PRIMAL-DUAL GRADIENT STRUCTURED FUNCTIONS: SECOND-ORDER RESULTS; LINKS TO EPI-DERIVATIVES AND PARTLY SMOOTH FUNCTIONS

ROBERT MIFFLIN* AND CLAUDIA SAGASTIZÁBAL†

Abstract. We give second-order expansions for quite general nonsmooth functions from the $\mathcal{VU}$-space decomposition point of view. The results depend on primal-dual gradient structure which we relate to general concepts of second-order epi-derivatives and partly smooth functions. Expressions for the associated second-order objects are given in terms of $\mathcal{U}$−subspace Hessians.

Key words. second-order derivatives, $\mathcal{VU}$-decomposition, nonsmooth analysis, subdifferential.

AMS subject classifications. Primary: 90C31, 49J52; Secondary: 65K10, 58C20

1. Introduction and motivation. Nonsmooth analysis is essential for the understanding of optimization problems. In applications, nonsmoothness of functions typically appears, not in a general way, but in a structured manner. The problem of defining "good" structures for nonsmooth functions has been addressed by many authors in different ways. A good structure should be special enough to result in smooth behavior on manifolds, but also general enough to apply to a broad class of functions. Ideally, the structure should characterize the manifolds and associated Hessians.

In this paper, we extend primal-dual gradient (`pdg`) structure, first defined in [8], to a class of not necessarily convex functions. This subclass of lower semicontinuous functions encompasses many examples, such as extreme value functions that have an underlying C^2-structure. It is similar to the amenable class [15, p. 442] in the sense of having desirable smooth substructure and containing functions that are pointwise maxima of finite collections of smooth functions [6]. It is different from the amenable class and the fully amenable subclass [15, p. 443] due to containing functions that are not regular [15, p. 260] and containing regular ones that are not fully amenable, for example, maximum eigenvalue functions and other functions arising from infinite collections of smooth functions [7]. A related class that includes maximum eigenvalue functions, as in Example 3.2 below, is introduced in [16]. However, it should be noted that this new class does not include Examples 3.1 and 3.3 in Section 3.

Primal-dual gradient structure is tied closely to $\mathcal{VU}$−space decomposition where $\mathcal{V}$ is the subspace parallel to a function's Clarke subdifferential [2] at a point and $\mathcal{U}$ is orthogonal to $\mathcal{V}$. The structure provides a finite set of vectors that span $\mathcal{V}$, even though the subdifferential may have a continuum of extreme points. The $\mathcal{U}$−component of the subdifferential is a singleton and, hence, the function appears to be differentiable relative to $\mathcal{U}$ at the point of definition. This $\mathcal{VU}$−terminology is meant to describe the shape of a function's graph near the point in question. However, since in our (possibly) nonconvex setting the corresponding V and U shapes could be inverted, it is perhaps best to call them "sharp" and "smooth", respectively.

*Department of Pure and Applied Mathematics, Washington State University, Pullman, WA 99164-3113 (mifflin@math.wsu.edu). The research of this author was supported by National Science Foundation grant DMS-0071459.

†IMPA, Estrada Dona Castorina 110, Jardim Botânico, 22460-320 Rio de Janeiro-RJ, Brazil. (sagastiz@impa.br). On leave from INRIA, B.P. 105, 78153 Le Chesnay, France (Claudia.Sagastizabal@inria.fr). The research of this author was supported by FAPERJ (Brazil) grant E26/150.205/98.

For a `pdg`-structured function we develop second-order results depending on basic index sets. These correspond to "active manifolds" such as those defined by active constraints in nonlinear programming or those defined by constant eigenvalue multiplicity in eigenvalue optimization; see [5]. The structure provides us with one or more of these sets whose associated vectors span a subspace of $\mathcal{V}$ and generate an implicit function therein from which a smooth trajectory tangent to $\mathcal{U}$ can be defined. This information also generates multipliers that are smooth functions of $u \in \mathbb{R}^{dim\,\mathcal{U}}$ and depend on the $\mathcal{V}$ subspace component of a parameter vector $y \in \mathbb{R}^n$. Combining these elements and making a primal feasiblity assumption leads to the definition of a Lagrangian-like "primal-dual" function that is C^2 in u for each y. This function then provides expansions of second-order in u for its associated `pdg`-structured function. Because y need not be a subgradient at the point in question, the final expansion contained in Theorem 5.2 below is new even for a convex function. In the nonconvex case it is useful for obtaining the result that certain primal-dual $\mathcal{U}-$Hessians associated with a local minimizer are positive semidefinite.

Under a $\mathcal{V}-$minimality assumption for a certain subgradient, the corresponding primal-dual function becomes a $\mathcal{U}-$Lagrangian, along the lines of [4], even for the nonconvex case. In addition, we are able to give expressions for second-order epi-derivatives [14] in $\mathcal{U}-$space directions in terms of the related $\mathcal{U}-$Hessians. For an example from [5] we demonstrate below that the second-order behavior of the function can be captured well either by using epi-derivatives, defined via complicated epi-limits, or by ordinary second-derivatives from `pdg` structure.

We also connect regular members of the `pdg`-structured family to partly smooth functions [5]. An important feature of the partly smooth class is the existence of a sensitivity theory akin to that of nonlinear programming and of eigenvalue optimization. A partly smooth function is smooth on a manifold in the sense that on the manifold it equals a representative function that is C^2. Our `pdg` structure provides expressions for these objects and for the corresponding representative Hessian, in terms of a primal-dual Hessian associated with a basic index set that satisfies both primal and dual feasibility conditions.

The paper is organized as follows: We define the `pdg`-structured class in Section 2 and relate the structure to $\mathcal{VU}-$space decomposition. Section 3 provides three example functions. The last one, from [5], is used in subsequent sections to illustrate the principal results. Smooth trajectories, manifolds and associated multiplier functions are studied in Section 4. Section 5 gives second-order expansions for `pdg`-structured functions. The connection between $\mathcal{U}-$Hessians and second-order epi-derivatives is made in Section 6. Section 7 relates the previous results to functions that are partly smooth and gives expressions for manifold restricted Hessians.

We use notation from [8] and the results therein which depend on the subdifferential being a convex set rather than the function being convex. Much of the additional notation comes from [15]. For algebraic purposes we consider (sub)gradients to be column vectors. For a vector function $v(\cdot)$, its Jacobian $Jv(\cdot)$ is a matrix, each row of which is the transposed gradient of the corresponding component of $v(\cdot)$. For a given set Y, we denote its convex hull by con$Y$ and its linear hull by lin$Y$.

2. Function structure and space decomposition. For variational analysis we use the Clarke normal cone and subgradients, as defined in [15, Def. 6.3 and eq. 8(32)].

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 $\langle v^\nu, z - x^\nu \rangle \le o(|z - x^\nu|)$ 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 $\bar{N}_C(x)$.

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, and 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

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

When f is Lipschitz around x, from [2, Theorem 2.5.1], $\bar{\partial} f(x)$ 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 $\bar{\partial} f(x)$ equals the Mordukhovich subdifferential ([11]) denoted by $\partial f(x)$ in [15]; see p. 337 and Definition 8.3 therein. If f is convex (and, hence, regular) this common subdifferential is the forerunner from convex analysis.

In order to give second-order expansions for f, it is essential to properly describe all of the subgradients in the Clarke subdifferential $\bar{\partial} f$, at least on a subset of $\mathbb{R}^n$. This is the purpose of the structure depending on a set $\mathcal{P}$ introduced next. In particular, the presence of the functions φ_ℓ and corresponding multipliers allows the subdifferential to be nonpolyhedral and, hence, permits the class to contain functions that are not fully amenable [15, pp. 443-444] such as the two convex example functions considered throughout [8]. In addition, the way in which these functions enter the structure allows for the possibility of $\bar{\partial} f$ being unbounded at some points.

2.1. `pdg` structure. In the following $\mathcal{P}$ is a subset of $\mathbb{R}^n$ containing $\bar{x}$. 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 exists $m_1 + 1 + m_2$ *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 about $\bar{x}$, denoted by $B(\bar{x})$, and a *dual multiplier set* $\Delta \subset \mathbb{R}^{m_1+1+m_2}$ 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 $\alpha := (\alpha_0, \ldots, \alpha_{m_1}, \alpha_{m_1+1}, \ldots, \alpha_{m_1+m_2}) \in \Delta$, 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 $\mathbb{R}^{m_1+1+m_2}$;

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) $g \in \bar{\partial} f(x)$ if and only if

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

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

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

and

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

As a consequence of item (iii), for all $x \in \mathcal{P}$ f is finite-valued at x and $\bar{\partial} f(x)$ is a nonempty set depending on a multiplier set Δ that is independent of x and is the part of the structure which reflects the shape of the subdifferential.

This pdg definition is broader than its convex forerunner in [8], because the set $\mathcal{P}$ considered here can be larger than the one defined there; see the first two examples in Section 3 below.

It is not difficult to show, as in [8], that this definition covers a function that is the pointwise maximum (or minimum) value of a finite collection of C^2-functions by taking $m_1 + 1$ to be the number of "active" functions at $\bar{x}$, m_2 to be 0, Δ to be Δ_1 and $\mathcal{P}$ to be a ball about $\bar{x}$ chosen small enough to exclude the "inactive" functions from the local structure. Such a function can be nonconvex and for the minimum value case can be nonregular [15, p. 260] at points of nondifferentiability and, as a consequence, not amenable [15, pp. 442-444].

Before giving additional examples, we show the connection between pdg structure and $\mathcal{VU}$-space decomposition.

2.2. Relation to $\mathcal{VU}$-decomposition. The $\mathcal{VU}$-theory introduced in [4] and further studied in [6], [8], [7], [10], is based on decomposing $\mathbb{R}^n$ into two orthogonal subspaces $\mathcal{V}$ and $\mathcal{U}$ depending on a point such that $\mathcal{V}$ being nontrivial indicates non-smooth behavior of the function at the point. As a result of (2.2) below, from the $\mathcal{U}$-subspace point of view, the function appears to be differentiable.

