$\mathcal{VU}$-Smoothness and Proximal Point Results for Some Nonconvex Functions.

Robert Mifflin*

Department of Mathematics, Washington State University, Pullman, WA 99164-3113 (mifflin@math.wsu.edu)

Claudia Sagastizábal†

IMPA, Estrada Dona Castorina 110, Jardim Botânico, 22460-320 Rio de Janeiro-RJ, Brazil. On leave from INRIA, B.P. 105, 78153 Le Chesnay, France (sagastiz@impa.br).

Dedicated to Olvi Mangasarian who provided the first author with an excellent course in Nonlinear Programming (and 17 theorems of alternative).

This paper is concerned with a function f having primal-dual gradient structure at a point $\bar{x}$ which satisfies a property called strong transversality. The structure is related to $\mathcal{VU}$-space decomposition, depending on the subdifferential of f at $\bar{x}$. It is shown that there exists a C^2 primal track leading to $\bar{x}$ and a space decomposition mapping that is C^1. As a result, there exists a second order expansion of f on the primal track, an associated subdifferential that is C^1 in a certain sense, and a corresponding dual track. For $\bar{x}$ a minimizer, conditions on f are given to ensure that for any point near $\bar{x}$ its corresponding proximal point is on the primal track.

AMS Subject Classification. Primary: 90C31, 49J52; Secondary: 65K10, 47N10

KEY WORDS: proximal point, $\mathcal{VU}$-decomposition, second-order derivatives.

1 Introduction

Consider the problem of minimizing a convex function f on $\mathbb{R}^n$. Because there is no well-defined second-order object at those points where f is not differentiable, a straightforward application of a Newton-type method for minimization is not possible. However, nonsmoothness typically appears not in a general way, but in a structured manner. In many applications such structures result in the identification of manifolds where the function behaves "smoothly" and may even have some kind of Hessian.

The problem of defining smooth substructures for general nonsmooth functions has been addressed by many authors in various ways, and for different purposes. One such purpose is for stating second-order optimality conditions, such as in [7],

* The research of this author was supported by NSF grant DMS-0071459.
† The research of this author was supported by CNPq (Brazil) grant 303540/03-6.

[8], [22]. More recently, the concept of *partial smoothness* was introduced in [13] to develop calculus rules and sensitivity analysis for lower semicontinuous functions. Essentially, a partly smooth function has a smooth representative on an "activity" manifold, with a subdifferential that is continuous thereon. An equivalent concept, defined for convex functions in [26] for showing convergence of various nonlinear programming algorithms, is *identifiability*; see [13, Thm. 6.3]. Another closely related structure is the one called *primal-dual gradient* (`pdg`) structure in [18]. With this structure, it is possible to identify smooth tracks that are shown in [4, Ch. 2] to be equivalent to identifiability and partial smoothness. We refer to [5] for a thorough analysis of the relations among these and other structure notions, including one in [24]. To illustrate these concepts, consider the following two variable example from [12]:

$$f(u,v) := \max\left\{u^2 + v^2 - v, v\right\}.$$

This function is minimized at $\bar{x} = (0,0)^\top$, and the subgradient $\bar{g} = (0,0)^\top$ is in the relative interior of its subdifferential at $\bar{x}$, the line segment $\{0\} \times [-1,1]$. In fact, the function fails to be differentiable only along the "active" circle defined by $u^2 + v^2 - v = v$, i.e., by $v = 1 \pm \sqrt{1-u^2}$ and $|u| \leq 1$. In particular, when close to $\bar{x}$, points of the form $\chi := (u, 1 - \sqrt{1-u^2})^\top$ are on a semicircle shown in Fig. 1. When u is taken as a parameter, the trajectory $\chi(u)$ corresponds to what we call a *primal* track and defines the manifold on which f is partly smooth:

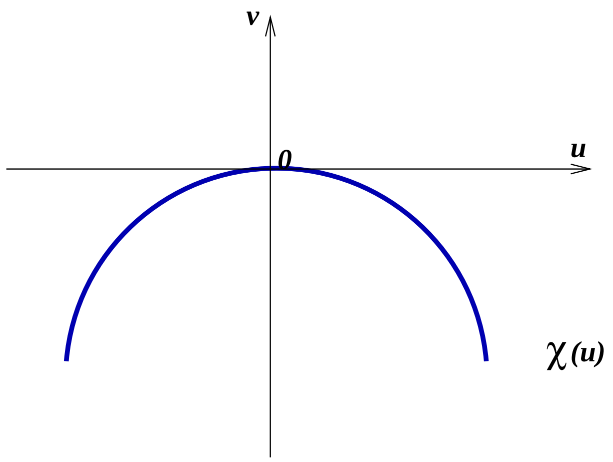

FIGURE 1: Smooth primal track

On this smooth primal track, the function of u given by $f(\chi(u)) = 1 - \sqrt{1-u^2}$ is C^2, graphs as a parabola and can be rapidly minimized with a Newton method. Each corresponding Newton step in $\mathbb{R}^2$ makes use of a "gradient" that is tangent to the track. Accordingly, there is a *dual* track $\gamma(u)$ corresponding to the primal track $\chi(u)$. As u converges to 0, the primal-dual track leads to $(\bar{x}, \bar{g})$, the minimizer and zero subgradient. In Fig. 2 below some members of the dual track are displayed as arrows having tails attached to the primal track:

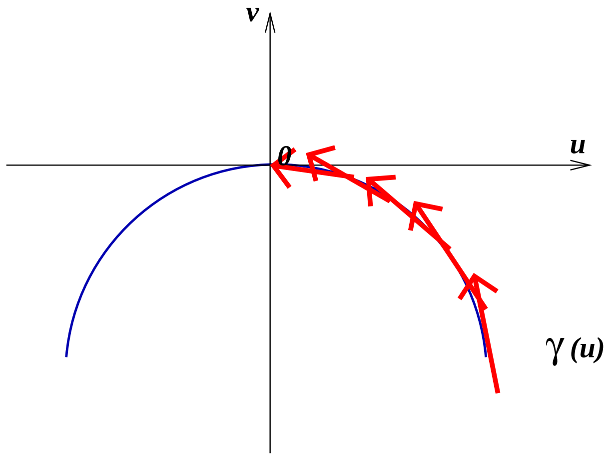

FIGURE 2: Dual track subgradient vectors

For minimization algorithms, it is important to study how to approximate smooth primal-dual tracks when the only available information is given by a black box whose ouput for each input x is $f(x)$ and one subgradient of f at x. A first important step in this direction is given in [19], where it is shown that, for a point near a minimizer, its corresponding proximal point lands on the primal track. This purely theoretical result is fundamental for minimization algorithms and their implementations, since it is known that a sequence of null steps from a (black box based) bundle mechanism can approximate proximal points with any desired accuracy, [1], [3], [9]. A corresponding dual track subgradient also can be approximated as a bundle by-product, by finding the mininum norm vector in the convex hull of the subgradients collected for estimating the proximal point.

The purpose of this paper is to extend to the nonconvex setting primal-dual track results from [20] and [19]. Here we focus on a conceptual definition of a primal-dual track for a lower semicontinuous function which is sound for future algorithmic development. Our approach is based on the $\mathcal{VU}$-theory introduced in [11] for convex functions. The starting point is to decompose $\mathbb{R}^n$ into two orthogonal subspaces, $\mathcal{V}$ and $\mathcal{U}$. Relative to the former, the graph of f is sharp or V-shaped, while for the latter the graph can appear U-shaped. For our simple example these subspaces are $\mathcal{V} = \{0\} \times \mathbb{R}$ and $\mathcal{U} = \mathbb{R} \times \{0\}$, on which the function f near $(0,0)^\top$ behaves, respectively, like an absolute value function and like a quadratic function.

Once the space decomposition is found, it is possible, as in our example, to identify a smooth primal track $\chi(u)$ leading to $\chi(0) = \bar{x}$ and tangent to the $\mathcal{U}$-subspace there whenever f has enough structure. In this case, for u small enough in the $\mathcal{U}$-subspace $f(\chi(u))$ has a second-order expansion about $u = 0$. A dual track depends on a local $\mathcal{V}(u)\mathcal{U}(u)$-decomposition along a primal track, as shown in Fig. 3.

