Proximal Points are on the Fast Track

Robert Mifflin* and Claudia Sagastizábal†

June 1, 2001

Abstract

For a convex function, we consider a space decomposition that allows us to identify a subspace on which a Lagrangian related to the function appears to be smooth. We study a particular trajectory, that we call a fast track, on which a certain second-order expansion of the function can be obtained. We show how to obtain such fast tracks for a general class of convex functions having primal-dual gradient structure. Finally, we show that for a point near a minimizer its corresponding proximal point is on the fast track.

Keywords Convex minimization, proximal points, second-order derivatives, $\mathcal{VU}$-decomposition.

AMS Subject Classification Primary: 49K35, 49M27. Secondary: 65K10, 90C25.

1 Introduction and motivation

Consider the problem

$$\min_{x \in \mathbb{R}^n} f(x)\,, \tag{1}$$

where f is a lower semicontinuous convex function. A classical conceptual algorithm for solving (1) is the proximal point method, based on the Moreau-Yosida regularization of f, [Mor65], [Yos64], [Mar70], [Roc76]. Implementable forms of the method can be obtained by means of a bundle technique, alternating serious steps with sequences of null steps, [Aus87], [Fuk84], [HUL93].

More recently, new conceptual schemes for solving (1) have been developed by using an approach that is somewhat different from Moreau-Yosida regularization.

*Department of Pure and Applied Mathematics, Washington State University, Pullman, WA 99164-3113. Research supported by the National Science Foundation under Grant Nos. DMS-9703952 and DMS-0071459 `mifflin@math.wsu.edu`

†IMPA, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro RJ 22460-320, Brazil. Also INRIA, BP 105, 78153 Le Chesnay, France. Research supported by FAPERJ (Brazil) under Grant No.E26/150.581/00 and by CNPq (Brazil) under Grant No. 301800/96. `sagastiz@impa.br`

This is the $\mathcal{VU}$-theory introduced in [LOS00] and further studied in [MS99], [MS00b], [MS00a], [Ous00]. The basic idea is to decompose $\mathbb{R}^n$ into two orthogonal subspaces $\mathcal{V}$ and $\mathcal{U}$ depending on a point in such a way that near the point f's nonsmoothness is concentrated essentially in $\mathcal{V}$. When f satisfies certain structural properties, it is possible to find smooth trajectories, tangent to $\mathcal{U}$, yielding a second-order expansion for f. The resulting $\mathcal{VU}$-algorithms make a step in the $\mathcal{V}$-subspace, followed by a $\mathcal{U}$-Newton move in order to obtain superlinear convergence. However, implementability depends on being able to approximate these subspaces well enough.

The purpose of this paper is to determine links between the two schemes mentioned above. More specifically, we uncover the relation between the proximal point mapping and a particular smooth trajectory, or *fast track*, defined from $\mathcal{VU}$-space decomposition theory. We address purely theoretical questions here as a prerequisite for future development of rapidly convergent implementable algorithms. Along the lines of [LOS00, §§ 4.3] and [MS99, § 3], such algorithms will exploit the smoothness on the $\mathcal{U}$-subspace of a certain Lagrangian for f in order to apply Newton-like methods on the fast track. Implementability via bundling techniques will follow from the crucial relation between fast tracks and the proximal points established in Theorem 5 below.

Our paper is organized as follows. We start by recalling the main elements of the $\mathcal{VU}$-space decomposition theory in § 2. The new concept of a fast track is defined in section 2.3. Initial properties of fast tracks are given in section 3. We show how to obtain fast tracks for a large class of functions in section 4. This class, consisting of functions with *primal-dual gradient* (`pdg`) structure, was introduced and thoroughly studied in [MS00b]. In Theorem 3 we show that a `pdg`-structured function that satisfies strong transversality at a minimizer has a fast track. Moreover, for such fast tracks the $\mathcal{VU}$-decomposition has certain basis matrix functions that are C^1 on a ball about $0 \in \mathcal{U}$. In section 5 we give our main result, relating proximal points to fast tracks. In addition, we show that a Newton-step based on the Moreau-Yosida regularization is equivalent to a proximal step followed by a Newton-step in the $\mathcal{U}$-subspace. We finish in section 6 with some concluding remarks on current algorithmic research.

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 $v(\cdot)$. The identity matrices in $\mathbb{R}^n$, $\mathbb{R}^{dim\,\mathcal{U}}$, and $\mathbb{R}^{dim\,\mathcal{V}}$ are denoted, respectively, by I, $I_\mathcal{U}$, and $I_\mathcal{V}$. Given a set Y, we denote by $\mathrm{lin}Y$ its linear hull.

2 Some elements of $\mathcal{VU}$-theory

Here we introduce some important concepts needed for our development, namely, $\mathcal{VU}$-space decomposition, $\mathcal{U}$-Lagrangians and fast tracks.

2.1 $\mathcal{VU}$-space decomposition

We start by recalling some concepts from [LOS00] and [MS00b]. For a convex function f, let g be any subgradient in $\partial f(\bar{x})$, the subdifferential of f at $\bar{x} \in \mathbb{R}^n$. Then the orthogonal subspaces

$$\mathcal{V} := \text{lin}(\partial f(\bar{x}) - g) \qquad \text{and} \qquad \mathcal{U} := \mathcal{V}^{\perp} \tag{2}$$

define the $\mathcal{VU}$-space decomposition at $\bar{x}$ of [LOS00, §2]: $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{V}$. From this definition, the relative interior of $\partial f(\bar{x})$, denoted by $\text{ri}\partial f(\bar{x})$, is the interior of $\partial f(\bar{x})$ relative to its affine hull, a manifold that is parallel to $\mathcal{V}$.

Letting $\bar{V}$ be a basis matrix for $\mathcal{V}$, not necessarily orthonormal, and letting $\bar{U}$ be an orthonormal basis matrix for $\mathcal{U}$ every $x \in \mathbb{R}^n$ can be decomposed into components $x_{\mathcal{U}}$ and $x_{\mathcal{V}}$ as follows:

$$\begin{array}{rclcccl} \mathbb{R}^n \ni x & = & \bar{U} \ (\bar{U}^\top x) & + & \bar{V} & \left([\bar{V}^\top \bar{V}]^{-1}\bar{V}^\top x\right) & \\ & = & \bar{U} \quad x_{\mathcal{U}} & + & \bar{V} & x_{\mathcal{V}} & \\ & = & \quad x_{\mathcal{U}} & \oplus & & x_{\mathcal{V}} & \in \mathbb{R}^{dim\,\mathcal{U}} \times \mathbb{R}^{dim\,\mathcal{V}} . \end{array}$$

The reason why we do not assumte that $\bar{V}$ is orthonormal is because typical $\mathcal{V}$-basis matrix approximations made by bundle methods are not orthonormal.