More precisely, given an lsc function f and a point $\bar{x} \in \mathbb{R}^n$ such that $f(\bar{x})$ finite and $\bar{\partial} f(\bar{x})$ is nonempty, let $g \in \bar{\partial} f(\bar{x})$ be an arbitrary subgradient and define the orthogonal subspaces

$$\mathcal{V} := \operatorname{lin}(\bar{\partial} f(\bar{x}) - g) \quad \text{and} \quad \mathcal{U} := \mathcal{V}^{\perp}. \tag{2.1}$$

Note that $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{V}$. Thus, letting $\bar{U}$ be an orthonormal basis matrix for $\mathcal{U}$, the $\mathcal{U}$-component of $x \in \mathbb{R}^n$ is given by $x_{\mathcal{U}} := \bar{U}^\top x$. In particular, from (2.1), the $\mathcal{U}$-component of a subgradient $g \in \bar{\partial} f(\bar{x})$ 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 $\bar{g}_{\mathcal{U}}$:

$$\bar{g}_{\mathcal{U}} := \bar{U}^\top g \text{ for any } g \in \bar{\partial} f(\bar{x}). \tag{2.2}$$

In addition, note that $\mathcal{V}$, depending on $\bar{x}$, is independent of the particular choice of $g \in \bar{\partial} f(\bar{x})$. Furthermore, when f has pdg structure it is possible to completely characterize $\mathcal{V}$ in terms of the gradients of the primal functions. More precisely, relations (4.3) and (4.5) and Lemma 4.1 in [8], imply the following:

LEMMA 2.1. *If f has* pdg *structure at $\bar{x}$ relative to $\mathcal{P}$ then*

$$\nabla f_i(\bar{x}) \in \bar{\partial} f(\bar{x}) \textit{ for } i = 0, 1, \ldots, m_1, \nabla \varphi_\ell(\bar{x}) \in \mathcal{V} \textit{ for } \ell = 1, \ldots, m_2 \tag{2.3}$$

and the subspace $\mathcal{V}$ *from* (2.1) *can be written as*

$$\mathcal{V} = \mathrm{lin}\left(\{\nabla f_i(\bar{x}) - \nabla f_0(\bar{x})\}_{i=0}^{m_1} \cup \{\nabla \varphi_\ell(\bar{x})\}_{\ell=1}^{m_2}\right).$$

3. Specific functions with pdg structure. In general, pdg structure is not unique and it is desirable to have $\mathcal{P}$ as large as possible in order to exhibit interesting structure of a function. The following examples illustrate three different ways to define $\mathcal{P}$. The initial example shows that it is possible for $\mathcal{P}$ to be larger than $\{x \in B(\bar{x}) : \varphi_\ell(x) = 0 \text{ for } \ell = 1, \ldots, m_2\}$, whereas the second example has $\mathcal{P}$ equal to this set. In addition, the first example provides a function for which $\bar{\partial} f(\bar{x})$ is unbounded.

Example 3.1. For $x \in \mathbb{R}$ let $f(x) := \begin{cases} -x & \text{if } |x| \le 1 \\ -x + \sqrt{x^2 - 1} & \text{if } |x| \ge 1 \end{cases}$ and let $\bar{x} := 1$. This function from [1] is the *spectral abscissa* (the largest real part of all eigenvalues) of the matrix $\begin{pmatrix} 0 & 1 \\ -1 & -2x \end{pmatrix}$ and $\bar{x} = 1$ is a minimizing point at which f is nondifferentiable.

Take $B(\bar{x}) := (0, 2)$, $m_1 := 0$, $f_0(x) := -x$, $m_2 := 1$, $\varphi_1(x) := x^2 - 1$, $\mathcal{P} := \{x \in B(\bar{x}) : \varphi_1(x) \le 0\} = (0, 1]$ and $\Delta := \{(\alpha_0, \alpha_1) : \alpha_0 = 1, \alpha_1 \ge 0\}$. Then $f_0(1) = -1 = f(1)$, $\varphi_1(1) = 0$ and

$$\begin{aligned} \bar{\partial} f(1) &= \{\gamma : -1 \le \gamma\} \\ &= \{\alpha_0(-1) + \alpha_1(2) : \alpha_0 = 1, \alpha_1 \ge 0\} \\ &= \{\alpha_0 f_0'(1) + \alpha_1 \varphi_1'(1) : (\alpha_0, \alpha_1) \in \Delta\}, \end{aligned}$$

$\mathcal{V} = \mathbb{R}$ and $\mathcal{U} = \{0\}$. Finally, for $x \in \mathcal{P} \backslash \bar{x} = (0, 1)$, $f_0(x) = -x = f(x)$, $\varphi_1(x) = x^2 - 1 \neq 0$ and

$$\bar{\partial} f(x) = \{-1\} = \{\alpha_0 f_0'(x) + \alpha_1 \varphi_1'(x) : (\alpha_0, \alpha_1) \in \Delta, \alpha_1 = 0\},$$

so f has the desired pdg-structure.

Example 3.2. For $x \in \mathbb{R}^n$ let $f(x)$ be the maximum eigenvalue of a symmetric matrix whose elements are C^2-functions of x. Then from the analysis in [8, Section 3.2] f has pdg structure at any $\bar{x} \in \mathbb{R}^n$, because the subdifferential used there, from [12], is the Clarke subdifferential for this locally Lipschitz function. At each $\bar{x}$, $m_1 + 1$ is the multiplicity of the maximum eigenvalue and $m_2 = m_1(m_1 + 1)/2$. The f_is and φ_ℓs depend on some C^2 vector functions that form a basis for a subspace spanned by certain eigenvectors; see [17]. The set Δ corresponds to a set of positive semidefinite matrices having unit traces and $\mathcal{P} = \{x \in B(\bar{x}) : \varphi_\ell(x) = 0 \text{ for } \ell = 1, \ldots, m_2\}$.

Finally, we give a pdg structure for a locally Lipschitz function $f : \mathbb{R}^2 \to \mathbb{R}$ from [5, Section 7].

Example 3.3. For $x = (x_1, x_2)^\top$ consider the following function, defined on a partition of $\mathbb{R}^2$:

$$f(x) := \begin{cases} f_0(x) & \text{on} \quad S_0 := \{(x_1, x_2) \in \mathbb{R}^2 : x_2 \le 0\} \\ h_1(x) & \text{on} \quad S_1 := \{(x_1, x_2) \in \mathbb{R}^2 : 0 < x_2 < 2x_1^2\} \\ f_2(x) & \text{on} \quad S_2 := \{(x_1, x_2) \in \mathbb{R}^2 : 0 < 2x_1^2 \le x_2 \le 4x_1^2\} \\ f_3(x) & \text{on} \quad S_3 := \{(x_1, x_2) \in \mathbb{R}^2 : 4x_1^2 < x_2\} \end{cases}$$

where

$$\begin{aligned}f_0(x) &:= x_1^2 - x_2,\\ h_1(x) &:= \sqrt{x_1^4 + 2x_1^2 x_2 - x_2^2},\\ f_2(x) &:= 3x_1^2 - x_2,\\ f_3(x) &:= -5x_1^2 + x_2\end{aligned}$$

and the sets S_i defined here correspond to the sets S_{i+1} in [5].

Take $\bar{x} = (0,0)^\top$, and note that $f(\bar{x}) = 0$. As shown in [5],

$$\bar{\partial} f(\bar{x}) = \operatorname{con}\{(0,-1)^\top, (0,1)^\top\}.$$

Thus,

$$\mathcal{V} = \operatorname{lin}\{(0,1)^\top\} \quad \text{and} \quad \mathcal{U} = \operatorname{lin}\{(1,0)^\top\}$$

and letting $\bar{U} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ gives in (2.2) $\bar{g}_{\mathcal{U}} = 0$. Furthermore, since $(0,0)^\top \in \bar{\partial} f(\bar{x})$, $\bar{x}$ is a stationary point for f.

The difficulty in determining a `pdg` structure for f at $\bar{x}$ comes from the function h_1 that is not C^1 on any ball about $\bar{x}$ and yet defines f on S_1. Accordingly, we omit S_1 by taking $\mathcal{P} := S_0 \cup S_2 \cup S_3$ and consider f's behavior on the regions where the closure of S_1 intersects S_0 or S_2, i.e., on

$$\mathcal{M}_1 := \{(x_1, x_2) : x_2 = 0\} \tag{3.1}$$

and on $\mathcal{R} := \{(x_1, x_2) : x_2 = 2x_1^2, x_1 \neq 0\}$. The gradient of h_1 is given by

$$\nabla h_1(x) = \left(1 + 2\frac{x_2}{x_1^2} - \left(\frac{x_2}{x_1^2}\right)^2\right)^{-1/2} \left(2x_1\left(1 + \frac{x_2}{x_1^2}\right), 1 - \frac{x_2}{x_1^2}\right)^\top.$$

Since on $\mathcal{R}$, $\nabla h_1(x) = (6x_1, -1)^\top = \nabla f_2(x)$, f has a continuous gradient and there is no loss with respect to $\mathcal{R}$ from omitting h_1. Thus, we replace h_1 by a C^2-function f_1 whose value and gradient agree with those of h_1 on $\mathcal{M}_1 \backslash \bar{x}$. Since

$$h_1(x_1, 0) = x_1^2 \quad \text{and} \quad \nabla h_1(x_1, 0) = (2x_1, 1)^\top$$

the desired function f_1 is given by

$$f_1(x) := x_1^2 + x_2.$$

In order to show that f has a `pdg` structure at $\bar{x}$ relative to $\mathcal{P}$, we let the ball $B(\bar{x})$ be $\mathbb{R}^2$ and consider the primal functions f_0, f_1, f_2, f_3 as defined above (so $m_1 = 3$ and $m_2 = 0$). As for the dual multiplier set Δ, it is the canonical simplex Δ_1 in $\mathbb{R}^4$. We now check satisfaction of the three items defining `pdg` structure in Section 2.1. Items (i), (ii) and (iii)(a) follow immediately from the definition of f and those of the structure objects f_i, $\mathcal{P}$ and Δ.