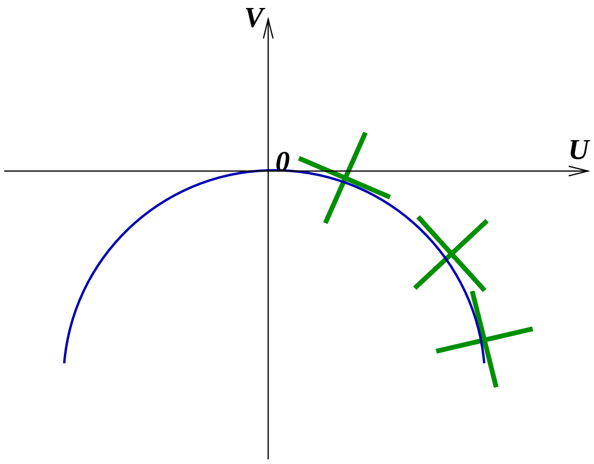

FIGURE 3: Local $\mathcal{V}(u)\mathcal{U}(u)$-decomposition

As in the convex case [18], our definition of a primal-dual track is built upon a suitably strong substructure for the function f. This is called a *strongly transversal* `pdg` structure; see Section 2 below. To show our main result relating primal tracks and proximal points, we employ three generalized convexity assumptions on f, namely, subdifferential- and prox-regularity and prox-boundedness; see [23, Defs. 8.3, 13.27, 1.23]. Prox-regularity and -boundedness are required in the nonconvex case in order to have well a defined Lipschitz continuous proximal point mapping. The remaining assumption is used to establish the link between primal tracks and proximal point mappings, via partial smoothness results from [6].

Our paper is organized as follows: we define the class of `pdg`-structured functions in Section 2 and relate such structure to $\mathcal{V}\mathcal{U}-$space decomposition. Smooth primal tracks and related primal results are studied in Section 3. New dual results, extending those in [20], are given in Section 4. The connection between proximal points and primal tracks is made in Section 5.

We use notation from [20] and the results therein. Much of the additional notation comes from [23]. For algebraic purposes we consider (sub)gradients to be column vectors. For (sub)gradients, the symbols ∂ and ∇ stand for differentiation with respect to $x \in \mathbb{R}^n$, while ∇_u indicates differentiation with respect to $u \in \mathbb{R}^{dim\,\mathcal{U}}$. Given a vector function $v(u)$ with components $v_i(u)$, the Jacobian $Jv(u)$ is a matrix whose i^{th} row is the transposed gradient $\nabla_u v_i(u)^\top$. For a given set Y, we denote its linear hull by $\mathrm{lin}Y$.

2 Notation. Function structure and space decomposition

For variational analysis we use the Clarke normal cone and subgradients, as defined in [23, eqs. 6(19) and 8(32)] depending on the basic normal cone [23, Def. 6.3] introduced in [21].

More precisely, for a set $C \subset \mathbb{R}^n$ and a point $x \in C$, a vector v is *normal* to C at x if there are sequences $x^\nu \to_C x$ and $v^\nu \to v$ such that $\langle v^\nu, z - x^\nu \rangle \leq o(|z - x^\nu|)$ for all $z \in C$. The set of normal vectors is denoted by $N_C(x)$. The closure of its convex hull yields the Clarke normal cone, denoted by $\bar{N}_C(x)$. A set C is said to be Clarke regular at x when C is locally closed at x and each normal vector v satisfies $\langle v, z - x \rangle \leq o(|z - x|)$ for all $z \in C$.

Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be a lower semicontinuous (lsc) function, so that its epigraph, denoted and defined by $\operatorname{epi} f := \{(x, \beta) \in \mathbb{R}^n \times \mathbb{R} : \beta \geq f(x)\}$, is a closed set in $\mathbb{R}^{n+1}$.

Take $x \in \mathbb{R}^n$ where f is finite-valued. The function f is said to be subdifferentially regular at x if $\operatorname{epi} f$ is a Clarke regular set at $(x, f(x))$. Consider the Clarke cone normal to the epigraph of f at $(x, f(x))$. The set of Clarke subgradients of f at x is denoted and defined by

$$\overline{\partial} f(x) := \left\{ g : (g, -1) \in \bar{N}_{\operatorname{epi} f}(x, f(x)) \right\} .$$

Being the closure of a point-to-set mapping, the normal cone $\bar{N}_{\operatorname{epi} f}(x, f(x))$ is outer semicontinuous (osc); see p. 154 in [23]. As a result, $\overline{\partial} f(x)$ is an osc point-to-set mapping at each x where f is continuous.

Furthermore, when f is Lipschitz around x, from [2, Theorem 2.5.1], $\overline{\partial} f(x)$ is the convex hull of all possible limits of gradients at points of differentiability of f in sequences converging to x. When f is subdifferentially regular at x, then $\overline{\partial} f(x)$ equals the Mordukhovich subdifferential ([21]) denoted by $\partial f(x)$ in [23]; see p. 337 and Definition 8.3 therein. If f is convex (and, hence, subdifferentially regular) this common subdifferential is the forerunner from convex analysis.

In order give a useful characterization of all of the subgradients in the Clarke subdifferential $\overline{\partial} f$, at least on a subset of $\mathbb{R}^n$, we endow f with a specific structure that we call *primal-dual gradient* structure.

2.1 `pdg` *structure*

In the following $\mathcal{P}$ is a subset of $\mathbb{R}^n$ containing $\bar{x}$ and m_1 and m_2 are nonnegative integers. We say that an lsc function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ has `pdg` *structure at* $\bar{x}$ *relative to* $\mathcal{P}$ if there exist $1 + m_1 + m_2$ real-valued *primal functions*

$$\{f_i(x)\}_{i=0}^{m_1} \quad \text{and} \quad \{\varphi_\ell(x)\}_{\ell=1}^{m_2}$$

that are C^2 on a ball in $\mathbb{R}^n$ about $\bar{x}$, denoted by $B(\bar{x})$, and a *dual multiplier set* $\Delta \subset \mathbb{R}^{1+m_1+m_2}$ satisfying the following conditions:

(i) $\bar{x} \in \left\{ x \in B(\bar{x}) : \left[\begin{array}{ll} f_i(x) = f(x) & \text{for } i = 0, 1, \ldots, m_1 \\ \varphi_\ell(x) = 0 & \text{for } \ell = 1, \ldots, m_2 \end{array} \right. \right\} \subseteq \mathcal{P} \subseteq B(\bar{x})$;

(ii) Δ is a closed convex set such that

(a) if $\alpha := (\alpha_0, \ldots, \alpha_{m_1}, \alpha_{1+m_1}, \ldots, \alpha_{m_1+m_2}) \in \Delta$, then $(\alpha_0, \ldots, \alpha_{m_1})$ is an element of the canonical simplex

$$\Delta_1 := \left\{ (\alpha_0, \alpha_1, \ldots, \alpha_{m_1}) : \sum_{i=0}^{m_1} \alpha_i = 1, \alpha_i \geq 0, i = 0, 1, \ldots, m_1 \right\};$$

(b) for each $i = 0, 1, \ldots, m_1$, $\mathbf{1}_{i+1} \in \Delta$, where $\mathbf{1}_j$ is the jth unit vector in $\mathbb{R}^{1+m_1+m_2}$;

and

(c) for each $\ell = 1, 2, \ldots, m_2$, there exists $\alpha^\ell \in \Delta$ such that $\alpha^\ell_{m_1+\ell} \neq 0$ and $\alpha^\ell_{m_1+i} = 0$ for $i \in \{1, 2, \ldots, m_2\} \backslash \{\ell\}$;