Note that, since $[\bar{V}|\bar{U}]$ is a basis for $\mathbb{R}^n$, the identity in $\mathbb{R}^n$ can be written as $I = \bar{U}\bar{U}^\top + \bar{V}[\bar{V}^\top\bar{V}]^{-1}\bar{V}^\top$.

2.2 $\mathcal{U}$-Lagrangians of convex functions

Given a subgradient $\bar{g} \in \partial f(\bar{x})$ with $\mathcal{V}$-component $\bar{g}_{\mathcal{V}} = ([\bar{V}^\top\bar{V}]^{-1}\bar{V}^\top)\bar{g}$, the $\mathcal{U}$-Lagrangian of f, depending on $\bar{g}_{\mathcal{V}}$, is defined by

$$\mathbb{R}^{dim\,\mathcal{U}} \ni u \mapsto L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}}) := \min_{v \in \mathbb{R}^{dim\,\mathcal{V}}} \left\{ f(\bar{x} + \bar{U}u + \bar{V}v) - \bar{g}^\top \bar{V} v \right\}. \tag{3}$$

Note that employing the scalar product induced by $\bar{V}^\top\bar{V}$ yields the $\mathcal{U}$-Lagrangian expression from [LOS00], $L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}}) = \min_v\{f(\bar{x} + u \oplus v) - \langle \bar{g}_{\mathcal{V}}, v\rangle\}$. The vector $\bar{g}_{\mathcal{V}}$ in our notation $L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}})$ plays the role of a multiplier vector, such as one that occurs in a Lagrangian from constrained optimization, because multipliers coming from the subspace minimization in (3) depend on $\bar{g}_{\mathcal{V}}$.

Each $\mathcal{U}$-Lagrangian is a convex function that is differentiable at $u = 0$ with

$$\nabla L_{\mathcal{U}}(0;\bar{g}_{\mathcal{V}}) = \bar{g}_{\mathcal{U}} = \bar{U}^\top \bar{g} = \bar{U}^\top g \quad \text{ for all } g \in \partial f(\bar{x}). \tag{4}$$

For $u \neq 0$ Theorem 3.3(i) in [LOS00] gives the following expression for each subdifferential:

$$\partial L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}}) = \{ g_{\mathcal{U}} : g_{\mathcal{U}} \oplus \bar{g}_{\mathcal{V}} \in \partial f(\bar{x} + u \oplus w)\} \tag{5}$$

where $\bar{V}w$ is an arbitrary element in $W(u;\bar{g}_{\mathcal{V}})$, the set of $\mathcal{V}$-space minimizers defined by

$$W(u;\bar{g}_{\mathcal{V}}) := \{ \bar{V}v : L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}}) = f(\bar{x} + \bar{U}u + \bar{V}v) - \bar{g}^\top\bar{V}v\}. \tag{6}$$

Whenever $\nabla^2 L_{\mathcal{U}}(0;\bar{g}_{\mathcal{V}})$ exists, those $\bar{x}+u\oplus w$ with $\bar{V}w \in W(u;\bar{g}_{\mathcal{V}})$ yield the following expansion of f:

$$f(\bar{x}+u\oplus w) = f(\bar{x}) + \bar{g}^\top(u\oplus w) + \frac{1}{2}u^\top \nabla^2 L_{\mathcal{U}}(0;\bar{g}_{\mathcal{V}})u + o(\|u\|^2)\,. \tag{7}$$

Furthermore, when $\bar{g} \in \mathrm{ri}\,\partial f(\bar{x})$, each $w \in W(u;\bar{g}_{\mathcal{V}})$ is $o(\|u\|)$ ([LOS00, Corollary 3.5]).

Next we develop and name a particular trajectory $\bar{x}+u\oplus w$.

2.3 Fast tracks to minimizers

Definition 1 *Let $f:\mathbb{R}^n \to \mathbb{R}$ be a convex function with minimizer $\bar{x}\in\mathbb{R}^n$ so that $0\in\partial f(\bar{x})$. We say that $\bar{x}+u\oplus v(u)$ is a* fast track *leading to a minimizer of f if for all u small enough*

(i) $v:\mathbb{R}^{dim\,\mathcal{U}} \mapsto \mathbb{R}^{dim\,\mathcal{V}}$ is a C^2-function satisfying $\bar{V}v(u) \in W_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}})$ for all $\bar{g}\in\mathrm{ri}\,\partial f(\bar{x})$; and

(ii) the particular $\mathcal{U}$-Lagrangian $L_{\mathcal{U}}(u;0)$ is a C^2-function.

When we write $v(u)$ we implicitly assume that $dim\,\mathcal{U}\geq 1$ and $dim\,\mathcal{V}\geq 1$. For the remaining two cases we define the fast track to be the singleton $\{\bar{x}\}$ if $dim\,\mathcal{U}=0$ and an n-dimensional ball about $\bar{x}$ on which $L_{\mathcal{U}}(u;0)=f(\bar{x}+\bar{U}u)$ is C^2 if $dim\,\mathcal{V}=0$. $\square$

Theorem 3.2(iv) in [LOS00] establishes that $W_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}})$ is nonempty for all $\bar{g}\in\mathrm{ri}\,\partial f(\bar{x})$. We now show that a similar nonemptiness result holds for $\bar{g}=0\in\partial f(\bar{x})$, not necessarily in the relative interior, whenever condition (i) above holds.