We conclude by demonstrating satisfaction of (iii)(b). The gradients of the primal functions are given by

$$\nabla f_0(x) = \begin{pmatrix} 2x_1 \\ -1 \end{pmatrix}, \nabla f_1(x) = \begin{pmatrix} 2x_1 \\ 1 \end{pmatrix}, \nabla f_2(x) = \begin{pmatrix} 6x_1 \\ -1 \end{pmatrix}, \nabla f_3(x) = \begin{pmatrix} -10x_1 \\ 1 \end{pmatrix}. \tag{3.2}$$

First consider the case $x \in S_0$, with $x \neq \bar{x}$ (for $x = \bar{x}$, this item is easily verified). Because $\nabla h_1(x_1, 0) = (2x_1, 1)^\top$ we have that

$$\bar{\partial} f(x) = \begin{cases} \operatorname{con}\left((2x_1, 1)^\top, (2x_1, -1)^\top\right) & \text{for } x_2 = 0 \\ \{(2x_1, -1)^\top\} & \text{for } x_2 < 0. \end{cases} \tag{3.3}$$

To write the subdifferential with **pdg** structure, the multipliers α_i must satisfy complementary slackness conditions, as given in the following table:

$x \in S_0 \backslash \bar{x}$	$f_i = f$	$f_i \neq f$	$\alpha_i = 0$
$x_2 = 0, x_1 \neq 0$	f_0, f_1	f_2, f_3	α_2, α_3
$x_2 < 0, x_1 \neq 0$	f_0	f_1, f_2, f_3	$\alpha_1, \alpha_2, \alpha_3$
$x_2 < 0, x_1 = 0$	f_0, f_2	f_1, f_3	α_1, α_3.

As a result, the respective convex combinations of gradients are given by

$x \in S_0 \backslash \bar{x}$	$\sum_i \alpha_i \nabla f_i$
$x_2 = 0, x_1 \neq 0$	$\alpha_0 \nabla f_0(x) + (1 - \alpha_0) \nabla f_1(x)$
$x_2 < 0, x_1 \neq 0$	$\nabla f_0(x)$
$x_2 < 0, x_1 = 0$	$\alpha_0 \nabla f_0(x) + (1 - \alpha_0) \nabla f_2(x) = \nabla f_0(x)$

where the last equality follows from the fact that f_0 and f_2 have the same gradient if $x_1 = 0$. Comparing these results to (3.3) via (3.2) shows that item (iii)(b) holds for all $x \in S_0$.

Similar calculations can be carried out for S_2 and S_3, whose closures intersect at

$$\mathcal{M}_2 := \{(x_1, x_2) : x_2 = 4x_1^2\}. \tag{3.4}$$

Table 3.1 below summarizes the main results from such calculations.

$x (\neq \bar{x})$	$\bar{\partial} f(x)$	$f_i = f$	$f_i \neq f$	$\alpha_i = 0$
$x \in S_2 \backslash \mathcal{M}_2$	$\nabla f_2(x)$	f_2	f_0, f_1, f_3	$\alpha_0, \alpha_1, \alpha_3$
$x \in \mathcal{M}_2$	$\operatorname{con}(\nabla f_2(x), \nabla f_3(x))$	f_2, f_3	f_1, f_2	α_1, α_2
$x \in S_3$	$\nabla f_3(x)$	f_3	f_0, f_1, f_2	$\alpha_0, \alpha_1, \alpha_2$

TABLE 3.1
Results for S_2 and S_3.

4. Basic trajectories and multipliers. Throughout the remainder of the paper we assume that f has **pdg** structure at $\bar{x}$ relative to $\mathcal{P}$. Lemma 2.1 gives a spanning set of generators for $\mathcal{V}$ based on the structure. Next we consider linearly independent subsets of these generating vectors which lead to defining associated smooth trajectories and multiplier functions.

4.1. Basic index sets. An index set K of the form $K = K_f \cup K_\varphi \subseteq \{0, 1, \ldots, m_1\} \cup \{m_1 + 1, \ldots, m_1 + m_2\}$ with $0 \in K_f$ is called a *basic index set* if the $(n+1)$-dimensional vectors

$$\left\{ \begin{bmatrix} \nabla f_i(\bar{x}) \\ 1 \end{bmatrix} \right\}_{i \in K_f} \bigcup \left\{ \begin{bmatrix} \nabla \varphi_{i - m_1}(\bar{x}) \\ 0 \end{bmatrix} \right\}_{i \in K_\varphi}$$

are linearly independent.

In the theory that follows, we always suppose that $K = K_f \cup K_\varphi$ is a basic index set and, if necessary, (with the exception of Example 3.3), the f_is are reindexed so that the nonempty set K_f contains $i = 0$.

Associated with a basic index set K we define a full column rank $n \times (|K_f| - 1 + |K_\varphi|)$ matrix

$$\bar V := [\{\nabla f_i(\bar x) - \nabla f_0(\bar x)\}_{0 \neq i \in K_f} \cup \{\nabla \varphi_{i-m_1}(\bar x)\}_{i \in K_\varphi}].$$

This is a basis matrix for the $|K_f| - 1 + |K_\varphi|$ dimensional subspace of $\mathcal{V}$ defined by

$$\mathcal{V}_K := \operatorname{lin}\left(\{\nabla f_i(\bar x) - \nabla f_0(\bar x)\}_{i \in K_f} \cup \{\nabla \varphi_{i-m_1}(\bar x)\}_{i \in K_\varphi}\right) \tag{4.1}$$

with the interpretation that if K is a singleton then $\bar V$ is vacuous and $\mathcal{V}_K = \{0\}$.

When K is such that $\mathcal{V}_K = \mathcal{V}$, we say that K is a *transversal* basic index set.

Remark 4.1. *Recall that $\bar U$ is an orthogonal basis matrix for $\mathcal{U}$. Corresponding to a basic index set K and its associated subspace $\mathcal{V}_K \subseteq \mathcal{V}$ from* (4.1), *if K is transversal then $\bar V$ is a basis matrix for $\mathcal{V}$ which is not necessarily orthonormal. These basis matrices can be used to express the identity matrix in $\mathbb{R}^n$ as the sum of the projections onto the subspaces $\mathcal{U}$ and $\mathcal{V} = \mathcal{V}_K$:*

$$I = \bar U \bar U^\top + \bar V[\bar V^\top \bar V]^{-1} \bar V^\top. \tag{4.2}$$

Accordingly, for any $x \in \mathbb{R}^n$, $\bar U^\top x$ and $[\bar V^\top \bar V]^{-1}\bar V^\top x$ are, respectively, the $\mathcal{U}$ and $\mathcal{V}$ components of x.

Example **(continuation of Example 3.3)** For convenience we do not reindex the f_is so that index $i = 0$ is always in a basic index set K. Since for this function the vectors $\begin{bmatrix} \nabla f_i(\bar x) \\ 1 \end{bmatrix}$ are equal to $\begin{bmatrix} 0 \\ (-1)^{i+1} \\ 1 \end{bmatrix}$ for $i = 0, \ldots, 3$, there are no basic index sets with three or four elements. All of the singleton index sets are basic, but none of them is transversal, because $dim\, \mathcal{V} = 1$. Finally, the only two-element sets that are basic are

$$\{0,1\}, \{0,3\}, \{1,2\} \text{ and } \{2,3\}.$$

These four sets are all transversal and for each of them we take $\bar V = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to be the corresponding basis matrix for $\mathcal{V}$.

4.2. Smooth trajectories. The purpose of introducing basic index sets is to identify trajectories along which f behaves in a smooth manner. Our next result shows how to find a smooth trajectory that is tangent to $\mathcal{U}$ using the `pdg` structure of f and an implicit function theorem to parameterize the trajectory in terms of $u \in \mathbb{R}^{dim\, \mathcal{U}}$.

From here on, we assume that $\mathcal{U} \neq \{0\}$ so that $\dim \mathcal{U} \geq 1$ and $\mathcal{V} \neq \mathbb{R}^n$.

THEOREM 4.2. *Let f have* `pdg` *structure at $\bar x$ relative to $\mathcal{P}$ and suppose $K = K_f \cup K_\varphi$ is a nonsingleton basic index set. For all u small enough,*

(i) *the nonlinear system with variables* (u, v)

$$\begin{cases} f_i(\bar{x} + \bar{U}u + \bar{V}v) - f_0(\bar{x} + \bar{U}u + \bar{V}v) = 0, & 0 \neq i \in K_f, \\ \varphi_{i-m_1}(\bar{x} + \bar{U}u + \bar{V}v) = 0, & i \in K_\varphi, \end{cases} \tag{4.3}$$

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

(ii) *the trajectory*

$$\chi(u) := \bar{x} + \bar{U}u + \bar{V}v_K(u), \tag{4.4}$$

has a C^1 *Jacobian* $J\chi(u) = \bar{U} - \bar{V}(V(u)^\top \bar{V})^{-1}V(u)^\top \bar{U}$, *where*

$$V(u) := [\{\nabla f_i(\chi(u)) - \nabla f_0(\chi(u))\}_{0\neq i \in K_f} \bigcup \{\nabla \varphi_{i-m_1}(\chi(u))\}_{i\in K_\varphi}];$$

(iii) *in particular,* $v_K(0) = 0$, $\chi(0) = \bar{x}$, $V(0) = \bar{V}$, $Jv_K(0) = 0$, $J\chi(0) = \bar{U}$;

(iv) $v_K(u) = O(|u|^2)$ *and the trajectory* $\chi(u)$ *is tangent to* $\mathcal{U}$ *at* $\chi(0) = \bar{x}$.

Proof. These results follow from the asssumption that the structure functions are C^2 along the lines of [8, Theorem 5.1], by applying a second-order Implicit Function Theorem; see, for example, [3, p. 364]. ☐

When K is a singleton, $\bar{V}$ and $V(u)$ are vacuous. In this case, it is useful to define $\bar{V}v_K(u) := 0$, so that $\chi(u) = \bar{x} + \bar{U}u$ with Jacobian $\bar{U}$.