(iii) for each $x \in \mathcal{P}$,

(a) $f(x) \in \{f_i(x) : i = 0, 1, \ldots, m_1\}$;
(b) $g \in \overline{\partial} f(x)$ if and only if

$$g = \sum_{i=0}^{m_1} \alpha_i \nabla f_i(x) + \sum_{\ell=1}^{m_2} \alpha_{m_1+\ell} \nabla \varphi_\ell(x),$$

where the multipliers $\alpha_0, \alpha_1, \ldots, \alpha_{m_1+m_2}$ satisfy

complementary slackness: $\alpha_i = 0$ if $\left\{ \begin{array}{ll} f_i(x) \neq f(x) & \text{for } i \leq m_1 \\ \varphi_{i-m_1}(x) \neq 0 & \text{for } i > m_1 \end{array} \right.$

and

dual feasibility: $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_{m_1+m_2}) \in \Delta$. □

The pdg-structured class includes many nonsmooth functions f of the prototypical form

$$f(x) = \max\{F(x,t) : t \in T\}$$

where T is a closed set and F is sufficiently smooth on $\mathbb{R}^n \times T$. For example, the maximum eigenvalue of a symmetric matrix whose elements are C^2-functions of x is such a function. Our analysis in [18, Sec. 3.2] shows that such a function has pdg structure at any $\bar{x} \in \mathbb{R}^n$. At each $\bar{x}$, $1 + m_1$ is the multiplicity of the maximum eigenvalue and $m_2 = m_1(1 + m_1)/2$. The f_i's and φ_ℓ's depend on some C^2 vector functions that form a basis for a subspace spanned by certain eigenvectors; see [25]. The set Δ corresponds to a set of positive semidefinite matrices having unit traces and $\mathcal{P} = \{x \in B(\bar{x}) : \varphi_\ell(x) = 0 \text{ for } \ell = 1, \ldots, m_2\}$.

We next show the vital connection between `pdg` structure and $\mathcal{VU}$-space decomposition.

2.2 Relation to $\mathcal{VU}$-decomposition

Given an lsc function f, a point $\bar{x} \in \mathbb{R}^n$ where $f(\bar{x})$ is finite and $\overline{\partial} f(\bar{x})$ is nonempty, and an arbitrary subgradient $g \in \overline{\partial} f(\bar{x})$, the orthogonal subspaces

$$\mathcal{V} := \operatorname{lin}(\overline{\partial} f(\bar{x}) - g) \quad \text{and} \quad \mathcal{U} := \mathcal{V}^{\perp} \tag{1}$$

define the $\mathcal{VU}$-space decomposition, introduced in [11] and further studied in [16], [17], [18], [19], and [20].

Note that $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{V}$ and, from (1), the $\mathcal{U}$-component of a subgradient $g \in \overline{\partial} f(\bar{x})$ is the same as that of any other subgradient at $\bar{x}$; we call it the *$\mathcal{U}$-gradient* of f at $\bar{x}$ and denote it by $\bar{g}_{\mathcal{U}}$:

$$\bar{g}_{\mathcal{U}} := \bar{U}^{\top} g \text{ for any } g \in \overline{\partial} f(\bar{x}), \tag{2}$$

where $\bar{U}$ is an orthonormal basis matrix for $\mathcal{U}$. In addition, note that $\mathcal{V}$, depending on $\bar{x}$, is independent of the particular choice of $g \in \overline{\partial} f(\bar{x})$. Also, every $x \in \mathbb{R}^n$ can be decomposed into components $x_{\mathcal{U}}$ and $x_{\mathcal{V}}$ as follows:

$$\begin{aligned} \mathbb{R}^n \ni x &= \bar{U}(\bar{U}^{\top} x) + \bar{V}\Big([\bar{V}^{\top}\bar{V}]^{-1}\bar{V}^{\top} x\Big) \\ &= \bar{U} \quad x_{\mathcal{U}} \quad + \bar{V} \qquad x_{\mathcal{V}} \,. \end{aligned} \tag{3}$$

From (1), the relative interior of $\overline{\partial} f(\bar{x})$, denoted by $\operatorname{ri} \overline{\partial} f(\bar{x})$, is the interior of $\overline{\partial} f(\bar{x})$ relative to its affine hull, a manifold that is parallel to $\mathcal{V}$ (cf. [11, relation 2.1 and Prop. 2.2]). Accordingly,

$$\bar{g} \in \operatorname{ri} \overline{\partial} f(\bar{x}) \quad \Longrightarrow \quad \bar{g} + \Big(B(0,\eta) \cap \mathcal{V}\Big) \subset \overline{\partial} f(\bar{x}) \text{ for some } \eta > 0, \tag{4}$$

where $B(0,\eta)$ denotes a ball in $\mathbb{R}^n$ centered at 0, with radius η.

When f has `pdg` structure it is possible to completely characterize $\mathcal{V}$ in terms of the gradients of the primal functions ([20, Lemma 2.1]:

$$\mathcal{V} = \operatorname{lin}\Big(\{\nabla f_i(\bar{x}) - \nabla f_0(\bar{x})\}_{i=0}^{m_1} \cup \{\nabla \varphi_\ell(\bar{x})\}_{\ell=1}^{m_2}\Big). \tag{5}$$

We say that a function f having `pdg` structure at $\bar{x}$ relative to $\mathcal{P}$ satisfies *strong transversality* if the $n \times (m_1 + m_2)$ matrix

$$\bar{V} := \Big[\{\nabla f_i(\bar{x}) - \nabla f_0(\bar{x})\}_{i=1}^{m_1} \cup \{\nabla \varphi_\ell(\bar{x})\}_{\ell=1}^{m_2}\Big]$$

has full column rank.

In the strongly transversal case, $\bar{V}$ is a basis matrix for $\mathcal{V}$, with $\mathit{dim}\,\mathcal{V} = m_1 + m_2$ and $\bar{U}^{\top}\bar{V} = 0$. In the terminology of [20, p. 1181], $I := \{0, 1, \ldots, m_1 + m_2\}$, the set of all primal function indices, is a "basic" index set that is "transversal".

The following result shows how strong transversality relates interior subgradients to interior multipliers in Δ. This relation yields desirable properties for dual track subgradients; see Corollary 4.3 below.

Lemma 2.1. *Suppose $x \in \mathcal{P}$ satisfies*

$$\begin{cases} f_i(x) = f(x) & \textit{for all } i = 0, 1, \ldots, m_1 \\ \varphi_\ell(x) = 0 & \textit{for all } \ell = 1, \ldots, m_2 \end{cases}$$

and the $n \times (m_1 + m_2)$ matrix

$$V := \Big[\{\nabla f_i(x) - \nabla f_0(x)\}_{i=1}^{m_1} \cup \{\nabla \varphi_\ell(x)\}_{\ell=1}^{m_2} \Big]$$

has full column rank. Then

$$g := \sum_{i=0}^{m_1} \alpha_i \nabla f_i(x) + \sum_{\ell=1}^{m_2} \alpha_{m_1+\ell} \nabla \varphi_\ell(x) \tag{6}$$

is in the relative interior of $\overline{\partial} f(x)$ if and only if α is in the interior of Δ relative to the hyperplane

$$\mathcal{H} := \left\{ \tilde{\alpha} \in \mathbb{R}^{1+m_1+m_2} : \sum_{i=0}^{m_1} \tilde{\alpha}_i = 1 \right\}.$$

As a result, if $g \in \operatorname{ri} \overline{\partial} f(x)$, then $\alpha_i > 0$ for all $i = 0, 1, \ldots, m_1$.