Lemma 1 *Let $f:\mathbb{R}^n\to\mathbb{R}$ be a convex function with minimizer $\bar{x}\in\mathbb{R}^n$. Suppose that $\bar{x}+u\oplus v(u)$ satisfies condition (i) of Definition 1. Then $\bar{V}v(u)\in W_{\mathcal{U}}(u;0)$ and $L_{\mathcal{U}}(u;0)=f(\bar{x}+u\oplus v(u))$.*

Proof. If $0\in\mathrm{ri}\,\partial f(\bar{x})$, the result follows from (3), (6) with $\bar{g}=0$ and condition (i). Otherwise, $0\in\mathrm{cl}\,\partial f(\bar{x})$, the closure of $\partial f(\bar{x})$. From Lemma III.2.1.6 in [HUL93], for any $\bar{g}\in\mathrm{ri}\,\partial f(\bar{x})$, the convex combination $\alpha\bar{g}+(1-\alpha)0=\alpha\bar{g}$ is in $\mathrm{ri}\,\partial f(\bar{x})$ for all $\alpha\in(0,1]$. For purposes of contradiction, suppose that the result does not hold. Then there exists $w\in\mathbb{R}^{dim\,\mathcal{V}}$ such that $f(\bar{x}+u\oplus v(u)) > f(\bar{x}+\bar{U}u+\bar{V}w)$. Choose α positive and so small that the small quantity $\alpha\bar{g}^\top\bar{V}(v(u)-w)$ satisfies

$$f(\bar{x}+u\oplus v(u)) > f(\bar{x}+\bar{U}u+\bar{V}w) + \alpha\bar{g}^\top\bar{V}(v(u)-w)\,.$$

Then $f(\bar{x}+u\oplus v(u)) - \alpha\bar{g}^\top\bar{V}v(u) > f(\bar{x}+\bar{U}u+\bar{V}w) - \alpha\bar{g}^\top\bar{V}w$, which contradicts the fact that $\bar{V}v(u)\in W_{\mathcal{U}}(u;\alpha\bar{g}_{\mathcal{V}})$, since $\alpha\bar{g}\in\mathrm{ri}\,\partial f(\bar{x})$. $\square$

Remark 1 The above proof can be modified to show that $\bar{V}v(u) \in W_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}})$ for all $\bar{g} \in \partial f(\bar{x})$. So, the symbol "ri" could be deleted in Definition 1(i). However, we prefer to keep it as written in order to have a smaller set of vectors to deal with when showing satisfaction of (i), as for example in the proof of Theorem 3 below. $\square$

The terminology "fast track" can be understood by examining the limiting differential behavior of the trajectory under consideration:

Tangency to $\mathcal{U}$: By [LOS00, Corollary 3.5], since $\mathrm{ri}\,\partial f(\bar{x})$ is nonempty, condition (i) implies that $v(u) = o(\|u\|)$ and, hence,

$$v(0) = 0 \qquad \text{and} \qquad Jv(0) = 0\,. \tag{8}$$

Thus, $\bar{x} + u \oplus v(u)$ is a trajectory leading to $\bar{x}$ which is tangent to $\mathcal{U}$.

Minimality: Since $\bar{x}$ minimizes f, $0 \in \partial f(\bar{x})$, and relation (4) with $g = 0$ gives

$$\nabla L_{\mathcal{U}}(0;\bar{g}_{\mathcal{V}}) = 0 \quad \text{for all } \bar{g} \in \partial f(\bar{x})\,. \tag{9}$$

Thus, each $L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}})$ is minimized at $u = 0$ and any subgradient $\bar{g} \in \partial f(\bar{x})$ has the form $\bar{g} = 0 \oplus \bar{g}_{\mathcal{V}}$.

Second-order smoothness: Lemma 1, combined with the fact that $L_{\mathcal{U}}(0;0) = f(\bar{x})$ and with (9) gives the second-order expansion

$$L_{\mathcal{U}}(u;0) = f(\bar{x} + u \oplus v(u)) = f(\bar{x}) + \frac{1}{2}u^{\top}\nabla^2 L_{\mathcal{U}}(0;0)u + o(\|u\|^2)\,. \tag{10}$$

Equivalents of the above fast track properties are employed by many authors to obtain rapidly convergent algorithms for minimizing maximum eigenvalue functions, see [OW95], [SF95], [LO96], [Ous99], [Ous00].

3 Additional properties of fast tracks

Here we give some consequences of our new $\mathcal{VU}$-theory definition. Associated with a fast track we define the matrices

$$B_{\mathcal{U}}(u) := \bar{U} + \bar{V}Jv(u) \quad \text{and} \quad B_{\mathcal{V}}(u) := \bar{V} - \bar{U}B_{\mathcal{U}}(u)^{\top}\bar{V} \tag{11}$$

of sizes $n \times \mathit{dim}\,\mathcal{U}$ and $n \times \mathit{dim}\,\mathcal{V}$, respectively. Note that $B_{\mathcal{U}}(u)$ is the Jacobian matrix of the fast track $\bar{x} + \bar{U}u + \bar{V}v(u)$. Also, since $\bar{U}$ and $\bar{V}$ are orthogonal, and $\bar{U}^{\top}\bar{U} = I_{\mathcal{U}}$,

$$\bar{U}^{\top}B_{\mathcal{U}}(u) = I_{\mathcal{U}}\,, \quad \text{and} \quad B_{\mathcal{V}}(u)^{\top}\bar{V} = \bar{V}^{\top}\bar{V}\,. \tag{12}$$

Other properties of these two matrices are given in the following proposition.

Proposition 1 *Let $\bar{x} + u \oplus v(u)$ be a fast track as described in Definition 1. For all u small enough the following hold:*

(i) both $B_{\mathcal{U}}(u)$ and $B_{\mathcal{V}}(u)$ are C^1, with $B_{\mathcal{U}}(u) \to \bar{U}$ and $B_{\mathcal{V}}(u) \to \bar{V}$ as $u \to 0$;

(ii) both $B_{\mathcal{U}}(u)^\top B_{\mathcal{U}}(u)$ and $B_{\mathcal{V}}(u)^\top B_{\mathcal{V}}(u)$ are invertible matrices;

(iii) $[B_{\mathcal{V}}(u)|B_{\mathcal{U}}(u)]$ is a basis for $\mathbb{R}^n$ with $B_{\mathcal{V}}(u)$ and $B_{\mathcal{U}}(u)$ orthogonal; and

(iv) for any $g := g_{\mathcal{U}} \oplus g_{\mathcal{V}} \in \mathbb{R}^n$

$$B_{\mathcal{U}}(u)^\top g = g_{\mathcal{U}} + B_{\mathcal{U}}(u)^\top \bar{V} g_{\mathcal{V}} \qquad \text{and} \qquad B_{\mathcal{V}}(u)^\top g = B_{\mathcal{V}}(u)^\top \bar{U} g_{\mathcal{U}} + \bar{V}^\top \bar{V} g_{\mathcal{V}} \,. \tag{13}$$