Remark 4.3. *When f is a convex function (not necessarily* pdg-*structured) the* $\mathcal{U}$-*Lagrangian introduced in [4] provides an alternative way to obtain smooth trajectories. More specifically, when* $\bar{V}$ *is a basis matrix for* $\mathcal{V}$, *trajectories* $\chi(u) = \bar{x} + \bar{U}u + \bar{V}v(u)$ *are given by means of the following minimization problem:*

$$v(u) \in \operatorname*{Argmin}_{v\in\mathbb{R}^{dim\,\mathcal{V}}}\{f(\bar{x} + \bar{U}u + \bar{V}v) - y^\top \bar{V}v\} = \operatorname*{Argmin}_{v\in\mathbb{R}^{dim\,\mathcal{V}}}\{f(\bar{x} + \bar{U}u + \bar{V}v) - z^\top v\},$$

where $y \in \partial f(\bar{x})$ *and* $z = \bar{V}^\top y$. *Such smooth trajectories give a convex function, called the* $\mathcal{U}$-*Lagrangian, which is defined by*

$$L_{\mathcal{U}}(u; z) := f(\chi(u)) - z^\top v(u). \tag{4.5}$$

When a $\mathcal{U}$-*Lagrangian depending on* $z = \bar{V}^\top y$ *has a Hessian in* u, *it is possible to obtain an expansion for* f *that is superlinear in* u, *since* $v(u) = o(|u|)$.

When f *is* pdg-*structured (not necessarily convex) the solution of system* (4.3), *depending on a basic index set* K, *gives* $v_K(u)$ *and a corresponding trajectory. In Section 5 below we give an associated "*$\mathcal{U}$*-Lagrangian-related" function whose definition is analogous to* (4.5), *except that it depends on* K. *This function is denoted by* $L_K(u; z)$ *and is called a "primal-dual" function. From* (4.5), *since both* $L_K(u; z)$ *and* $v_K(u)$ *are* C^2-*functions, we are able to obtain an expansion for* f *that is of second order in* u *(rather than only superlinear as in the case without special structure); see Theorems 5.1 and 5.2 below. When the* $\mathcal{V}$-*minimality condition of Theorem 6.1 below is satisfied, the function* $v_K(u)$ *becomes an above* $v(u)$, *even for a nonconvex function; see [8, Section 6] for sufficient conditions for* $\mathcal{V}$-*minimality when* f *is convex.*

We will show in Section 7 below that a pdg-structured function f is partly smooth relative to certain manifolds $\mathcal{M}_K$. More precisely, depending on a basic index set K we consider the following smooth manifold:

$$\mathcal{M}_K := \left\{ x \in B(\bar{x}) : \left\{ \begin{array}{lcl} f_i(x) - f_0(x) & = & 0, \quad 0 \neq i \in K_f \\ \varphi_{i-m_1}(x) & = & 0, \quad i \in K_\varphi \end{array} \right\} \right\}. \tag{4.6}$$

In order to establish a relation between a manifold $\mathcal{M}_K$ and the function f, the set K is required to satisfy a feasibility assumption depending on the following definition:

A basic index set K with corresponding trajectory $\chi(u)$ is *primal feasible* if for all $u \in \mathbb{R}^{\dim \mathcal{U}}$ small enough $\chi(u) \in \mathcal{P}$ and $f(\chi(u)) = f_i(\chi(u))$ for some (and, hence, all) $i \in K_f$.

LEMMA 4.4. *Let f have* pdg *structure at $\bar{x}$ relative to $\mathcal{P}$ and suppose K is a basic index set. Then the following hold:*

(i) $N_{\mathcal{M}_K}(\bar{x}) = \mathcal{V}_K$, *the subspace defined in* (4.1);

(ii) *if K is transversal and primal feasible, then for all x close enough to $\bar{x}$*

$$x \in \mathcal{M}_K \iff x = \bar{x} + \bar{U}u(x) + \bar{V}v_K(u(x)) \text{ where } u(x) := \bar{U}^\top(x - \bar{x}).$$

As a result, for $x \in \mathcal{M}_K$

(a) $x \in \mathcal{P}$ *and*

(b) *if $i \in \{0, 1, \ldots, m_1 + m_2\}$ is such that $f_i(x) \neq f(x)$ or $\varphi_{i-m_1}(x) \neq 0$, then $i \notin K$.*

Proof. From [15, p. 203], the subspace $\mathcal{V}_K$ as defined in (4.1) is the normal cone to $\mathcal{M}_K$ at $\bar{x}$, so (i) follows. Now we show that (ii) holds.
By the transversality assumption, $N_{\mathcal{M}_K}(\bar{x}) = \mathcal{V}$ and $\bar{V}$ is a basis matrix for $\mathcal{V}$. Take any $x \in \mathcal{M}_K$ and define $v(x) := [\bar{V}^\top\bar{V}]^{-1}\bar{V}^\top(x - \bar{x})$. From (4.2),

$$x - \bar{x} = \bar{U}\bar{U}^\top(x - \bar{x}) + \bar{V}[\bar{V}^\top\bar{V}]^{-1}\bar{V}^\top(x - \bar{x}), \text{ so } x = \bar{x} + \bar{U}u(x) + \bar{V}v(x).$$

With this notation, by the definition of $\mathcal{M}_K$ in (4.6), $(u, v) = (u(x), v(x))$ satisfies nonlinear system (4.3). Since the solution $v_K(u)$ is unique for a given u, $v(x) = v_K(u(x))$, so any $x \in \mathcal{M}_K$ has the form $x = \bar{x} + \bar{U}u(x) + \bar{V}v_K(u(x))$. The converse is immediate from (4.6) and Theorem 4.2(i) with $u = u(x)$. The remaining results follow from the assumption that K is primal feasible. ☐

Example **(continuation of Example 3.3)** Although singleton basic sets could be considered, we focus on the two-element sets that are basic and transversal, i.e., on $\{0, 1\}$, $\{0, 3\}$, $\{1, 2\}$ and $\{2, 3\}$.

Next we perform some calculations for $K = \{2, 3\}$ only and give a summary for all four sets in Table 4.1 below.

Since for this example

$$\bar{x} + \bar{U}u + \bar{V}v = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u + \begin{bmatrix} 0 \\ 1 \end{bmatrix} v = \begin{pmatrix} u \\ v \end{pmatrix},$$

when $K = \{2, 3\}$, system (4.3) corresponds to (recall that index $0 \notin K$ here)

$$f_2(u, v) - f_3(u, v) = 0 \quad \iff \quad 3u^2 - v - (-5u^2 + v) = 0,$$

which is solved for v by $v_K(u) = 4u^2$. Therefore, $\chi(u) = (u, 4u^2)^\top$. For $u \neq 0$, $\chi(u) \in S_2 \subset \mathcal{P}$, so $f(\chi(u)) = f_2(\chi(u)) = f_3(\chi(u))$. Thus, $K = \{2, 3\}$ is primal feasible.

Note that for none of the trajectories $\chi(u)$ from Table 4.1 is $f(\chi(u))$ equal to the maximum or the minimum over all four functions $f_i(\chi(u))$. Also, we see that only

$$K_1 := \{0, 1\} \quad \text{and} \quad K_2 := \{2, 3\} \tag{4.7}$$

K	$v_K(u)$	$\chi(u) \in S_i$	$f(\chi(u))$	$f_0(\chi(u))$	$f_1(\chi(u))$	$f_2(\chi(u))$	$f_3(\chi(u))$
$0,1$	0	$\begin{pmatrix} u \\ 0 \end{pmatrix} \in S_0$	u^2	u^2	u^2	$3u^2$	$-5u^2$
$0,3$	$3u^2$	$\begin{pmatrix} u \\ 3u^2 \end{pmatrix} \in S_2$	0	$-2u^2$	$4u^2$	0	$-3u^2$
$1,2$	u^2	$\begin{pmatrix} u \\ u^2 \end{pmatrix} \in S_1$	$\sqrt{2}u^2$	0	$2u^2$	$2u^2$	$-4u^2$
$2,3$	$4u^2$	$\begin{pmatrix} u \\ 4u^2 \end{pmatrix} \in S_2$	$-u^2$	$-3u^2$	$5u^2$	$-u^2$	$-u^2$

TABLE 4.1
Trajectories for Example 3.3.

are primal feasible basic index sets. Since they are also transversal, letting $u = \bar{U}x = x_1$, the manifolds corresponding to (4.6) given via Lemma 4.4 and Table 4.1 are

$$\{x = (x_1, x_2) \in \mathbb{R}^2 : f_0(x) = f_1(x)\} = \{(u, 0) : u \in \mathbb{R}\} = \mathcal{M}_1$$

and

$$\{x = (x_1, x_2) \in \mathbb{R}^2 : f_2(x) = f_3(x)\} = \{(u, 4u^2) : u \in \mathbb{R}\} = \mathcal{M}_2,$$

i.e.., the smooth manifolds from (3.1) and (3.4), respectively, containing the points of nondifferentiability of f.

4.3. Multiplier functions. So far we have developed only a primal object $\chi(u)$ depending on the pdg structure primal functions and on a basic index set K. Now we turn our attention to an associated dual object that is also a smooth function of $u \in \mathbb{R}^{dim\,\mathcal{U}}$. It is a multiplier vector $\alpha(u)$ which depends on structure function gradients as well as on $\chi(u)$ and on an arbitrary parameter vector y that need not be a subgradient at $\bar{x}$. The next result follows along the lines of Theorem 5.2 in [8].

THEOREM 4.5. *Suppose f has* pdg *structure at $\bar{x}$ with respect to $\mathcal{P}$. Corresponding to a basic index set $K = K_f \cup K_\varphi$ with trajectory $\chi(u) = \bar{x} + \bar{U}u + \bar{V}v_K(u)$ and to a parameter vector $y \in \mathbb{R}^n$, for each u small enough, the linear system with variables α_i*

$$\begin{cases} \bar{V}^\top \left(\sum_{i \in K_f} \alpha_i \nabla f_i(\chi(u)) + \sum_{i \in K_\varphi} \alpha_i \nabla \varphi_{i-m_1}(\chi(u)) \right) = \bar{V}^\top y \in \mathbb{R}^{|K|-1}, \\ \sum_{i \in K_f} \alpha_i = 1 \end{cases}$$

has a unique solution $\alpha = \alpha(u)$, given by

$$\begin{aligned} \{\alpha_i(u)\}_{0 \neq i \in K} &= (\bar{V}^\top V(u))^{-1} \bar{V}^\top (y - \nabla f_0(\chi(u))), \\ \alpha_0(u) &= 1 - \sum_{0 \neq i \in K_f} \alpha_i(u), \end{aligned}$$

where $V(u)$ is defined in Theorem 4.2(ii) if K is not a singleton and $\alpha_0(u) = 1$ otherwise.