Proof. The "only if" part follows from [18, Thm. 6.1 (i)] with $\bar{x}$ and the convex set $\partial f(\bar{x})$ there replaced, respectively, by x and the convex set $\overline{\partial} f(x)$, because in the terminology of [18, p.563] for the *basic index set I* the assumption on V is called *strong transversality* and the desired conclusion is called *strong dual feasibility.*
To see the "if" part, suppose $\alpha \in \mathbb{R}^{1+m_1+m_2}$ is in the interior of Δ relative to $\mathcal{H}$ and let $g \in \mathbb{R}^n$ be defined by (6). Then, since $\sum_{i=0}^{m_1} \alpha_i = 1$, there exists an open ball about $0 \in \mathbb{R}^{1+m_1+m_2}$, denoted by B_0, such that for all $\beta = (\beta_0, \ldots, \beta_{m_1+m_2})^\top \in B_0$ with $\sum_{i=0}^{m_1} \beta_i = 0$ the vector $\alpha^+ := (\alpha_0 + \beta_0, \alpha_1 + \beta_1, \ldots, \alpha_{m_1+m_2} + \beta_{m_1+m_2}) \in \Delta$. Let $g^+ := \sum_{i=0}^{m_1} \alpha_i^+ \nabla f_i(x) + \sum_{\ell=1}^{m_2} \alpha_{m_1+\ell}^+ \nabla \varphi_\ell(x)$. Then from (6), the fact that $\beta_0 = -\sum_{i=1}^{m_1} \beta_i$ and the definition of V

$$\begin{aligned} g^+ &= g + \sum_{i=0}^{m_1} \beta_i \nabla f_i(x) + \sum_{\ell=1}^{m_2} \beta_{m_1+\ell} \nabla \varphi_\ell(x) \\ &= g + \beta_0 \nabla f_0(x) + \sum_{i=1}^{m_1} \beta_i \nabla f_i(x) + \sum_{\ell=1}^{m_2} \beta_{m_1+\ell} \nabla \varphi_\ell(x) \\ &= g + \sum_{i=1}^{m_1} \beta_i \Big(\nabla f_i(x) - \nabla f_0(x) \Big) + \sum_{\ell=1}^{m_2} \beta_{m_1+\ell} \nabla \varphi_\ell(x) \\ &= g + V\beta. \end{aligned}$$

Since $\alpha^+ \in \Delta$, the assumptions on x and item (iii) of the `pdg` structure definition imply that $g^+ = g + V\beta \in \overline{\partial} f(x)$ for all $\beta \in B_0$. Since V has full rank, it is a basis matrix for the subspace parallel to the affine hull of $\overline{\partial} f(x)$. Since the relative interior of $\overline{\partial} f(x)$ is the interior of $\overline{\partial} f(x)$ relative to its affine hull, $g \in \operatorname{ri} \overline{\partial} f(x)$.
Finally, suppose for contradiction purposes that for $g \in \operatorname{ri} \overline{\partial} f(x)$ $\alpha_j = 0$ for some

$j \le m_1$. Then $m_1 \ge 1$. We may choose $\beta \in B_0$ such that $\sum_{i=0}^{m_1} \beta_i = 0$ and $\beta_j < 0$. Thus, $\alpha_j^+ = \alpha_j + \beta_j < 0$. This contradicts α^+ being in Δ. □

We next show how a strongly transversal pdg structure allows us to extend in a natural way objects such as $\mathcal{V}$ and $\bar{V}$ to certain mappings $\mathcal{V}(u)$ and $V(u)$ for $u \in \mathbb{R}^{dim\,\mathcal{U}}$ associated with a primal track leading to $\bar{x}$.

3 Primal results for strongly transversal pdg-structured functions

Throughout the remainder of the paper we assume that f has a strongly transversal pdg-structure at $\bar{x}$ relative to $\mathcal{P}$.

For our simple example, the primal track is given by the set of kinks, i.e., those points where the primal structure functions $f_0(u,v) := u^2 + v^2 - v$ and $f_1(u,v) := v$ agree; see Fig. 1. Likewise, for a general pdg-structured f, a second-order implicit function theorem ([10, p. 364]) gives a smooth primal track that is tangent to $\mathcal{U}$.

Theorem 3.1. *Let f be a function with* pdg *structure at $\bar{x}$ relative to $\mathcal{P}$ satisfying strong transversality and suppose that* $\dim \mathcal{V} \ge 1$ *and* $\dim \mathcal{U} \ge 1$. *For all u small enough,*

(i) *the nonlinear system, with variable v and parameter u,*

$$\begin{cases} f_i(\bar{x} + \bar{U}u + \bar{V}v) - f_0(\bar{x} + \bar{U}u + \bar{V}v) = 0, & i = 1, \dots, m_1, \\ \varphi_\ell(\bar{x} + \bar{U}u + \bar{V}v) = 0, & \ell = 1, \dots, m_2, \end{cases}$$

has a unique solution $v = v(u)$ satisfying $v(0) = 0$, where $v : \mathbb{R}^{dim\,\mathcal{U}} \to \mathbb{R}^{dim\,\mathcal{V}}$ is a C^2-function;

(ii) *the trajectory $\chi(u) := \bar{x} + \bar{U}u + \bar{V}v(u)$ has a C^1-Jacobian $J\chi(u) = \bar{U} - \bar{V}(V(u)^\top \bar{V})^{-1} V(u)^\top \bar{U}$, where*

$$V(u) := [\{\nabla f_i(\chi(u)) - \nabla f_0(\chi(u))\}_{i=1}^{m_1} \bigcup \{\nabla \varphi_\ell(\chi(u))\}_{\ell=1}^{m_2}].$$

(iii) *In particular, $v(0) = 0$, $\chi(0) = \bar{x}$, $V(0) = \bar{V}$, $Jv(0) = 0$, $J\chi(0) = \bar{U}$.*

(iv) *The trajectory $\chi(u)$ is tangent to $\mathcal{U}$ at $\chi(0) = \bar{x}$, with $v(u) = O(|u|^2)$;*

(v) *$\chi(u) \in \mathcal{P}$ and $f(\chi(u)) = f_i(\chi(u))$ for $i = 0, 1, \dots, m_1$.*

Proof. Items (i)-(iv) follow from Theorem 4.2 in [20], with its basic index set $K = I$. Item (v) follows from the definition of $\chi(u)$ and parts (i) and (iii)(a) of the pdg structure definition. □

For completeness, if *dim* $\mathcal{U}$(resp. *dim* $\mathcal{V}$)$= 0$, we define $\bar{V}$ (resp. $\bar{U}$) to be vacuous and $\chi(u) := \bar{x}$ (resp. $\chi(u) := \bar{x} + \bar{U}u$).

We call item (v) in Theorem 3.1 *primal feasibility.* In this case, it is possible to define the following $\mathcal{VU}$-like subspaces:

$$\mathcal{V}(u) := \text{lin}\left(\overline{\partial} f(\chi(u)) - g\right) \quad \text{and} \quad \mathcal{U}(u) := \mathcal{V}(u)^\perp,$$

where g is any subgradient in $\overline{\partial} f(\chi(u))$. These are important for expressing subgradients at $\chi(u)$ and defining the dual track $\gamma(u)$ (recall Figs. 3 and 2).

We now give basis matrices for the corresponding $\mathcal{V}(u)\mathcal{U}(u)$-space decomposition.

Proposition 3.2. *Let f be a function with* `pdg` *structure at $\bar{x}$ relative to $\mathcal{P}$ satisfying strong transversality, with corresponding mappings $v(u)$, $\chi(u)$, and $V(u)$ from Theorem 3.1. Then for all u small enough the $n \times \dim \mathcal{V}$ matrix $V(u)$ is a basis for $\mathcal{V}(u)$, while the $n \times \dim \mathcal{U}$ matrix $J\chi(u) = \bar{U} + \bar{V} Jv(u)$ is a basis for $\mathcal{U}(u)$.*

Proof. Since $V(u) \to \bar{V}$ as $u \to 0$, the matrix $V(u)$ has full rank for u sufficiently small. Since $\chi(u) \in \mathcal{P}$, the fact that $V(u)$ is a basis for $\mathcal{V}(u)$ follows essentially from (5), written with $\bar{x}$ and $\bar{V}$ replaced by $\chi(u)$ and $\mathcal{V}(u)$, respectively.
By Theorem 3.1, $J\chi(u)$ converges to $\bar{U}$ as $u \to 0$. It follows that $J\chi(u)$ has full rank for u near 0. Furthermore, the orthogonality result

$$V(u)^\top J\chi(u) = V(u)^\top \bar{U} - V(u)^\top \bar{V}(V(u)^\top \bar{V})^{-1} V(u)^\top \bar{U} = 0$$

implies that $J\chi(u)$ is a basis matrix for $\mathcal{U}(u)$. □

The previous result shows how `pdg` structure provides C^1 mappings for basis matrices corresponding to the local space decomposition. In the next section we will use `pdg` structure to obtain C^1 mappings for subgradients of f along the primal track $\chi(u)$.