Proof. From Definition 1(i), $B_{\mathcal{U}}(\cdot)$ is C^1 and from (8), as $u \to 0$, $B_{\mathcal{U}}(u) \to \bar{U}$ and, hence, $\bar{V}^\top B_{\mathcal{U}}(u) \to 0$. As a result, $B_{\mathcal{V}}(\cdot)$ is also a C^1 function and $B_{\mathcal{V}}(u) \to \bar{V}$ as $u \to 0$ and (i) follows.
Since $[\bar{V}|\bar{U}]$ is a basis for $\mathbb{R}^n$, from (i) it follows that, for all u small enough, the columns of $B_{\mathcal{V}}(u)$ and of $B_{\mathcal{U}}(u)$ are also linearly independent, so (ii) follows.
We now abbreviate notation and drop the symbol (u) so that, for example, $B_{\mathcal{V}}(u)$ becomes $B_{\mathcal{V}}$. From (11) and the first equality in (12),

$$B_{\mathcal{V}}^\top B_{\mathcal{U}} = \left(\bar{V}^\top - \bar{V}^\top B_{\mathcal{U}} \bar{U}^\top\right) B_{\mathcal{U}} = \bar{V}^\top B_{\mathcal{U}} - \bar{V}^\top B_{\mathcal{U}} \bar{U}^\top B_{\mathcal{U}} = \bar{V}^\top B_{\mathcal{U}} - \bar{V}^\top B_{\mathcal{U}} = 0 \,,$$

which is the orthogonality needed to show (iii).
Finally, to see that both equalities in (13) hold, write $g = \bar{U} g_{\mathcal{U}} + \bar{V} g_{\mathcal{V}}$ and use, respectively, the transpose of the first equality and the second equality in (12). □

From here on we assume that $0 \in \mathrm{ri}\partial f(\bar{x})$ in order to eventually obtain differentiability of basis matrix functions for the $\mathcal{VU}$-decomposition on a fast track and to be able to apply the Implicit Function Theorem in the proof of Theorem 4 below.

Lemma 2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function with minimizer $\bar{x} \in \mathbb{R}^n$. Suppose that $0 \in \mathrm{ri}\partial f(\bar{x})$ and let $\bar{x} + u \oplus v(u)$ be a fast track as described in Definition 1. For each $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$ and all u small enough the following hold:*

(i) the gradient of $L_{\mathcal{U}}(\cdot; \bar{g}_{\mathcal{V}})$ has the expression

$$\nabla L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) = \nabla L_{\mathcal{U}}(u; 0) - B_{\mathcal{U}}(u)^\top \bar{V} \bar{g}_{\mathcal{V}} \,.$$

Furthermore, $\nabla L_{\mathcal{U}}(\cdot;\cdot)$ is a C^1-function on a ball about $(u; \bar{g}_{\mathcal{V}}) = (0; 0) \in \mathcal{U} \times \mathcal{V}$; and

(ii) the vector $\nabla L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) \oplus \bar{g}_{\mathcal{V}}$ is in $\partial f(\bar{x} + u \oplus v(u))$.

Proof. [(i)] From item (i) in Definition 1, $\bar{V}v(u)$ is a minimizer in (3), which means that

$$L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}}) = f(\bar{x}+u\oplus v(u)) - \bar{g}^\top \bar{V}v(u)\,.$$

Lemma 1 together with the identity $(\bar{V}^\top\bar{g})^\top = (\bar{V}^\top\bar{V}\bar{g}_{\mathcal{V}})^\top$ yields

$$L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}}) = L_{\mathcal{U}}(u;0) - \bar{g}_{\mathcal{V}}^\top\bar{V}^\top\bar{V}v(u)\,.$$

Because the right hand side of this equation is differentiable with respect to u,

$$\nabla L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}}) = \nabla L_{\mathcal{U}}(u;0) - (\bar{V}Jv(u))^\top\bar{V}\bar{g}_{\mathcal{V}}\,.$$

The first stated result follows from the fact that, by (11), $(\bar{V}Jv(u))^\top\bar{V} = (B_{\mathcal{U}}(u) - \bar{U})^\top\bar{V} = B_{\mathcal{U}}(u)^\top\bar{V}$, because $\bar{U}^\top\bar{V} = 0$.
Since $0 \in \mathrm{ri}\partial f(\bar{x})$, the first stated result in (u) holds for all $\bar{g}_{\mathcal{V}}$ in a ball about $0 \in \mathcal{V}$. Thus, because $\nabla L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}})$ is affine on $\bar{g}_{\mathcal{V}}$, and both $\nabla L_{\mathcal{U}}(u;0)$ and $B_{\mathcal{U}}(u)$ are C^1 in u, by Definition 1 (ii) and (i), respectively, $\nabla L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}})$ is C^1 in the pair $(u;\bar{g}_{\mathcal{V}})$ about $(0;0)$.
[(ii)] Since here $\partial L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}})$ is the singleton $\{\nabla L_{\mathcal{U}}(u;\bar{g}_{\mathcal{V}})\}$, taking $w = v(u)$ in (5) gives the desired result. □

Our next two theorems lay the groundwork for later obtaining $\mathcal{VU}$ basis matrix functions along a fast track for certain functions with *primal-dual gradient* structure; see section 4 below.

We let $\mathcal{V}(u)$ be the $\mathcal{V}$-subspace relative to $\bar{x}+u\oplus v(u)$ defined from $\partial f(\bar{x}+u\oplus v(u))$ as in (2) with $\bar{x}$ replaced by $\bar{x}+u\oplus v(u)$, i.e., for any $g_0 \in \partial f(\bar{x}+u\oplus v(u))$

$$\mathcal{V}(u) := \mathrm{lin}(\partial f(\bar{x}+u\oplus v(u)) - g_0)\,. \tag{14}$$

Theorem 1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function with minimizer $\bar{x} \in \mathbb{R}^n$. Suppose that $0 \in \mathrm{ri}\partial f(\bar{x})$ and let $\bar{x}+u\oplus v(u)$ be a fast track as described in Definition 1. For all u small enough the following three statements are equivalent:*

(i) the matrix $B_{\mathcal{V}}(u)$ is a basis for $\mathcal{V}(u)$;