Note that multipliers $\alpha_i(u)$ depend on K and $\bar{V}^\top y$, so we should think of the abbreviated vector notation $\alpha(u)$ as standing for something like $\alpha_K(u; \bar{V}^\top y)$. An expression for the Jacobian of $\alpha(u)$ is given in [8].

Example **(continuation of Example 3.3)** Table 4.2 below shows the results for obtaining multiplier functions for the basic index sets K_1 and K_2 defined in (4.7). The results follow from Table 4.1, (3.2) and the choice of $\bar{V} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ which implies that $\bar{V}^\top y = y_2$.

	$K_1 = \{0,1\}$	$K_2 = \{2,3\}$
$\chi(u)$	$\begin{pmatrix} u \\ 0 \end{pmatrix}$	$\begin{pmatrix} u \\ 4u^2 \end{pmatrix}$
$\sum_{i\in K} \alpha_i \nabla f_i(\chi(u))$	$\alpha_0 \begin{pmatrix} 2u \\ -1 \end{pmatrix} + \alpha_1 \begin{pmatrix} 2u \\ 1 \end{pmatrix}$	$\alpha_2 \begin{pmatrix} 6u \\ -1 \end{pmatrix} + \alpha_3 \begin{pmatrix} -10u \\ 1 \end{pmatrix}$
Linear system in Theorem 4.5	$\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = \begin{pmatrix} y_2 \\ 1 \end{pmatrix}$	same, except the variables are α_2, α_3
$\alpha(u)$	$\alpha_0(u) = (1 - y_2)/2$ $\alpha_1(u) = (1 + y_2)/2$	$\alpha_2(u) = (1 - y_2)/2$ $\alpha_3(u) = (1 + y_2)/2$

TABLE 4.2
Multipliers for Example 3.3 with $\bar{V}^\top = [0 \;\; 1]$.

5. Primal-dual functions; 2^{nd} order expansions. Consider a primal feasible basic index set K with associated quantities $\bar{V}$, $v_K(u)$ and $\chi(u)$, and let $z = \bar{V}^\top y$ where y is a parameter vector in $\mathbb{R}^n$. In view of Remark 4.3, in particular definition (4.5), we define for $u \in \mathbb{R}^{\dim \mathcal{U}}$ small enough the following primal-dual function:

$$L_K(u; z) := f(\chi(u)) - z^\top v_K(u). \tag{5.1}$$

Next we show that this function is C^2 and give explicit expressions for its first and second derivatives with respect to u.

THEOREM 5.1. *Let f have* pdg *structure at $\bar{x}$ with respect to $\mathcal{P}$. Suppose $K = K_f \cup K_\varphi$ is a primal feasible basic index set with corresponding trajectory $\chi(u)$ and basis matrix $\bar{V}$. Let $y \in \mathbb{R}^n$ be a parameter vector and consider the primal-dual function defined in* (5.1) *and the multiplier functions $\alpha_i(u)$ from Theorem 4.5 corresponding to K and to $z = \bar{V}^\top y \in \mathbb{R}^{\dim \mathcal{V}_K}$.*

Then for all u small enough we have the following:

(i) *$L_K(u; z)$ is a C^2-function of u satisfying the particular Lagrangian-like result that*

$$L_K(u; 0) = \left(\sum_{i \in K_f} \alpha_i(u) f_i(\chi(u)) + \sum_{i \in K_\varphi} \alpha_i(u) \varphi_{i-m_1}(\chi(u)) \right).$$

(ii) *The gradient of L_K is given by*

$$\nabla L_K(u; z) = \bar{U}^\top g_K(u; z),$$

where $g_K(u; z)$ is defined by

$$g_K(u; z) := \left(\sum_{i \in K_f} \alpha_i(u) \nabla f_i(\chi(u)) + \sum_{i \in K_\varphi} \alpha_i(u) \nabla \varphi_{i-m_1}(\chi(u)) \right). \tag{5.2}$$

(iii) *The Hessian of* L_K *is given by*

$$\nabla^2 L_K(u;z) = J\chi(u)^\top M_K(u;z) J\chi(u),$$

where $M_K(u;z)$ *is the* $n \times n$ *matrix function defined by*

$$M_K(u;z) := \sum_{i \in K_f} \alpha_i(u)\nabla^2 f_i(\chi(u)) + \sum_{i \in K_\varphi} \alpha_i(u)\nabla^2 \varphi_{i-m_1}(\chi(u)), \tag{5.3}$$

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

(iv) *In particular, for all* z, $L_K(0;z) = f(\bar{x})$, $\nabla L_K(0;z) = \bar{g}_{\mathcal{U}}$, *the* $\mathcal{U}$*-gradient of* f *at* $\bar{x}$ *defined in* (2.2), *and* $\nabla^2 L_K(0;z) = \bar{U}^\top M_K(0;z)\bar{U}$.

Proof. Since K is primal feasible, using (5.1) with $z = 0$ gives

$$L_K(u;0) = f_i(\chi(u)) \qquad \text{for all } i \in K_f.$$

In addition, since

$$0 = \varphi_{i-m_1}(\chi(u)) \qquad \text{for all } i \in K_\varphi,$$

multiplying each of the above equations by the appropriate multiplier $\alpha_i(u)$ and then summing and recalling that $\sum_{i\in K_f} \alpha_i(u) = 1$ gives the Lagrangian-like expression in item (i).

The result that L_K is C^2 in u and the next two corresponding items follow along the lines of [8, Theorem 6.3$(iii) - (iv)$], with $L_{\mathcal{U}}$ and $\bar{g}$ therein replaced by L_K and y, respectively.

Item (iv) follows because $v_K(0) = 0$, $\chi(0) = \bar{x}$, $J\chi(0) = \bar{U}$, $\sum_{i\in K_f} \alpha_i(0) = 1$ and for each $i \in K$, by (2.3):

$$\begin{array}{llcrcl} \text{either} & \nabla f_i(\bar{x}) \in \bar{\partial} f(\bar{x}) & \text{so} & \bar{U}^\top \nabla f_i(\bar{x}) & = & \bar{g}_{\mathcal{U}} \\ \text{or} & \nabla \varphi_{i-m_1}(\bar{x}) \in \mathcal{V} & \text{so} & \bar{U}^\top \nabla \varphi_{i-m_1}(\bar{x}) & = & 0. \end{array} \qquad \square$$

If K is primal feasible and y is such that $z = \bar{V}^\top y = 0$, we call the corresponding Hessian of L_K at $u = 0$ a *basic* $\mathcal{U}$*-Hessian for* f *at* $\bar{x}$ and denote it by $\bar{H}_K := \nabla^2 L_K(0;0)$. Second-order $\mathcal{U}-$derivatives are useful for specifying second-order expansions for f and giving related necessary conditions for optimizers, as is shown next.

THEOREM 5.2. *Let* f *have* pdg *structure at* $\bar{x}$ *with respect to* $\mathcal{P}$. *Suppose* K *is a primal feasible basic index set with corresponding trajectory* $\chi(u) = \bar{x} + \bar{U}u + \bar{V}v_K(u)$. *Then for all* u *small enough and all* $y \in \mathbb{R}^n$

$$f(\chi(u)) = \quad f(\bar{x}) + \bar{g}_{\mathcal{U}}^\top u + y^\top \bar{V} v_K(u) + \tfrac{1}{2} u^\top \nabla^2 L_K(0; \bar{V}^\top y) u + o(|u|^2),$$

where $\bar{g}_{\mathcal{U}}$ *is the* $\mathcal{U}$*-gradient defined in* (2.2).
In particular, when $y = \bar{g} \in \bar{\partial} f(\bar{x})$,

$$\begin{aligned} f(\chi(u)) = f(\bar{x}) + \bar{g}^\top(\chi(u) - \bar{x}) &+ \tfrac{1}{2}(\chi(u)-\bar{x}))^\top \bar{U} \nabla^2 L_K(0; \bar{V}^\top \bar{g}) \bar{U}^\top (\chi(u) - \bar{x}) \\ &+ o(|\chi(u) - \bar{x}|^2), \end{aligned}$$

or, when $\bar{V}^\top y = 0$,

$$f(\chi(u)) = f(\bar{x}) + \bar{g}_{\mathcal{U}}^\top u + \frac{1}{2} u^\top \bar{H}_K u + o(|u|^2).$$

Proof. The first expansion follows from expanding $L_K(u;z)$ with $z = \bar{V}^\top y$ about $u = 0$ and using (5.1) and Theorem 5.1(ii),(iii) and (iv). The second then follows from (2.2) and the result from Theorem 4.2 that $v_K(u) = O(|u|^2)$. The third expansion is a consequence of the definition of $\bar{H}_K$. ∎

COROLLARY 5.3. *Let f have* `pdg` *structure at $\bar{x}$ with respect to $\mathcal{P}$ and suppose $\bar{x}$ is a local minimizer of f. Then $0 \in \bar{\partial} f(\bar{x})$ and for any primal feasible basic index set K the associated basic $\mathcal{U}$-Hessian $\bar{H}_K$ is positive semidefinite.*

Example **(continuation of Example 3.3)** To compute the primal-dual functions and their derivatives given in Table 5.1 below, we substitute into (5.1) the relevant expressions from Table 4.1 and differentiate with respect to u. Equivalently, we could use the multipliers $\alpha(u)$ given in Table 4.2, and the formulæ from Theorem 5.1. Recall that $z = \bar{V}^\top y = y_2$ for this example.