4 Dual and primal-dual results for strongly transversal `pdg`-structured functions

We now consider dual objects associated with $\chi(u)$ which are C^1 functions of $u \in \mathbb{R}^{\dim \mathcal{U}}$. First, we introduce a multiplier vector $\alpha(u) \in \mathbb{R}^{1+m_1+m_2}$ (depending on the structure function gradients as well as on $\chi(u)$ and on an arbitrary parameter vector $y \in \mathbb{R}^n$), given by the unique solution of the linear system

$$\begin{cases} \bar{V}^\top \left(\sum_{i=0}^{m_1} \alpha_i \nabla f_i(\chi(u)) + \sum_{\ell=1}^{m_2} \alpha_{m_1+\ell} \nabla \varphi_\ell(\chi(u)) \right) = \bar{V}^\top y, \\ \sum_{i=0}^{m_1} \alpha_i = 1. \end{cases} \tag{7}$$

To see that (7) has a unique solution, consider the reduced system obtained by eliminating $\alpha_0(u)$ and using the unknown $\beta := (\alpha_i(u)\,, i \neq 0)$. Then β is the unique solution of the nonsingular system $\bar{V}^\top V(u)\beta = -\bar{V}^\top(\nabla f_0(\chi(u)) - y)$ and $\alpha_0(u) = 1 - \sum_i \beta_i$.

Although not indicated by the notation, multipliers $\alpha_i(u)$ depend on $\bar{V}^\top y$. Algebraic expressions for $\alpha(u)$ and its Jacobian can be found in [18, Thm. 5.2(i) and eq. (5.9)]. Multipliers are essential for obtaining second order expansions for f. In particular, in the strongly transversal case, they yield an expression for each subgradient along the primal track which is a combination of the primal function gradients. Since both the gradients and the multipliers are C^1, it is possible to obtain a second order expansion for f, simply from derivatives of these objects. The next result summarizes such relations for the strongly transversal case.

Theorem 4.1. *Let f be a function with* pdg *structure at $\bar{x}$ relative to $\mathcal{P}$ satisfying strong transversality. Let $y \in \mathbb{R}^n$ be a parameter vector, $z := \bar{V}^\top y$, and consider the primal-dual function*

$$L_I(u; z) := f(\chi(u)) - z^\top v(u).$$

Then for all u small enough we have the following:

(i) *$L_I(u; z)$ is a C^2-function of u with gradient $\nabla_u L_I(u; z) = \bar{U}^\top g_I(u; z)$, where $g_I(u; z)$ is defined by*

$$g_I(u; z) := \left(\sum_{i=0}^{m_1} \alpha_i(u) \nabla f_i(\chi(u)) + \sum_{\ell=1}^{m_2} \alpha_{m_1+\ell}(u) \nabla \varphi_\ell(\chi(u)) \right)$$

with $\alpha(u)$ solving (7) depending on $z = \bar{V}^\top y$ and with $\chi(u)$ from Theorem 3.1(ii).

(ii) *For all $g \in \overline{\partial} f(\chi(u))$*

$$g = g_I(u; z_g) \quad \text{with } \alpha(u) \in \Delta \text{ depending on } z_g := \bar{V}^\top g. \tag{8}$$

As a result, $J\chi(u)^\top g = \nabla L_I(u; 0)$.

(iii) *The Hessian of L_I is given by $\nabla^2 L_I(u; z) = J\chi(u)^\top M_I(u; z) J\chi(u)$, where $M_I(u; z)$ is the $n \times n$ matrix function defined by*

$$M_I(u; z) := \sum_{i=0}^{m_1} \alpha_i(u) \nabla^2 f_i(\chi(u)) + \sum_{\ell=1}^{m_2} \alpha_{m_1+\ell}(u) \nabla^2 \varphi_\ell(\chi(u)),$$

which depends on z via $\alpha(u)$ even when $u = 0$.

(iv) *In particular, $L_I(0; z) = f(\bar{x})$, $\nabla_u L_I(0; z) = \bar{g}_{\mathcal{U}}$, the $\mathcal{U}$-gradient of f at $\bar{x}$ defined in (2), and $\nabla_u^2 L_I(0; z) = \bar{U}^\top M_I(0; z) \bar{U}$.*

(v) *When $z = 0$ we have the second order expansion*

$$f(\chi(u)) = f(\bar{x}) + \bar{g}_{\mathcal{U}}^\top u + \frac{1}{2} u^\top \nabla_u^2 L_I(0; 0) u + o(|u|^2). \qquad \square$$

Proof. Items (i), (iii) and (iv) are a rewriting of [20, Thm. 5.1(ii)-(iv)] with its basic index set $K = I$. Likewise for item (v), which corresponds to [20, Thm. 5.2]. It only remains to show item (ii).

By primal feasibility, $\chi(u) \in \mathcal{P}$ for all u small enough. Fix u and take $g \in \overline{\partial} f(\chi(u))$. Then by conditions (iii)(b) and (ii)(a) in the pdg definition, there exists a vector $\alpha \in \Delta$ such that

$$g = \left(\sum_{i=0}^{m_1} \alpha_i \nabla f_i(\chi(u)) + \sum_{\ell=1}^{m_2} \alpha_{m_1+\ell} \nabla \varphi_\ell(\chi(u)) \right) \text{ with } \sum_{i=0}^{m_1} \alpha_i = 1.$$

Multiplying both sides of the left hand side relation on the left by $\bar{V}^\top$ gives linear system (7), with $y = g$. Since $\alpha(u)$ is the unique solution of (7), $\alpha(u) = \alpha \in \Delta$ and, thus, (8) holds.
Finally, since $J\chi(u)^\top = \bar{U}^\top + Jv(u)^\top \bar{V}^\top$,

$$J\chi(u)^\top g_I(u;\bar{V}^\top g) \quad = \quad \bar{U}^\top g_I(u;\bar{V}^\top g) + Jv(u)^\top \bar{V}^\top g_I(u;\bar{V}^\top g).$$

From (8) and item (i) written with $z = \bar{V}^\top g$, this implies that

$$J\chi(u)^\top g \quad = \quad \nabla_u L_I(u;\bar{V}^\top g) + Jv(u)^\top \bar{V}^\top g. \tag{9}$$

Since $L_I(u;0) = f(\chi(u))$, from the definition of L_I, we have that

$$L_I(u;z) = L_I(u;0) - z^\top v(u)$$

for all $z = \bar{V}^\top y$ and $y \in \mathbb{R}^n$. By differentiating this last relation with respect to u we obtain

$$\nabla_u L_I(u;z) = \nabla_u L_I(u;0) - Jv(u)^\top z.$$

This equality written with $z = \bar{V}^\top g$ and (9) give the desired result. $\square$

As a consequence of item (ii) in the above theorem, we see that a result as in (2) holds along the primal track, i.e., the $\mathcal{U}(u)$-component of the subdifferential is a singleton, namely $\nabla_u L_I(u;0)$.

Due to (8) in Theorem 4.1, we say that *dual feasibility* holds for all subgradients in $\overline{\partial} f(\chi(u))$. The following result is an important consequence of dual feasibility. In particular, item (iii) below gives explicit subgradient expressions that are useful for expressing dual track gradients in Corollary 4.3.