(ii) the matrix $B_{\mathcal{U}}(u)$ is a basis for $\mathcal{V}(u)^\perp$; and

(iii) for all $g \in \partial f(\bar{x}+u\oplus v(u))$

$$B_{\mathcal{U}}(u)^\top g = \nabla L_{\mathcal{U}}(u;0)\,.$$

Proof. [$(i) \Leftrightarrow (ii)$] This is straightforward from Proposition 1(iii).
In order to show [$(ii) \Rightarrow (iii) \Rightarrow (i)$] we use Lemma 2$(ii)$ with $\bar{g} = 0 \in \mathrm{ri}\partial f(\bar{x})$ to write

$$\bar{U}\nabla L_{\mathcal{U}}(u;0) = \bar{U}\nabla L_{\mathcal{U}}(u;0) + \bar{V}0 = \nabla L_{\mathcal{U}}(u;0) \oplus 0 \in \partial f(\bar{x}+u\oplus v(u))\,. \tag{15}$$

If (ii) holds, then $B_{\mathcal{U}}(u)^\top(g_1 - g_2) = 0$ for any $g_1, g_2 \in \partial f(\bar{x} + u \oplus v(u))$, since $g_1 - g_2 \in \mathcal{V}(u)$. Thus, $B_{\mathcal{U}}(u)^\top g$ is the same constant vector (depending on u) for all $g \in \partial f(\bar{x} + u \oplus v(u))$. Since from (15) the particular $g = g_0 := \nabla L_{\mathcal{U}}(u; 0) \oplus 0$ is in $\partial f(\bar{x} + u \oplus v(u))$, the left hand side equality in (13) gives the right hand side constant in (iii) as follows:

$$B_{\mathcal{U}}(u)^\top g_0 = B_{\mathcal{U}}(u)^\top \Big(\nabla L_{\mathcal{U}}(u; 0) \oplus 0\Big) = \nabla L_{\mathcal{U}}(u; 0) + B_{\mathcal{U}}(u)^\top \bar{V} 0 = \nabla L_{\mathcal{U}}(u; 0) .$$

If (iii) holds, then for any $g := \bar{U} g_{\mathcal{U}} + \bar{V} g_{\mathcal{V}} \in \partial f(\bar{x} + u \oplus v(u))$ the left hand side equality in (13) yields $\nabla L_{\mathcal{U}}(u; 0) = B_{\mathcal{U}}(u)^\top g = g_{\mathcal{U}} + B_{\mathcal{U}}(u)^\top \bar{V} g_{\mathcal{V}}$ and, hence,

$$\begin{array}{rcl} g - \bar{U} \nabla L_{\mathcal{U}}(u; 0) & = & \bar{U}\Big(g_{\mathcal{U}} - \nabla L_{\mathcal{U}}(u; 0)\Big) + \bar{V} g_{\mathcal{V}} \\ & = & -\bar{U} B_{\mathcal{U}}(u)^\top \bar{V} g_{\mathcal{V}} + \bar{V} g_{\mathcal{V}} \\ & = & \Big(\bar{V} - \bar{U} B_{\mathcal{U}}(u)^\top \bar{V}\Big)\, g_{\mathcal{V}} \\ & = & B_{\mathcal{V}}(u) g_{\mathcal{V}} , \end{array}$$

by (11). Thus, each vector of the form

$$g - \bar{U} \nabla L_{\mathcal{U}}(u; 0) \quad \text{with } g \in \partial f(\bar{x} + u \oplus v(u)) \tag{16}$$

is a linear combination of the columns of $B_{\mathcal{V}}(u)$. From (15), $g_0 := \bar{U} \nabla L_{\mathcal{U}}(u; 0) \in \partial f(\bar{x} + u \oplus v(u))$, so, from (14), all vectors of form (16) are in $\mathcal{V}(u)$ and, since $\mathcal{V}(u)$ is a linear combination of all such vectors, $B_{\mathcal{V}}(u)$ is a basis for $\mathcal{V}(u)$. $\square$

Theorem 2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function with minimizer $\bar{x} \in \mathbb{R}^n$. Suppose that $0 \in \mathrm{ri}\partial f(\bar{x})$ and let $\bar{x} + u \oplus v(u)$ be a fast track as described in Definition 1. For all u small enough let*

$$\gamma(u) := B_{\mathcal{U}}(u)[B_{\mathcal{U}}(u)^\top B_{\mathcal{U}}(u)]^{-1} \nabla L_{\mathcal{U}}(u; 0) . \tag{17}$$

Then the following hold:

- *(i) there exists a positive constant ρ such that for all $z \in \mathbb{R}^{dim\,\mathcal{V}}$ with $\|z\| \leq \rho$ the vector $\gamma(u) + B_{\mathcal{V}}(u) z$ belongs to $\partial f(\bar{x} + u \oplus v(u))$;*

and if, in addition, any one of the three statements in Theorem 1 holds, then

- *(ii) $\gamma(u) + B_{\mathcal{V}}(u) z \in \mathrm{ri}\partial f(\bar{x} + u \oplus v(u))$ for all z with $\|z\| \leq \frac{1}{2}\rho$; and*
- *(iii) each fast track point $\bar{x} + u \oplus v(u)$ is a unique minimizer of f on the corresponding affine set $\bar{x} + u \oplus v(u) + \mathcal{V}(u)$.*

Proof. To abbreviate notation, we again drop symbols such as (u) and $(u; 0)$ so that, for example, $\nabla L_{\mathcal{U}}(u; 0)$ becomes $\nabla L_{\mathcal{U}}$. By Proposition 1(iii), $B_{\mathcal{V}}^\top B_{\mathcal{V}}$ is invertible. Hence, $\ell := -[B_{\mathcal{V}}^\top B_{\mathcal{V}}]^{-1} B_{\mathcal{V}}^\top \bar{U} \nabla L_{\mathcal{U}}$ is well defined, with $\ell \to 0$ as $u \to 0$, by (9) and Proposition 1(i). Given $z \in \mathbb{R}^{dim\,\mathcal{V}}$ and letting $w := z + \ell$, we claim that

$$(\nabla L_{\mathcal{U}}(u; 0) - B_{\mathcal{U}}(u)^\top \bar{V} w) \oplus w = \gamma(u) + B_{\mathcal{V}}(u) z \tag{18}$$

for all u small enough. To prove this claim, we consider these two vectors with respect to $[B_{\mathcal{V}}|B_{\mathcal{U}}]$, which is a basis for $\mathbb{R}^n$, by Proposition 1(iii). Let G be the left hand side vector in (18). Taking $g = G$ in (13) gives