	$K_1 = \{0,1\}$	$K_2 = \{2,3\}$
$L_K(u;y_2)$	u^2	$-u^2 - 4y_2u^2$
$\nabla L_K(u;y_2)$	$2u$	$-2(1+4y_2)u$
$\nabla^2 L_K(u;y_2)$	2	$-2(1+4y_2)$
$\bar{H}_K$	2	-2

TABLE 5.1
Primal-dual information for Example 3.3.

6. Link to second-order epi-derivatives. We now show that some $\mathcal{U}$-Hessians give second-order epi-derivatives for $\mathcal{U}$-subspace directions. Thus, primal-dual functions L_K, having ordinary second derivatives, can capture much of the second-order epi-differential behavior of a `pdg`-structured function.

We start by recalling the formal definition of epigraphical convergence.

6.1. Characterization of epi-limits and epi-derivatives. The epigraphical convergence theory developed in [15, Ch. 7] includes the following useful characterization of epi-limits.

Proposition 7.2 and eq. 7(3) in [15]

Let $\{q^\nu\}$ be a sequence of functions on $\mathbb{R}^n$, and let w be any point in $\mathbb{R}^n$. The value $q(w)$ is the *epi-limit* of the sequence q^ν at w if and only if

$$\begin{cases} \liminf_\nu q^\nu(w^\nu) \geq q(w) & \text{for every sequence } w^\nu \to w\,, \\ \limsup_\nu q^\nu(w^\nu) \leq q(w) & \text{for some sequence } w^\nu \to w\,. \end{cases}$$

∎

Note that in the expression above the limit $q(w)$ can be infinity.

For a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ and point $\bar{x} \in$ *dom* h with $h(\bar{x})$ finite we consider various limits of the following first-order and second-order difference quotients:

$$\frac{h(\bar{x}+\tau\cdot) - h(\bar{x})}{\tau} \quad \text{and} \quad \frac{h(\bar{x}+\tau\cdot) - h(\bar{x}) - \tau y^\top \cdot}{\frac{1}{2}\tau^2}$$

for $\tau > 0$ and $y \in \mathbb{R}^n$.

The *(first) subderivative* of h at $\bar{x}$ in the direction w is denoted and defined by

$$\mathrm{d}h(\bar{x})(w) := \liminf_{\tau \searrow 0, w^\tau \to w} \frac{h(\bar{x} + \tau w^\tau) - h(\bar{x})}{\tau}.$$

When the first-order difference quotient has an epi-limit at w as $\tau \searrow 0$ then $\mathrm{d}h(\bar{x})(w)$ is this limit and it is then called the *first epi-derivative* of h at $\bar{x}$ in the direction w and h is said to be *epi-differentiable* at $\bar{x}$ for w.

Similarly, but for each $y \in \mathbb{R}^n$, when the second-order difference quotient has an epi-limit at w as $\tau \searrow 0$, it is denoted by $\mathrm{d}^2 h(\bar{x}|y)(w)$ and is called the *second epi-derivative* of h at $\bar{x}$ relative to y in the direction w. In this case

$$\mathrm{d}^2 h(\bar{x}|y)(w) = \liminf_{\tau \searrow 0, w^\tau \to w} \frac{h(\bar{x} + \tau w^\tau) - h(\bar{x}) - \tau y^\top w^\tau}{\frac{1}{2}\tau^2},$$

generally called the *second subderivative* of h at $\bar{x}$ relative to y in the direction w.

6.2. Connection with $\mathcal{U}$-Hessians. It is important to note that, in the epigraphical setting, the second-order epi-derivative provides a second-order approximation $f(\bar{x}+\tau w) \approx f(\bar{x}) + \tau y^\top w + \frac{1}{2}\tau^2 \mathrm{d}^2 f(\bar{x}|y)(w)$, not in the usual sense of local uniform convergence, but of closeness of the epigraphs of the second-order difference quotient function and $\mathrm{d}^2 f(\bar{x}|y)(\cdot)$; see [13]. In contrast, primal-dual functions $L_K(u; \bar{V}^\top y)$ provide second-order approximations in u in the classical sense. We now establish a relation between these second-order objects when a basic index set K satisfies a property relative to certain subgradients which for a convex function follows from the $\mathcal{V}$-optimality conditions given in [7, Section 6.1].

THEOREM 6.1. *Let f have* pdg *structure at $\bar{x}$ with respect to $\mathcal{P}$. Suppose K is a transversal and primal feasible basic index set with corresponding trajectory $\chi(u) = \bar{x} + \bar{U}u + \bar{V}v_K(u)$ and primal-dual Hessian $\nabla^2 L_K(u; z)$ from Theorem 5.1(iii). Let*

$$G_K := \left\{ \bar{g} \in \bar{\partial} f(\bar{x}) : \begin{array}{l} v_K(u) \in \mathrm{Argmin}_{v \in \mathbb{R}^{dim\,\mathcal{V}}} \{ f(\bar{x} + \bar{U}u + \bar{V}v) - \bar{g}^\top \bar{V} v \} \\ \textit{for all small enough } u \in \mathbb{R}^{dim\,\mathcal{U}} \end{array} \right\}. \quad (6.1)$$

Then for all $w \in \mathbb{R}^n$ the corresponding first and second subderivatives, respectively, satisfy

$$\mathrm{d}f(\bar{x})(w) \geq \sup\{\bar{g}^\top w : \bar{g} \in G_K\}$$

and

$$\mathrm{d}^2 f(\bar{x}|\bar{g})(w) \geq w^\top \bar{U} \nabla^2 L_K(0; \bar{V}^\top \bar{g}) \bar{U}^\top w \quad \textit{for all } \bar{g} \in G_K.$$

Furthermore, for all $w \in \mathcal{U}$ the corresponding first and second epi-derivatives, respectively, are given by

$$\mathrm{d}f(\bar{x})(w) = \bar{g}_{\mathcal{U}}^\top \bar{U}^\top w$$

and

$$\mathrm{d}^2 f(\bar{x}|\bar{g})(w) = w^\top \bar{U} \nabla^2 L_K(0; \bar{V}^\top \bar{g}) \bar{U}^\top w \quad \textit{for all } \bar{g} \in G_K.$$

Proof. We start by showing the second-order results. Suppose $\bar{g} \in G_K$. Then, for all $v \in \mathbb{R}^{dim\,\mathcal{V}}$, $f(\chi(u)) - \bar{g}^\top \bar{V} v_K(u) \le f(\bar{x} + \bar{U}u + \bar{V}v) - \bar{g}^\top \bar{V} v$. Subtracting $f(\bar{x}) + \bar{g}^\top \bar{U} u$ from both sides of the inequality gives

$$f(\chi(u)) - f(\bar{x}) - \bar{g}^\top(\chi(u) - \bar{x}) \le f(\bar{x} + \bar{U}u + \bar{V}v) - f(\bar{x}) - \bar{g}^\top(\bar{U}u + \bar{V}v).$$

Then, since K is primal feasible, Theorem 5.2 written with $y = \bar{g}$ implies that for all $u \in \mathbb{R}^{dim\,\mathcal{U}}$ small enough and all $v \in \mathbb{R}^{dim\,\mathcal{V}}$

$$\frac{1}{2}u^\top \nabla^2 L_K(0; \bar{V}^\top \bar{g})u + o(|u|^2) \le f(\bar{x} + \bar{U}u + \bar{V}v) - f(\bar{x}) - \bar{g}^\top(\bar{U}u + \bar{V}v). \qquad (6.2)$$

Take a sequence $w^\tau \to w$ as $\tau \searrow 0$ and let $u^\tau := \bar{U}^\top w^\tau$. Then, using (4.2) and the transversality of K,

$$w^\tau = \bar{U}\bar{U}^\top w^\tau + \bar{V}[\bar{V}^\top \bar{V}]^{-1}\bar{V}^\top w^\tau = \bar{U}u^\tau + \bar{V}[\bar{V}^\top \bar{V}]^{-1}\bar{V}^\top w^\tau.$$

With this notation

$$\bar{x} + \tau w^\tau = \bar{x} + \bar{U}(\tau u^\tau) + \bar{V}[\bar{V}^\top \bar{V}]^{-1}\bar{V}^\top(\tau w^\tau).$$

From (6.2) with $u = \tau u^\tau \in \mathbb{R}^{dim\,\mathcal{U}}$, τ small enough and $v = [\bar{V}^\top \bar{V}]^{-1}\bar{V}^\top(\tau w^\tau) \in \mathbb{R}^{dim\,\mathcal{V}}$ we obtain

$$\frac{1}{2}(\tau u^\tau)^\top \nabla^2 L_K(0; \bar{V}^\top \bar{g})\tau u^\tau + o(|\tau u^\tau|^2) \le f(\bar{x} + \tau w^\tau) - f(\bar{x}) - \tau \bar{g}^\top w^\tau. \qquad (6.3)$$

Dividing both sides of this inequality by $\frac{1}{2}\tau^2$ yields

$$(u^\tau)^\top \nabla^2 L_K(0; \bar{V}^\top \bar{g})u^\tau + \frac{o(\tau^2|u^\tau|^2)}{\frac{1}{2}\tau^2} \le \frac{f(\bar{x} + \tau w^\tau) - f(\bar{x}) - \tau \bar{g}^\top w^\tau}{\frac{1}{2}\tau^2}.$$

Note that since $w^\tau \to w$, the definition of u^τ implies that $u^\tau \to \bar{U}^\top w$. Hence, passing to the limit gives the following inequality involving the second subderivative:

$$w^\top \bar{U} \nabla^2 L_K(0; \bar{V}^\top \bar{g})\bar{U}^\top w \le \mathrm{d}^2 f(\bar{x}|\bar{g})(w).$$

To show that the left hand side is an epi-limit for $w \in \mathcal{U}$ we exhibit below a sequence w_K^τ converging to w with the property that $\bar{x} + \tau w_K^\tau = \chi(\tau u^\tau)$. We have just shown, using Theorem 5.2, that the second-order difference quotient function of such a direction vector w_K^τ converges to $w^\top \bar{U} \nabla^2 L_K(0; \bar{V}^\top \bar{g})\bar{U}^\top w$, which therefore equals the second-order epi-derivative (by the above definition of epi-limit, since the liminf is greater than or equal to this limit).
Consider the sequence

$$w_K^\tau := \bar{U}u^\tau + \bar{V}\frac{1}{\tau}v_K(\tau u^\tau)) \text{ which implies } \bar{x} + \tau w_K^\tau = \chi(\tau u^\tau).$$