Theorem 4.2. *Let f be a function with* `pdg` *structure at $\bar{x}$ relative to $\mathcal{P}$ satisfying strong transversality. Suppose that $\bar{g} \in \overline{\partial} f(\bar{x})$ and $\bar{\alpha} \in \Delta$ is the multiplier vector corresponding to $\bar{g}$, i.e., such that $\bar{\alpha} = \alpha(0)$ depending on $z = \bar{V}^\top \bar{g}$ and satisfying*

$$\bar{g} = \sum_{i=0}^{m_1} \bar{\alpha}_i \nabla f_i(\bar{x}) + \sum_{\ell=1}^{m_2} \bar{\alpha}_{m_1+\ell} \nabla \varphi_\ell(\bar{x}).$$

For u small enough let the vector function $s(u)$, depending on $\bar{\alpha}$ and on $\chi(u)$ from Theorem 3.1(ii), be defined by

$$s(u) := \sum_{i=0}^{m_1} \bar{\alpha}_i \nabla f_i(\chi(u)) + \sum_{\ell=1}^{m_2} \bar{\alpha}_{m_1+\ell} \nabla \varphi_\ell(\chi(u)).$$

Then the following results hold for all u small enough:

(i) *$s(u)$ is a C^1 function satisfying $s(u) \in \overline{\partial} f(\chi(u))$, $J\chi(u)^\top s(u) = \nabla_u L_I(u;0)$, and $s(u) \to \bar{g}$ as $u \to 0$.*

(ii) *The point-to-set mapping $\overline{\partial} f(\chi(u))$ is continuous at $u = 0$.*

(iii) *If, in addition,* $\bar g \in \mathrm{ri}\,\overline{\partial} f(\bar x)$ *and* $w : \mathbb{R}^{dim\,\mathcal{U}} \to \mathbb{R}^{dim\,\mathcal{V}}$ *is any* C^1 *function such that* $w(0) = 0$*, then*

$$G(u) := J\chi(u)\,[J\chi(u)^\top J\chi(u)]^{-1}\,\nabla_u L_I(u;0) + V(u)\Big(\bar g_{\mathcal{V}} + w(u)\Big)$$

defines a C^1 *function, depending on* $V(u)$ *from Theorem 3.1(ii) and on* $\bar g_{\mathcal{V}} = [\bar V^\top \bar V]^{-1}\bar V^\top \bar g$*, which satisfies* $G(u) \in \mathrm{ri}\,\overline{\partial} f(\chi(u))$ *and* $G(u) \to \bar g$ *as* $u \to 0$*. In particular,* $s(u) \in \mathrm{ri}\,\overline{\partial} f(\chi(u))$*.*

Proof. To show item (i), first note that since the structure functions are C^2 and $\chi(u)$ is C^2 by item (ii) in Theorem 3.1, $s(u)$ is C^1. Since $\chi(u) \in \mathcal{P}$ by item (v) in Theorem 3.1 and $\bar\alpha \in \Delta$ by assumption, item (iii)(b) in the `pdg` definition implies that $s(u)$ belongs to $\overline{\partial} f(\chi(u))$. Hence, the second assertion follows from Theorem 4.1 (ii). Finally, since $\chi(u) \to \bar x$ as $u \to 0$, continuity of the primal function gradients implies that $s(u) \to \bar g$.
To show item (ii), first note that inner semicontinuity of $\overline{\partial} f(\chi(\cdot))$ at 0 follows from item (i). Outer semicontinuity for $\overline{\partial} f(x)$ holds at all x where f is continuous. The fact that $f(\chi(\cdot))$ is continuous follows from item (v) in Theorem 3.1, the continuity of $\chi(\cdot)$ and that of all the structure functions f_i.
Item (iii) is proved as follows: by construction, the function $G(u)$ is C^1 because from Theorem 3.1 $J\chi(u)$ and $V(u)$ are C^1 with $J\chi(0)$ equal to the full column rank matrix $\bar U$, $\nabla_u L_I(u;0)$ is C^1 by Theorem 4.1, and $w(u)$ is C^1 by assumption. In addition, since $J\chi(u) \to \bar U$, $\nabla_u L_I(u;0) \to \bar g_{\mathcal{U}}$, $V(u) \to \bar V$, and $w(u) \to 0$, we have, via decomposition (3), that $G(u) \to \bar U \bar g_{\mathcal{U}} + \bar V \bar g_{\mathcal{V}} = \bar g$. Together with the fact that $s(u) \to \bar g$, this means that

$$G(u) - s(u) \to 0 \quad \text{as } u \to 0. \tag{10}$$

Consider the following (nonsingular) linear system of order $m_1 + m_2$:

$$V(u)^\top V(u)\beta = V(u)^\top \Big(G(u) - s(u)\Big). \tag{11}$$

Let $\beta = (\beta_1(u), \ldots, \beta_{m_1+m_2}(u))^\top$ be its unique solution and define $\beta_0(u) := -\sum_{i=1}^{m_1} \beta_i(u)$. Then the multiplier vector function

$$\alpha(u) := \bar\alpha + \beta(u) = (\bar\alpha_0 + \beta_0(u), \ldots, \bar\alpha_{m_1+m_2} + \beta_{m_1+m_2}(u))$$

is an element of $\mathcal{H}$ which converges to $\bar\alpha$ as $u \to 0$, since (10) and (11) imply that $\beta(u) \to 0$ as $u \to 0$. The assumption that $\bar g \in \mathrm{ri}\,\overline{\partial} f(\bar x)$ and the "only if" part of Lemma 2.1 with $x = \bar x$, $V = \bar V$ and $\alpha = \bar\alpha$ implies that $\bar\alpha$ is in the interior of Δ relative to $\mathcal{H}$. Thus, for u small enough $\alpha(u)$ is also in this relative interior and the "if" part of Lemma 2.1 with $x = \chi(u)$, $V = V(u)$ and $\alpha = \alpha(u)$ shows that

$$g := \sum_{i=0}^{m_1} \alpha_i(u)\nabla f_i(\chi(u)) + \sum_{\ell=1}^{m_2} \alpha_{m_1+\ell}(u)\nabla\varphi_\ell(\chi(u)) \in \mathrm{ri}\,\overline{\partial} f(\chi(u)). \tag{12}$$

By the definitions of $s(u)$ and $\beta_0(u)$,

$$\begin{aligned} g &= s(u) + \sum_{i=0}^{m_1} \beta_i(u)\nabla f_i(\chi(u)) + \sum_{\ell=1}^{m_2} \beta_{m_1+\ell}(u)\nabla\varphi_\ell(\chi(u)) \\ &= s(u) + \beta_0(u)\nabla f_0(\chi(u)) + \sum_{i=1}^{m_1} \beta_i(u)\nabla f_i(\chi(u)) + \sum_{\ell=1}^{m_2} \beta_{m_1+\ell}(u)\nabla\varphi_\ell(\chi(u)) \\ &= s(u) + \sum_{i=1}^{m_1} \beta_i(u)\Big(\nabla f_i(\chi(u)) - \nabla f_0(\chi(u))\Big) + \sum_{\ell=1}^{m_2} \beta_{m_1+\ell}(u)\nabla\varphi_\ell(\chi(u)). \end{aligned}$$

From (11), the last relation means that

$$\begin{aligned} g &= s(u) + V(u)(\beta_1(u), \ldots, \beta_{m_1+m_2}(u))^\top \\ &= s(u) + V(u)\,[V(u)^\top V(u)]^{-1}\,V(u)^\top\Big(G(u) - s(u)\Big). \end{aligned}$$

From considering the $\mathcal{V}(u)\mathcal{U}(u)$-components of g we see that

$$V(u)^\top g = V(u)^\top\Big(s(u) + G(u) - s(u)\Big) = V(u)^\top G(u);$$

while, by Theorem 4.1(ii) and the definition of $G(u)$

$$J\chi(u)^\top g = \nabla_u L_I(u;0) = J\chi(u)^\top G(u).$$

We conclude that $g = G(u)$ so, from (12), $G(u) \in \mathrm{ri}\,\overline{\partial} f(\chi(u))$.
Finally, we will demonstrate a particular choice for $w(u)$ for which $s(u) = G(u) \in \mathrm{ri}\,\overline{\partial} f(\chi(u))$. In general, by item (i) and the definition of $G(u)$, $s(u)$ and $G(u)$ have the same $J\chi(u)$-component. So we need only choose $w(u)$ so that they have the same $V(u)$-component, i.e., such that $V(u)^\top s(u) = V(u)^\top V(u)(\bar{g}_{\mathcal{V}} + w(u))$. This equality is satisfied by taking $w(u) := [V(u)^\top V(u)]^{-1}\,V(u)^\top s(u) - \bar{g}_{\mathcal{V}}$, which is C^1, because both $s(u)$ and $V(u)$ are C^1, and satisfies $w(0) = 0$, because $s(0) = \bar{g}$ and $V(0) = \bar{V}$. $\square$

The next result reveals a particular subgradient at $\chi(u)$ which is closest to $\bar{g}$ and is also a C^1 function of u. When $\bar{g} = 0$, this subgradient is important for a minimization algorithm, since it defines a dual track leading to a zero subgradient.

