theta_{Y_*, B_*} = \theta_{Y_*, B_*}(0) = 0$. It follows that $\theta_{Y_*, B_*}(-A_* x') \geq 0$ for all x', with strict inequality holding unless $-A_* x' \in Y_*^*$.

Thus, the first condition in (4.3) reduces to the first condition in (4.2). By virtually the same argument, with only changes of notation, the second condition in (4.3) reduces to the second condition in (4.2). $\square$

Comparison of the condition in (4.2) with the one in (3.2) reveals the essential difference in what it takes to get single-valued Lipschitzian behavior under perturbations instead of mere calmness. The calmness criterion (4.2) could of course likewise be elaborated along the lines of Proposition 3.3 and Examples 3.4 and 3.5. For recent studies of calmness in nonlinear programming, see Klatte [12] and Levy [13].

References

1. R. T. Rockafellar and R. J-B Wets, *Variational Analysis*, Springer Verlag, Berlin, 1998.
2. A. L. Dontchev and R. T. Rockafellar, "Ample parameterization of variational inclusions," preprint.
3. R. T. Rockafellar, "Extended nonlinear programming," in *Nonlinear Optimization and Applications 2* (G. Di Pillo and F. Giannessi, eds.), Kluwer, 1999.
4. R. T. Rockafellar, "Lagrange multipliers and optimality," SIAM Review 35 (1993), 183-238.
5. R. T. Rockafellar and R. J-B Wets, "A Lagrangian finite-generation technique for solving linear-quadratic problems in stochastic programming," Math Programming Studies 28 (1986), 63–93.
6. R. T. Rockafellar, "Linear-quadratic programming and optimal control," SIAM J. Control Opt. 25 (1987), 781–814.
7. R. T. Rockafellar, "Proto-differentiability of set-valued mappings and its applications in optimization," 449–482 in *Analyse Non Linéaire*, H. Attouch et al. (eds.), Gauthier-Villars, Paris, 1989.
8. A. B. Levy and R. T. Rockafellar, "Proto-derivatives and the geometry of solution mappings in nonlinear programming," 249-260 in *Nonlinear Optimization and Applications* (G. Di Pillo and F. Giannessi, eds.), Plenum, New York, 1996.
9. S. M. Robinson, "Strongly regular generalized equations," Math. of Operations Research 5 (1980), 43–62.
10. A. J. King and R. T. Rockafellar, "Sensitivity analysis for nonsmooth generalized equations," Mathematical Programming 55 (1992), 193–212.
11. A. L. Dontchev and R. T. Rockafellar, "Characterizations of strong regularity for variational inequalities over polyhedral sets," SIAM J. Optimization 6 (1996), 1087–1105.
12. D. Klatte, "Upper Lipschitz behavior of solutions to perturbed $\mathcal{C}^{1,1}$ programs," preprint.
13. A. B. Levy, "Calm minima in parameterized finite-dimensional optimization," preprint.