$$B_{\mathcal{U}}^\top G = G_{\mathcal{U}} + B_{\mathcal{U}}^\top \bar{V} G_{\mathcal{V}} = \nabla L_{\mathcal{U}} - B_{\mathcal{U}}^\top \bar{V} w + B_{\mathcal{U}}^\top \bar{V} w = \nabla L_{\mathcal{U}}$$

and

$$\begin{array}{rcll} B_{\mathcal{V}}^\top G & = & B_{\mathcal{V}}^\top \bar{U} G_{\mathcal{U}} + \bar{V}^\top \bar{V} G_{\mathcal{V}} & \\ & = & B_{\mathcal{V}}^\top \bar{U}\left(\nabla L_{\mathcal{U}} - B_{\mathcal{U}}^\top \bar{V} w\right) + \bar{V}^\top \bar{V} w & \\ & = & -B_{\mathcal{V}}^\top B_{\mathcal{V}} \ell - B_{\mathcal{V}}^\top \bar{U} B_{\mathcal{U}}^\top \bar{V} w + \bar{V}^\top \bar{V} w & \text{[by definition of } \ell] \\ & = & -B_{\mathcal{V}}^\top B_{\mathcal{V}} \ell - B_{\mathcal{V}}^\top \left(\bar{V} - B_{\mathcal{V}}\right) w + \bar{V}^\top \bar{V} w & \text{[by (11)]} \\ & = & -B_{\mathcal{V}}^\top B_{\mathcal{V}} \ell - \bar{V}^\top \bar{V} w + B_{\mathcal{V}}^\top B_{\mathcal{V}} w + \bar{V}^\top \bar{V} w & \text{[by the second equality in (12)]} \\ & = & B_{\mathcal{V}}^\top B_{\mathcal{V}} z\,. & \text{[by definition of } w] \end{array}$$

Altogether, we have that

$$B_{\mathcal{U}}^\top G = \nabla L_{\mathcal{U}} \qquad \text{and} \qquad B_{\mathcal{V}}^\top G = B_{\mathcal{V}}^\top B_{\mathcal{V}} z\,.$$

Our claim is established, because from (17) and the orthogonality of $B_{\mathcal{U}}$ and $B_{\mathcal{V}}$ given in Proposition 1(iii), we have that

$$B_{\mathcal{U}}^\top (\gamma + B_{\mathcal{V}} z) = \nabla L_{\mathcal{U}} \qquad \text{and} \qquad B_{\mathcal{V}}^\top (\gamma + B_{\mathcal{V}} z) = B_{\mathcal{V}}^\top B_{\mathcal{V}} z\,.$$

[(i)] Since $0 \in \mathrm{ri}\partial f(\bar{x})$, there exists $\rho_0 > 0$ such that for $w \in \mathbb{R}^{dim\,\mathcal{V}}$

$$0 \oplus w \in \mathrm{ri}\partial f(\bar{x}) \quad \text{if } \|w\| \leq \rho_0\,.$$

By Lemma 2(i) and (ii), if $\|w\| \leq \rho_0$, then for all u small enough

$$(\nabla L_{\mathcal{U}}(u;0) - B_{\mathcal{U}}(u)^\top \bar{V} w) \oplus w \in \partial f(\bar{x} + u \oplus v(u))\,. \tag{19}$$

Since $\rho_0 > 0$, there exist small enough $\rho\,, \rho_1 > 0$ such that

$$\rho + \|\ell\| \leq \rho_0 \quad \text{for all } u \in \mathbb{R}^{dim\,\mathcal{U}} \text{ such that } \|u\| \leq \rho_1.$$

Let $z \in \mathbb{R}^{dim\,\mathcal{V}}$ have arbitrary direction and length $\|z\| \in [0, \rho]$ and let $w = z + \ell$. Then $\|w\| \leq \rho_0$ if $\|u\| \leq \rho_1$, so, by (18) and (19), $\gamma + B_{\mathcal{V}} z$ is a subgradient of f at $\bar{x} + u \oplus v(u)$ for all u small enough, and (i) follows.

[(ii)] Since any of the three equivalent statements in Theorem 1 holds, $B_{\mathcal{V}}(u)$ is a basis for $\mathcal{V}(u)$, a subspace parallel to $\partial f(\bar{x} + u \oplus v(u))$ by (14). As a result, the topology giving the relative interior of this subdifferential consists of relative balls of the form $\mathcal{B}_\delta := \left\{B_{\mathcal{V}}(u)t : \text{ for any } t \in \mathbb{R}^{dim\,\mathcal{V}} \text{ with } \|B_{\mathcal{V}}(u)t\| \leq \delta\right\}$ for positive δ. Take δ so small that any corresponding t satisfies $\|t\| \leq \rho/2$. Then, for all z with $\|z\| \leq \rho/2$, $\|z + t\| \leq \rho$ and from item (i), with z replaced by $z + t$, $\gamma + B_{\mathcal{V}}(z + t)$ is a subgradient of f at $\bar{x} + u \oplus v(u)$, i.e., $\gamma + B_{\mathcal{V}} z + \mathcal{B}_\delta \subset \partial f(\bar{x} + u \oplus v(u))$ and the result follows.

[(iii)] For any $x \in \bar{x} + u \oplus v(u) + \mathcal{V}(u)$, $d := x - (\bar{x} + u \oplus v(u)) \in \mathcal{V}(u)$. Since $B_{\mathcal{V}}(u)$ is a basis for $\mathcal{V}(u)$ and, by Proposition 1(iii) $B_{\mathcal{U}}(u)$ and $B_{\mathcal{V}}(u)$ are orthogonal, d and $\gamma(u)$ are orthogonal too. From item (ii), $\gamma(u) \in \mathrm{ri}\partial f(\bar{x} + u \oplus v(u))$, so there exists a positive η sufficiently small such that $\gamma(u) + \eta d \in \partial f(\bar{x} + u \oplus v(u))$. From the convexity of f it follows that

$$\begin{array}{rcl} f(x) & \geq & f(\bar{x} + u \oplus v(u)) + (\gamma(u) + \eta d)^\top d \\ & = & f(\bar{x} + u \oplus v(u)) + \eta \|d\|^2 . \end{array}$$

Thus, unless $x = \bar{x} + u \oplus v(u)$, $f(x) > f(\bar{x} + u \oplus v(u))$, which implies (iii). ☐

In the next section we show how to obtain fast tracks to certain minimizers of functions with *primal-dual gradient (*`pdg`*) structure*, as introduced in [MS00b].