By Theorem 4.2(iv), the term $\frac{1}{\tau}v_K(\tau u^\tau) = \frac{1}{\tau}O(\tau^2|u^\tau|^2)$ converges to 0 together with τ. As a result, $\lim_\tau w_K^\tau = \lim_\tau \bar{U}u^\tau = \lim_\tau \bar{U}\bar{U}^\top w^\tau = \bar{U}\bar{U}^\top w$. The result follows by noting from (4.2) that $w \in \mathcal{U} \iff w = \bar{U}\bar{U}^\top w$.
The first-order results are obtained in a similar manner, by dividing (6.3) by τ instead of $\frac{1}{2}\tau^2$ and noting that if $w \in \mathcal{U}$ then $\bar{g}^\top w = \bar{g}^\top \bar{U}\bar{U}^\top w = \bar{g}_{\mathcal{U}}^\top \bar{U}^\top w$ from (2.2). $\square$

Remark 6.2. *When $\bar{g} \in G_K$ then from* (6.1) *and the definition of the minimizer $v(u)$ in Remark 4.3, whether or not f is convex, $L_K(u; \bar{V}^\top \bar{g})$ can be called the $\mathcal{U}$-Lagrangian relative to $\bar{V}^\top \bar{g}$ and, hence, $\nabla^2 L_K(0; \bar{V}^\top \bar{g})$ is the related $\mathcal{U}$-Hessian of f at $\bar{x}$. A similar generalization from the convex case to the nonconvex can be made for a fast track [10], i.e. if $G_K = \bar{\partial} f(\bar{x})$ then the corresponding $\chi(u)$ can be called a fast track. We note in passing that the dependence of the $\mathcal{U}$-Lagrangian on the pair $(u; z) \in \mathbb{R}^{dim\mathcal{U}} \times \mathbb{R}^{dim\mathcal{V}} = \mathbb{R}^n$ is exploited in [10] in order to apply an Implicit Function Theorem to show that for a convex f proximal points are on the fast track.*

Example **(continuation of Example 3.3)** Recall that $\bar{\partial} f(\bar{x}) = \{\bar{g} = (0, \bar{g}_2) : -1 \leq \bar{g}_2 \leq 1\}$. We will see below that for the transversal and primal feasible sets K_1 and K_2 from (4.7) the associated sets G_{K_1} and G_{K_2} from (6.1) satisfy $\bar{\partial} f(\bar{x}) = G_{K_1} \cup G_{K_2}$.

In view of Table 4.1, we first find values of $\bar{g}_2 \in [-1, 1]$ which correspond to $v_{K_1}(u) = 0$ or to $v_{K_2}(u) = 4u^2$ being in

$$\operatorname*{Argmin}_{v \in \mathbb{R}} \{f(u, v) - \bar{g}_2 v\}. \tag{6.4}$$

Table 6.1 below summarizes the main results for computing the various minimizers in (6.4). To obtain these results from the definition of f we partition $\mathbb{R}$ into four v-subintervals. For all $u \in \mathbb{R}$ and $\bar{g}_2 \in [-1, 1]$, on each subinterval the objective in (6.4) is concave in v (actually, it is affine on three subintervals and strictly concave on $[0, 2u^2]$, since $\dfrac{\partial^2 h_1}{\partial v^2} < 0$ there) and, hence, corresponding minimizers are at subinterval endpoints.

subinterval	objective in (6.4)	argmin value	objective min value
$v \leq 0$	$u^2 - (1 + \bar{g}_2)v$	$v = 0$ for all $\bar{g}_2$	u^2
$0 \leq v \leq 2u^2$	$h_1(u, v) - \bar{g}_2 v$	$\left[\begin{array}{l} v = 0 \text{ or } v = 2u^2, \\ \text{depending on } \bar{g}_2 \end{array}\right.$	$\min\{u^2, (1 - 2\bar{g}_2)u^2\}$
$2u^2 \leq v \leq 4u^2$	$3u^2 - (1 + \bar{g}_2)v$	$v = 4u^2$ for all $\bar{g}_2$	$-(1 + 4\bar{g}_2)u^2$
$4u^2 \leq v$	$-5u^2 + (1 - \bar{g}_2)v$	$v = 4u^2$ for all $\bar{g}_2$	$-(1 + 4\bar{g}_2)u^2$

TABLE 6.1
Subinterval v-minimizers for Example 3.3

From the last column in Table 6.1 the minimum objective value in (6.4) is

$$\min\{u^2, (1 - 2\bar{g}_2)u^2, -(1 + 4\bar{g}_2)u^2\} = \begin{cases} u^2 & \text{for } \bar{g}_2 \in [-1, -1/2] \\ -(1 + 4\bar{g}_2)u^2 & \text{for } \bar{g}_2 \in [-1/2, 1]. \end{cases}$$

Combined with the third column, this implies that

$$G_{K_1} = \left\{\bar{g} = (0, \bar{g}_2)^\top : \bar{g}_2 \in [-1, -\frac{1}{2}]\right\} \text{ and } G_{K_2} = \left\{\bar{g} = (0, \bar{g}_2)^\top : \bar{g}_2 \in [-\frac{1}{2}, 1]\right\}.$$

Thus, from Theorem 6.1 with $\bar{U}^\top w = w_1$ and Table 5.1 with $y_2 = \bar{g}_2$ we find that for all $w = (w_1, 0)^\top \in \mathcal{U}$ and $\bar{g} = (0, \bar{g}_2)^\top \in \bar{\partial} f(\bar{x})$ the second-order epi-derivatives are given by

$$\mathrm{d}^2 f(\bar{x}|\bar{g})(w) = \begin{cases} 2w_1^2 & \text{if } \bar{g}_2 \in [-1, -\frac{1}{2}] \\ -2(1 + 4\bar{g}_2)w_1^2 & \text{if } \bar{g}_2 \in [-\frac{1}{2}, 1]. \end{cases} \tag{6.5}$$

7. Link to partly smooth functions. The concept of partial smoothness, introduced in [5], is closely related to $\mathcal{VU}$-space decomposition and `pdg` structure. The smooth and "sharp" (generalized U- and V-shaped, respectively) behavior of a function are decoupled via an "active" smooth manifold, as, for example, in (4.6).

More precisely, see [5, Definition 2.7], given a set $\mathcal{M} \subset X$ that contains $\bar{x}$, the function $h : X \to \bar{\mathbb{R}}$ is *partly smooth at $\bar{x}$ relative to $\mathcal{M}$* if $\mathcal{M}$ is a manifold around $\bar{x}$ and the following four properties hold:

(i) **(restricted smoothness)** the restriction $h|_{\mathcal{M}}$ equals a representative function that is C^2 around $\bar{x}$;
(ii) **(regularity)** at every point close to $\bar{x}$ in $\mathcal{M}$, the function h is (subdifferentially) regular and has a subgradient;
(iii) **(V-type sharpness)** for all nonzero directions $w \in N_{\mathcal{M}}(\bar{x})$ the first subderivative satisfies $\mathrm{d}h(\bar{x})(w) > -\mathrm{d}h(\bar{x})(-w)$.
(iv) **(subgradient continuity)** the subdifferential map ∂h is continuous at $\bar{x}$ relative to $\mathcal{M}$.

We will show that regular `pdg`-structured functions are partly smooth relative to certain structure-defined manifolds. We start by showing the properties of V-type sharpness and restricted smoothness.

LEMMA 7.1. *Let f have* `pdg` *structure at $\bar{x}$ relative to $\mathcal{P}$ and suppose K is any basic index set with associated manifold $\mathcal{M}_K$ from* (4.6)*. If f is regular at $\bar{x}$ then*

$$\mathrm{d}f(\bar{x})(w) > -\mathrm{d}f(\bar{x})(-w) \text{ for all nonzero } w \in N_{\mathcal{M}_K}(\bar{x}).$$

Proof. Suppose w is a nonzero vector in $\mathcal{V}$. Let $\bar{g}$ be a subgradient in the relative interior of $\bar{\partial} f(\bar{x})$, which, from (2.1), is the interior of $\bar{\partial} f(\bar{x})$ relative to the affine set $g + \mathcal{V}$ for any subgradient g. Then,

$$\bar{g} + \tau w \in \bar{\partial} f(\bar{x}) \text{ for some } \tau > 0.$$

The regularity assumption implies, by [15, p. 337 and Theorem 8.30], that $\mathrm{d}f(\bar{x})(w) = \sup\{g^\top w : g \in \bar{\partial} f(\bar{x})\}$. Thus, for all nonzero $w \in \mathcal{V}$

$$\mathrm{d}f(\bar{x})(w) \geq \bar{g}^\top w + \tau |w|^2 > \bar{g}^\top w = -\bar{g}^\top(-w) \geq -\mathrm{d}f(\bar{x})(-w).$$

The proof is ended by noting, from Lemma 4.4(i), that $N_{\mathcal{M}_K}(\bar{x}) = \mathcal{V}_K \subseteq \mathcal{V}$. ◻

Next we show that the richness of `pdg` structure allows us to exhibit a *smooth representative of $f|_{\mathcal{M}_K}$* (in terms of a primal-dual function with $z = 0$) and to give its first and second derivatives.

THEOREM 7.2. *Let f have* `pdg` *structure at $\bar{x}$ with respect to $\mathcal{P}$. Suppose K is a basic index set that is primal feasible and transversal. Then the following hold:*

(i) *$L_K(u(x); 0)$, with $u(x) := \bar{U}^\top(x - \bar{x})$, is C^2 around $\bar{x} \in \mathbb{R}^n$ and equals f on $\mathcal{M}_K$;*
(ii) *The gradient and Hessian of $L_K(u(x); 0)$ are given, respectively, by*

$$\bar{U}\bar{U}^\top g_K(u(x); 0) \quad \text{and} \quad \bar{U}\nabla^2 L_K(u(x); 0)\bar{U}^\top.$$

These derivatives depend on expressions in Theorem 5.1(ii) and (iii) and Theorem 4.2(ii) with u and $\chi(u)$ therein replaced by $u(x)$ and x, respectively.