Corollary 4.3. *Suppose the assumptions of Theorem 4.2 hold with $\bar{g} \in \mathrm{ri}\,\overline{\partial} f(\bar{x})$. Then for all u small enough*

$$\gamma(u) := J\chi(u)[J\chi(u)^\top J\chi(u)]^{-1}\nabla_u L_I(u;0) + V(u)\,[V(u)^\top V(u)]^{-1} V(u)^\top \bar{g}$$

defines a C^1 function satisfying $\gamma(u) \in \mathrm{ri}\,\overline{\partial} f(\chi(u))$, $\gamma(0) = \bar{g}$, and

$$\gamma(u) - \bar{g} = \operatorname{argmin}\left\{|g - \bar{g}|^2 : g \in \overline{\partial} f(\chi(u))\right\}. \tag{13}$$

Proof. Consider the following C^1 function of u:

$$w_\gamma(u) := \left\{[V(u)^\top V(u)]^{-1}V(u)^\top - \left[\bar V^\top \bar V\right]^{-1} \bar V^\top\right\} \bar g,$$

which satisfies $w_\gamma(0) = 0$ because, by Theorem 3.1(iii), $V(0) = \bar V$. By definition of $\bar g_{\mathcal{V}}$, $\bar g_{\mathcal{V}} + w_\gamma(u) = [V(u)^\top V(u)]^{-1}V(u)^\top \bar g$, so $\gamma(u)$ is the function $G(u)$ from Theorem 4.2 (iii), written with $w(u) = w_\gamma(u)$. Therefore, Theorem 4.2(iii) applies and $\gamma(u) \in \mathrm{ri}\, \overline{\partial} f(\chi(u))$. To show (13), recall from Proposition 3.2 that $J\chi(u)$ is a basis for $\mathcal{U}(u)$ and from Theorem 4.1(ii) that the $J\chi(u)$-component of any subgradient $g \in \overline{\partial} f(\chi(u))$ does not depend on g. This means that the minimization in (13) is only done on the subspace orthogonal to $\mathcal{U}(u)^\perp$, that is, on $\mathcal{V}(u)$. Since by Proposition 3.2, $V(u)$ is a basis for $\mathcal{V}(u)$, the solution to the minimization problem in (13) is given by the subgradient minimizing the norm of the $V(u)$-component of $g - \bar g$. The result follows because, by the definition of $\gamma(u)$, the $V(u)$-component of $\gamma(u) - \bar g$ is zero. □

In the next section we show that the properties of primal and dual feasibility, implied by strong transversality, are useful for relating the proximal point mapping to a primal track, via the concept of partial smoothness from [13].

5 Connection with proximal points

The smooth primal track $\chi(u)$ defines the smooth manifold

$$\mathcal{M} := \left\{p \in B(\bar x) : \left\{ \begin{array}{lcl} f_i(p) - f_0(p) & = & 0, \quad i = 1, \ldots, m_1 \\ \varphi_\ell(p) & = & 0, \quad \ell = 1, \ldots, m_2 \end{array} \right\}\right\}, \tag{14}$$

that is an example of an "activity" manifold in [13]. Indeed, since our assumption of strong transversality implies transversality and primal feasibility we have, from [20, Lemma 4.4], that $\mathcal{V} = N_{\mathcal{M}}(\bar x)$ and for all p close enough to $\bar x$

$$p \in \mathcal{M} \iff \begin{array}{lcl} p & = & \chi(u(p)) \\ & = & \bar x + \bar U u(p) + \bar V v(u(p)) \text{ where } u(p) := \bar U^\top (p - \bar x). \end{array} \tag{15}$$

In Lemma 5.1 and Theorem 5.3 below we assume that f is subdifferentially regular at each $p \in \mathcal{M}$ near $\bar x$, which implies that the Clarke and the Mordukhovich subdifferentials are the same. This is one of the defining properties of a function that is *partly smooth* at a point with respect to a manifold, [13, Def. 2.7]. Essentially, a partly smooth function has a smooth representative on a manifold with a subdifferential that is continuous thereon. For a strongly transversal `pdg`-structured function, these properties are illustrated, respectively, by item (v) in Theorem 4.1 and by item (ii) in Theorem 4.2. Moreover, [20, Thm. 7.4] shows that a strongly transversal `pdg`-structured function that is also subdifferentially regular on $\mathcal{M}$ is partly smooth at $\bar x$ relative to $\mathcal{M}$. In addition to the explicit expressions appearing in Theorems 4.1 and 4.2, the structure gives the C^1 mappings of Proposition 3.2, Theorem 4.2 and Corollary 4.3. These results are important for minimization algorithm research, especially the dual track $\gamma(u)$ from Corollary 4.3 written with $\bar{g} = 0$.

A concept equivalent to partial smoothness, which is defined for convex functions in [26], is one called *identifiability*, see [13, Thm. 6.3]. For the nonconvex case in [6], identifiability results follow from partial smoothness, provided the function is of the following type:
A lower semicontinuous function h is *prox-regular* at $\bar{x} \in dom\, h$ for a subgradient $\bar{g} \in \partial h(\bar{x})$ if there exists $\rho > 0$ such that

$$h(x') \geq h(x) + \langle g, x' - x\rangle - \frac{\rho}{2}|x' - x|^2$$

whenever x and x' are near $\bar{x}$ with $h(x)$ near $h(\bar{x})$ and $g \in \partial h(x)$ near $\bar{g}$. When this property holds for all subgradients in $\partial h(\bar{x})$, the function is prox-regular at $\bar{x}$.

Now we turn to the minimization context, where the zero vector is a subgradient at $\bar{x}$. The partial smoothness result [20, Thm. 7.4], combined with the identification result [6, Thm. 5.3] involving sequences, gives the following:

Lemma 5.1. *Let f be a function with* `pdg` *structure at $\bar{x}$ relative to $\mathcal{P}$ satisfying strong transversality. Also, suppose that f is subdifferentially regular at each $p \in \mathcal{M}$ near $\bar{x}$ where $\mathcal{M}$ is defined in* (14)*, that f is prox-regular at $\bar{x}$, and that* $0 \in \mathrm{ri}\, \partial f(\bar{x})$. *If $p_k \to \bar{x}$, with $f(p_k) \to f(\bar{x})$, then*

$$p_k \in \mathcal{M} \text{ for all large } k \text{ if and only if } dist\Big(0, \partial f(p_k)\Big) \to 0. \qquad \square$$

We also make use of the following result, extracted from [23, Prop. 13.37, p. 617], that establishes basic properties for the proximal point mapping of an lsc function.

Lemma 5.2. *Suppose that f is prox-regular at $\bar{x}$ for $\bar{g} = 0 \in \partial f(\bar{x})$ and that*

$$\exists r \in \mathbb{R} \text{ such that } f(\cdot) + \tfrac{1}{2}r|\cdot|^2 \text{ is bounded from below on } \mathbb{R}^n. \tag{16}$$

Then for all $\mu > 0$ sufficiently large there is a neighborhood of $\bar{x}$ on which the proximal point mapping

$$p_\mu(x) := \mathrm{argmin}_w \left\{ f(w) + \frac{1}{2}\mu|w - x|^2 \right\}$$

is well defined, single valued and Lipschitz continuous. In addition,

$$g_\mu(x) := \mu\,(x - p_\mu(x)) \in \partial f(p_\mu(x)) \quad \textit{and} \quad p_\mu(\bar{x}) = \bar{x}. \qquad \square$$

We are now in a position to establish our concluding result, the first equality of which was obtained independently in [4, Thm. 5.15]. It allows a minimization algorithm to approximate points on $\chi(u)$ without having to know this primal track.