4 A class of functions having fast tracks

To support the usefulness of our fast track definition, we next study a large class of functions for which such trajectories exist. The subdifferential of a function in this class is allowed to have a continuum of extreme points and its shape is reflected in the shape of a certain convex multiplier set Δ. If the underlying structure of such a function is known and the nonlinear system (20) given below can be solved (as is done for the example functions in [MS99], [MS00a]), then the fast track functions $v(u)$ and $\nabla^2 L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}})$ can be computed. However, for practical algorithm purposes we are content just to know that they exist.

4.1 Primal-dual gradient structure and $\mathcal{VU}$-decomposition

Definition 2 *We say that a convex function $f : \mathbb{R}^n \to \mathbb{R}$ has* primal-dual gradient structure *about $\bar{x} \in \mathbb{R}^n$ if the following conditions hold:*

There exists a ball *about $\bar{x}$, $B(\bar{x})$, $m_1 + 1 + m_2$* 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 $B(\bar{x})$ and a dual multiplier set *$\Delta \subset \mathbb{R}^{m_1+1+m_2}$ such that*

(i) $\bar{x} \in \mathcal{P} := \{x \in B(\bar{x}) : \varphi_\ell(x) = 0 \, for \, \ell = 1, \ldots, m_2\}$ and $f_i(\bar{x}) = f(\bar{x})$ for $i = 0, 1, \ldots, m_1$;

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

$$f(x) = \max\{f_i(x) : i = 0, 1, \ldots, m_1\} ;$$

(iii) 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 unit simplex in $\mathbb{R}^{m_1+1}$ given by

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

(b) *for each* $i = 0, 1, \ldots, m_1$ $\mathbf{1}_{i+1} \in \Delta$, *where* $\mathbf{1}_j$ *is the* j^{th} *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* $i \in \{1, 2, \ldots, m_2\} \backslash \{\ell\}$;

(iv) *for each* $x \in \mathcal{P}$, $g \in \partial f(x)$ *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*

complementary slackness: $\alpha_i = 0$ *if* $f_i(x) < f(x)$ *and* $i \leq m_1$,

and

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

□

The class of pdg-structured functions appears to be quite large and includes maximum eigenvalue functions as well as other convex functions such as some that are pointwise maxima of finite or infinite collections of smooth functions, [MS99], [MS00a].

We say that a pdg-structured function f satisfies *strong transversality* at $\bar{x}$ when the $n \times (m_1 + m_2)$ matrix

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

has full column rank. In this case, this $\bar{V}$ is a basis matrix for $\mathcal{V}$, see [MS00b, Lemma 4.4] with $K = I := \{0, 1, \ldots, m_1 + m_2\}$, the set of all primal function indices.

4.2 Fast tracks for pdg-structured functions

We now employ some results from [MS00b, §§5,6] to show how to obtain a fast track satisfying $\mathcal{VU}$-decomposition "differentiability" for a pdg-structured function f that satisfies strong transversality at a minimizer.

Relative to the function $v_I(\cdot)$ defined next, let $\mathcal{V}_I(u) := \mathcal{V}(u)$ when $v(u) = v_I(u)$ in (14).

Theorem 3 *Let f be a* pdg *-structured function satisfying strong transversality at a minimizer $\bar{x}$. Then, for all u small enough, the nonlinear system with variables u and v*

$$\left\{ \begin{array}{lcll} f_i(\bar{x} + \bar{U}u + \bar{V}v) \quad -f_0(\bar{x} + \bar{U}u + \bar{V}v) & = & 0\,, & i = 1, \ldots, m_1 \\ \varphi_{i-m_1}(\bar{x} + \bar{U}u + \bar{V}v) & = & 0\,, & i = m_1 + 1, \ldots, m_1 + m_2 \end{array} \right. \tag{20}$$

has a unique C^2-function solution $v = v_I(u)$ such that the corresponding matrix function

$$V(u) := \left[\left\{ \nabla f_i(\bar{x}+u \oplus v_I(u)) - \nabla f_0(\bar{x}+u \oplus v_I(u)) \right\}_{i=1}^{m_1} \cup \left\{ \nabla \varphi_{i-m_1}(\bar{x}+u \oplus v_I(u)) \right\}_{i=m_1+1}^{m_1+m_2} \right] \tag{21}$$

is a C^1-basis for $\mathcal{V}_I(u)$ and such that $\bar{V} v_I(u) \in W(u; \bar{g}_{\mathcal{V}})$ for all $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$.

If, in addition, $0 \in \mathrm{ri}\partial f(\bar{x})$, then the associated trajectory $\bar{x}+\bar{U}u+\bar{V}v_I(u)$ is a fast track such that all three statements of Theorem 1 and all three results of Theorem 2 hold with $v(u) = v_I(u)$, $\mathcal{V}(u) = \mathcal{V}_I(u)$ and $B_{\mathcal{U}}(u)$ and $B_{\mathcal{V}}(u)$ as defined in (11) *with $Jv(u) = Jv_I(u)$.*

Proof. The following cited Definition 4.2, Lemma 4.4 and Theorems 5.1, 5.2, 6.1 and 6.3 are all from [MS00b]. The assumption of strong transversality implies that the full index set I is a basic index set, as defined in Definition 4.2. Therefore, from Theorem 5.1, written with $K = I$, there exists a unique function $v_I(\cdot)$ solving (20) which satisfies $v_I(0) = 0$ and has a continuous Jacobian

$$Jv_I(u) = -(V(u)^\top \bar{V})^{-1} V(u)^\top \bar{U} \quad \text{satisfying} \quad Jv_I(0) = 0$$

where $V(u)$ is the matrix defined in (21) with linearly independent columns for u sufficiently small. As a result, the trajectory $\bar{x} + u \oplus v_I(u)$ has a continuous Jacobian $\bar{U} + \bar{V} Jv_I(u)$ (denoted by $Jx(u)$ in [MS00b]). Furthermore, $V(u)$ is a C^1-function, because the primal functions are C^2 and $\bar{x} + u \oplus v_I(u)$ is C^1. Thus, $Jv_I(u)$ is C^1, so $v_I(u)$ is C^2. Moreover, from Lemma 4.4 with $\bar{x}$ replaced by $\bar{x} + u \oplus v_I(u)$ and $\mathcal{V}$ replaced by $\mathcal{V}_I(u)$, $V(u)$ is a basis for $\mathcal{V}_I(u)$ satisfying $V(0) = \bar{V}$.