Proof. Take $x \in \mathcal{M}_K$ close enough to $\bar{x}$ so that $u(x)$ is small enough for $v_K(u(x))$ to be defined. For such $x \in \mathcal{M}_K$ Lemma 4.4(ii) gives $x = \chi(u(x))$, where $\chi(u)$ corresponds to K. Then (5.1) with $(u;z) = (u(x);0)$ gives $L_K(u(x);0) = f(x)$ for $x \in \mathcal{M}_K$ close enough to $\bar{x}$. Since the Jacobian of $u(x)$ is $\bar{U}^\top$, the results follow from Theorem 5.1 and the chain rule. □

In order to show continuity of the subdifferential on $\mathcal{M}_K$, we require an additional assumption on the basic index set K.

A basic index set $K = K_f \cup K_\varphi$ is called *dual feasible with respect to* $\bar{g} \in \bar{\partial} f(\bar{x})$ if there exists an $\bar{\alpha} \in \Delta$ such that

$$\bar{g} = \sum_{i \in K_f} \bar{\alpha}_i \nabla f_i(\bar{x}) + \sum_{i \in K_\varphi} \bar{\alpha}_i \nabla \varphi_{i-m_1}(\bar{x}) \text{ and } \bar{\alpha}_i = 0 \text{ for all } i \notin K. \tag{7.1}$$

THEOREM 7.3. *Let f have* pdg *structure at $\bar{x}$ with respect to $\mathcal{P}$. Suppose K is a basic index set that is dual feasible with respect to all $\bar{g} \in \bar{\partial} f(\bar{x})$. Then the following hold:*

(i) *K is transversal;*

(ii) *If, in addition, K is primal feasible, then $\bar{\partial} f$ is inner semicontinous at $\bar{x}$ relative to $\mathcal{M}_K$.*

Proof. To show (i), i.e., $\mathcal{V}_K = \mathcal{V}$, we only need to prove the inclusion $\mathcal{V} \subseteq \mathcal{V}_K$. Since the subgradient g used in defining $\mathcal{V}$ in (2.1) is arbitrary, take $g = \nabla f_0(\bar{x})$, which is in $\bar{\partial} f(\bar{x})$ by (2.3). Suppose $v \in \mathcal{V}$. Then there exist scalars β_j and subgradients $g^j \in \bar{\partial} f(\bar{x})$ such that

$$v = \sum_j \beta_j (g^j - \nabla f_0(\bar{x})). \tag{7.2}$$

By the assumption that K is dual feasible for all subgradients, there exist vectors $\alpha^j \in \Delta$ such that

$$g^j = \sum_{i \in K_f} \alpha_i^j \nabla f_i(\bar{x}) + \sum_{i \in K_\varphi} \alpha_i^j \nabla \varphi_{i-m_1}(\bar{x}) \text{ and } \alpha_i^j = 0 \text{ for all } i \notin K.$$

Then, from the definition of Δ, $\sum_{i \in K_f} \alpha_i^j = 1$ for each j, so

$$g^j - 1\nabla f_0(\bar{x}) = \sum_{i \in K_f} \alpha_i^j (\nabla f_i(\bar{x}) - \nabla f_0(\bar{x})) + \sum_{i \in K_\varphi} \alpha_i^j \nabla \varphi_{i-m_1}(\bar{x}).$$

Multiplying each of these equalities by β_j, and then summing over j, and using (7.2) gives

$$v = \sum_{i \in K_f} \gamma_i (\nabla f_i(\bar{x}) - \nabla f_0(\bar{x})) + \sum_{i \in K_\varphi} \gamma_i \nabla \varphi_{i-m_1}(\bar{x})$$

where γ_i is defined by $\gamma_i := \sum_j \beta_j \alpha_i^j$ for $i \in K$. Thus, from (4.1), $v \in \mathcal{V}_K$.

To show (ii), let $\bar{g} \in \bar{\partial} f(\bar{x})$ and let $\bar{\alpha} \in \Delta$ be a multiplier vector satisfying (7.1) in the definition of dual feasibility. Let $\{x^r\} \subset \mathcal{M}_K$ be a sequence converging to $\bar{x}$. Then the sequence $\{g^r\}$ defined by

$$g^r := \sum_{i \in K_f} \bar{\alpha}_i \nabla f_i(x^r) + \sum_{i \in K_\varphi} \bar{\alpha}_i \nabla \varphi_{i-m_1}(x^r),$$

converges to $\bar{g}$ by (7.1) and the continuity of the primal function gradients. To complete the proof, we need to show that $g^r \in \bar{\partial} f(x^r)$ using condition (iii)(b) of the `pdg` structure definition in Section 2.1. By item (i), K is transversal, so, by Lemma 4.4(ii), $x^r \in \mathcal{P}$ and if i is such that $f_i(x^r) \neq f(x^r)$ or $\varphi_{i-m_1}(x^r) \neq 0$, then $i \notin K$. From (7.1) $\bar{\alpha}_i = 0$ for $i \notin K$, so we have the complementary slackness between $x^r \in \mathcal{P}$ and $\bar{\alpha} \in \Delta$ needed to conclude from (7.1) and Section 2.1(iii)(b) with $x = x^r$ and $\alpha = \bar{\alpha}$ that $g^r \in \bar{\partial} f(x^r)$. ☐

We are now in a position to give the main result of this section.

THEOREM 7.4. *Let f have* `pdg` *structure at $\bar{x}$ with respect to $\mathcal{P}$. Suppose K is a primal feasible basic index set that is dual feasible with respect to all $\bar{g} \in \bar{\partial} f(\bar{x})$. If f is regular at all $x \in \mathcal{M}_K$ close to $\bar{x}$, then f is partly smooth at $\bar{x}$ relative to $\mathcal{M}_K$.*

Proof. Lemma 7.1 and Theorems 7.2 and 7.3 show all of the conditions for f to be partly smooth, except for outer semicontinuity of $\bar{\partial} f$ at $\bar{x}$ relative to $\mathcal{M}_K$. By the regularity assumption, on $\mathcal{M}_K$ $\bar{\partial} f(\cdot)$ equals the outer semicontinuous map $\partial f(\cdot)$; see [15, Proposition 8.7]. ☐

Example **(conclusion of Example 3.3)** It is shown in [5] that this example f is subdifferentially regular everywhere. From (3.2) and (3.3) with $(x_1, x_2) = (0, 0)$ and the fact that Δ is the canonical simplex in $\mathbb{R}^4$, it is easy to show that both of the primal feasible basic index sets $K_1 = \{0, 1\}$ and $K_2 = \{2, 3\}$ are dual feasible for all $\bar{g} \in \bar{\partial} f(\bar{x})$. Hence, this function is partly smooth at $\bar{x}$ relative to both $\mathcal{M}_1$ and $\mathcal{M}_2$. From Theorem 7.2 with $\bar{U}^\top = [1 \;\; 0]$ and Table 5.1 with $y_2 = 0$ the respective corresponding manifold restricted Hessians are

$$\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}.$$

8. Conclusion. In this paper we have related $\mathcal{U}-$subspace second-order smoothness results for the broad class of `pdg`-structured functions to the general concepts of second-order epi-derivatives and partly smooth functions. The principal results given here are the second-order expansions in u depending on a parameter y that need not be a subgradient. They are contained in Theorem 5.2 and based on a primal-dual function corresponding to a basic index set that is assumed only to be primal feasible. Without having to develop second-order epi-derivatives (as in Theorem 6.1) or manifold restricted Hessians (as in Theorem 7.2) for Example 3.3 we can conclude from Corollary 5.3 and Table 5.1 (having two basic $\mathcal{U}-$Hessians of opposite sign) that $\bar{x} = (0, 0)^\top$ is neither a local maximizer nor a local minimizer.

Theorem 6.1 provides somewhat more second-order information, since from (6.5), the second-order epi-derivatives for Example 3.3 vary from 2 to -10 as $\bar{g}_2$, the $\mathcal{V}$-component of the subgradients, varies from -1 to 1. This is at the expense of the additional assumption of $\mathcal{V}-$minimality, which provides the important link between primal-dual functions given here and $\mathcal{U}-$Lagrangians defined earlier for convex functions.

The connection to partly smooth functions in Theorem 7.4 requires a regularity assumption as well as dual feasibility for all subgradients. This latter assumption is important for obtaining inner semicontinuity of the subdifferential. This Theorem 7.3 result and its proof may have future application for rate of convergence analysis for minimization algorithms based on approximating fast tracks as in [9]. Such algorithms would try to exploit `pdg` structure implicitly without having to know it explicitly.

REFERENCES

[1] J. Burke, A. Lewis, and M. Overton, *Two numerical methods for optimizing matrix stability*, Linear Algebra and its Applications, 351-352 (2002), pp. 117–145.

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

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

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

[5] A. Lewis, *Active sets, nonsmoothness and sensitivity*, SIAM Journal on Optimization, (2002). To appear.

[6] 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.

[7] ———, *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.

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

[9] ———, *A $\mathcal{V}U$-proximal point algorithm for minimization.* In "http://www.sci.wsu.edu/math/faculty/mifflin/", 2001.

[10] ———, *Proximal points are on the fast track*, Journal of Convex Analysis, (2001). Accepted for publication.

[11] B. Mordukhovich, *Generalized differential calculus for nonsmooth and set-valued mappings*, J. Math. Anal. Appl., 183 (1994), pp. 250–288.

[12] M. Overton, *Large scale optimization of eigenvalues*, SIAM Journal on Optimization, 2 (1992), pp. 88–120.

[13] R. Poliquin and R. Rockafellar, *Second-order nonsmooth analysis in nonlinear programming*, in Recent advances in nonsmooth optimization, World Sci. Publishing, River Edge, NJ, 1995, pp. 322–349.

[14] R. Rockafellar, *First and second-order epi-differentiability in nonlinear programming*, Trans. Amer. Math. Soc., 307 (1988), pp. 75–108.

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

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

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