Theorem 5.3. *In addition to the assumptions of Lemma 5.1 suppose that* (16) *holds. Then for all* $\mu > 0$ *sufficiently large and for any sequence* $x_k \to \bar{x}$

$$p_\mu(x_k) = \chi(u_k) \quad \textit{and} \quad \gamma_k = \operatorname{argmin}\left\{|g|^2 : g \in \overline{\partial} f(p_\mu(x_k))\right\}$$

for all large k*, where*

$$u_k := \bar{U}^\top(p_\mu(x_k) - \bar{x}) \quad \textit{and} \quad \gamma_k := J\chi(u_k)[J\chi(u_k)^\top J\chi(u_k)]^{-1}\nabla_u L_I(u_k; 0).$$

Proof. Take μ large enough for Lemma 5.2 to apply. If $x_k \to \bar{x}$ then, since $p_\mu(\cdot)$ is a Lipschitz continuous function, $p_\mu(x_k) \to p_\mu(\bar{x}) = \bar{x}$, $g_\mu(x_k) \to 0$ and, thus, the distances $dist\Big(0, \partial f(p_\mu(x_k))\Big)$ converge to 0. The definition of p_μ implies that $f(p_\mu(x_k)) \le f(\bar{x}) + \frac{1}{2}\mu(|\bar{x} - x_k|^2 - |p_\mu(x_k) - x_k|^2)$. Since both x_k and $p_\mu(x_k)$ converge to $\bar{x}$, this inequality implies that $\limsup_k f(p_\mu(x_k)) \le f(\bar{x})$. Then, since f is lsc at $\bar{x}$, $f(p_\mu(x_k)) \to f(\bar{x})$, so Lemma 5.1 with $p_k := p_\mu(x_k)$ applies to give $p_\mu(x_k) \in \mathcal{M}$ for all large k. This last relation is equivalent to the desired result, by (15) written with $p := p_\mu(x_k)$ and $u(p) = u_k$. The second equality follows from Corollary 4.3 with $\bar{g} = 0$ and $u = u_k$. □

There is no guarantee that $\bar{x}$ from Theorem 5.3 is a minimizer of f. However, if f is semiconvex [15] on $\mathbb{R}^n$ (the nondifferentiable analog of pseudoconvex [14]), then $\bar{x}$ does minimize f.

6 Concluding remarks

A conceptual superlinearly convergent $\mathcal{VU}$-algorithm makes a minimizing step in the $\mathcal{V}$-subspace, followed by a $\mathcal{U}$-Newton move. A $\mathcal{V}$-step essentially amounts to finding a point on a primal track $\chi(u)$. A principal result here is Theorem 5.3, that shows for prox-regular and -bounded functions $\mathcal{V}$-steps can be replaced by proximal steps, at least in the locality of a minimizer. In order to define implementable $\mathcal{VU}$ algorithms, we are currently considering how to use a bundle method to approximate proximal points and related quantities such as the special dual track subgradient $\gamma(u)$ of Corollary 4.3 corresponding to $\bar{g} = 0$. This corollary comes from new pdg-structure results appearing here in Lemma 2.1, Proposition 3.2, Theorem 4.1(ii), and Theorem 4.2. These results concerning the subdifferential mapping $\overline{\partial} f(\chi(u))$ depend on a strong transversality assumption and not on any subdifferential regularity assumption.

ACKNOWLEDGEMENTS

We thank the referees for clarification suggestions.

REFERENCES

[1] A. Auslender, *Numerical methods for nondifferentiable convex optimization*, Math. Programming Stud., 30 (1987), pp. 102–126.

[2] F. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, 1983. Reprinted by SIAM, 1990.

[3] R. Correa and C. Lemaréchal, *Convergence of some algorithms for convex minimization*, Math. Program., 62 (1993), pp. 261–275.

[4] W. Hare, *Nonsmoth Optimization with Smooth Substructure*, PhD thesis, Department of Mathematics, Simon Fraser University, 2003. Preprint available at http://www.cecm.sfu.ca/ whare.

[5] ———, *Recent functions and sets of smooth substructure: Relationships and examples*. 2003.

[6] W. Hare and A. Lewis, *Identifying active contraints via partial smoothness and prox-regularity*, Journal of Convex Analysis, (2003). Submitted. Preprint available at http://www.cecm.sfu.ca/ whare.

[7] J.-B. Hiriart-Urruty, *Refinements of necessary optimality conditions in nondifferentiable programming I*, Applied Mathematics and Optimization, 5 (1979), pp. 63–82.

[8] ———, *Refinements of necessary optimality conditions in nondifferentiable programming II*, Math. Programming Stud., 19 (1982), pp. 120–139.

[9] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, no. 305-306 in Grund. der math. Wiss, Springer-Verlag, 1993. (two volumes).

[10] S. Lang, *Real and Functional Analysis*, Springer-Verlag, New York, third ed., 1993.

[11] C. Lemaréchal, F. Oustry, and C. Sagastizábal, *The $\mathcal{U}$-Lagrangian of a convex function*, Trans. Amer. Math. Soc., 352 (2000), pp. 711–729.

[12] C. Lemaréchal and C. Sagastizábal, *Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries*, SIAM Journal on Optimization, 7 (1997), pp. 367–385.

[13] A. Lewis, *Active sets, nonsmoothness and sensitivity*, SIAM Journal on Optimization, 13 (2002), pp. 702–725.

[14] O. Mangasarian, *Nonlinear Programming*, McGraw - Hill, 1969.

[15] R. Mifflin, *Semi-smooth and semi-convex functions in constrained optimization*, SIAM Journal on Control and Optimization, 15 (1977), pp. 959–972.

[16] R. Mifflin and C. Sagastizábal, *$\mathcal{VU}$-decomposition derivatives for convex max-functions*, in Ill-posed Variational Problems and Regularization Techniques, R. Tichatschke and M. Théra, eds., no. 477 in Lecture Notes in Economics and Mathematical Systems, Springer-Verlag Berlin Heidelberg, 1999, pp. 167–186.

[17] ———, *Functions with primal-dual gradient structure and $\mathcal{U}$-Hessians*, in Nonlinear Optimization and Related Topics, G. D. Pillo and F. Giannessi, eds., no. 36 in Applied Optimization, Kluwer Academic Publishers B.V., 2000, pp. 219–233.

[18] ———, *On $\mathcal{VU}$-theory for functions with primal-dual gradient structure*, SIAM Journal on Optimization, 11 (2000), pp. 547–571.

[19] ———, *Proximal points are on the fast track*, Journal of Convex Analysis, 9 (2002), pp. 563–579.

[20] ———, *Primal-dual gradient structured functions: second-order results; links to epi-derivatives and partly smooth functions*, SIAM Journal on Optimization, 13 (2003), pp. 1174–1194.

[21] B. S. Mordukhovich, *Maximum principle in the problem of time optimal response with nonsmooth constraints*, Prikl. Mat. Meh., 40 (1976), pp. 1014–1023.

[22] R. Rockafellar, *Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives*, Mathematics of Operations Research, 14 (1989), pp. 462–484.

[23] R. Rockafellar and R.-B. Wets, *Variational Analysis*, no. 317 in Grund. der math. Wiss, Springer-Verlag, 1998.

[24] A. Shapiro, *On a class of nonsmooth composite functions*. Working paper, 2002.

[25] A. Shapiro and M. Fan, *On eigenvalue optimization*, SIAM Journal on Optimization, 5 (1995), pp. 552–569.

[26] S. Wright, *Identifiable surfaces in constrained optimization*, SIAM Journal on Control and Optimization, 21 (1993), pp. 1063–1079.