Strong transversality also implies that I is a dual feasible basic index set for each $g \in \partial f(\bar{x})$, as defined in Definition 4.2 (ii). Thus, for each $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$, the combination of Theorem 6.1 with $K = I$ and Theorem 6.3 with $\bar{K} = I$ gives $\bar{V} v_I(u) \in W_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}})$, and, hence, $v_I(u)$ satisfies condition (i) of the fast track definition.
The same theorems yield the following expression for the Hessian of each $L_{\mathcal{U}}$:

$$\nabla^2 L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) = \left(\bar{U} + \bar{V} Jv_I(u) \right)^\top H(u; \bar{g}_{\mathcal{V}}) \left(\bar{U} + \bar{V} Jv_I(u) \right),$$

where the $n \times n$ matrix function

$$H(u; \bar{g}_{\mathcal{V}}) := \sum_{i=0}^{m_1} \alpha_i(u) \nabla^2 f_i(\bar{x} + u \oplus v_I(u)) + \sum_{i=m_1+1}^{m_1+m_2} \alpha_i(u) \nabla^2 \varphi_{i-m_1}(\bar{x} + u \oplus v_I(u))$$

depends on $\bar{g}_{\mathcal{V}}$ via the C^1 multiplier vector function $(\alpha_0(u), \ldots, \alpha_{m_1+m_2}(u))$ from Theorem 5.2. Finally, since the primal functions and v_I are C^2 functions, $\nabla^2 L_{\mathcal{U}}(\cdot; \bar{g}_{\mathcal{V}})$ is continuous for each $\bar{g} \in \mathrm{ri}\partial f(\bar{x})$. In particular, when $0 \in \mathrm{ri}\partial f(\bar{x})$, $\nabla^2 L_{\mathcal{U}}(\cdot; 0)$ exists and is continuous. Thus, the above results and the assumption that $0 \in \mathrm{ri}\partial f(\bar{x})$ imply that $\bar{x} + u \oplus v_I(u)$ is a fast track.

To show that the statements in Theorem 1 and all three results of Theorem 2 here hold, we next show that $\bar{U} + \bar{V} Jv_I(u)$ is a basis matrix for $\mathcal{V}_I(u)^\perp$. Since $Jv_I(u) \to 0$

as $u \to 0$, the columns of $\bar{U}+\bar{V}Jv_I(u)$ are linearly independent for all u small enough. Furthermore,

$$\begin{aligned}\left[\bar{U}^\top + Jv_I(u)^\top \bar{V}^\top\right] V(u) &= \bar{U}^\top V(u) - \bar{U}^\top V(u)(\bar{V}^\top V(u))^{-1}\bar{V}^\top V(u) \\ &= 0\,,\end{aligned}$$

so $\bar{U}+\bar{V}Jv_I(u)$ is a basis matrix for $\mathcal{V}_I(u)^\perp$. □

Remark 2 The above theorem shows that $B_\mathcal{V}(u)$ and $V(u)$ are both basis matrices for $\mathcal{V}_I(u)$ which converge to $\bar{V}$, a basis matrix for $\mathcal{V} = \mathcal{V}_I(0)$. Theorem 3 via Theorem 2(*iii*) also shows that $\bar{x} + u \oplus v_I(u)$ is the unique minimizer of f on the affine set $\bar{x} + u \oplus v_I(u) + \mathcal{V}_I(u)$. In addition, combined with the definition of $W(u; \bar{g}_\mathcal{V})$ in (6), it implies that this fast track point minimizes f on another affine set, namely $\bar{x}+u\oplus v_I(u)+\mathcal{V}$. Thus, all three matrices $B \in \{B_\mathcal{V}(u), V(u), \bar{V}\}$ satisfy the property that there exists a $g \in \partial f(\bar{x}+u\oplus v_I(u))$ (depending on B) such that $B^\top g = 0$. Among these three matrices, $V(u)$ is the one of practical importance, because it is the type of matrix that is approximated by bundle methods, see [Mif96, §5], for example. □

5 Proximal points and fast tracks

We are now in a position to start our development to show that fast tracks attract proximal points. In the following theorem we find implicit functions $u(x)$ and $\bar{g}_\mathcal{V}(x)$ such that $u(x) \to 0$ and $\bar{g}_\mathcal{V}(x) \to 0$ as $x \to \bar{x}$ and we make use of the $n \times n$ matrix defined by

$$\mathcal{J}(x) := B_\mathcal{U}(u(x)) \left[\frac{1}{\mu}\nabla^2 L_\mathcal{U}(u(x); \bar{g}_\mathcal{V}(x)) + B_\mathcal{U}(u(x))^\top B_\mathcal{U}(u(x))\right]^{-1} B_\mathcal{U}(u(x))^\top\,. \quad (22)$$

The inverse in this expression exists because $B_\mathcal{U}(u)^\top B_\mathcal{U}(u) \to \bar{U}^\top\bar{U} = I_\mathcal{U}$ as $u \to 0$ and, by convexity, $L_\mathcal{U}$ has positive semidefinite Hessians. Also, because of the first equality in (12), $\mathcal{J}(x)$ satisfies the following property:

$$B_\mathcal{U}(u(x))\bar{U}^\top \mathcal{J}(x) = \mathcal{J}(x)\,. \quad (23)$$

Theorem 4 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function with minimizer $\bar{x} \in \mathbb{R}^n$. Suppose that $0 \in \mathrm{ri}\partial f(\bar{x})$ and let $\bar{x} + u \oplus v(u)$ be a fast track as described in Definition 1. Given a positive parameter μ, for all x close enough to $\bar{x}$ the nonlinear system with variables $(u, \bar{g}_\mathcal{V})$ and x given by*

$$\frac{1}{\mu}\Big(\nabla L_\mathcal{U}(u; \bar{g}_\mathcal{V}) \oplus \bar{g}_\mathcal{V}\Big) - \Big(x - (\bar{x} + u \oplus v(u))\Big) = 0$$

has a unique C^1-function solution $(u, \bar{g}_\mathcal{V}) = \big(u(x), \bar{g}_\mathcal{V}(x)\big)$ such that $0 \oplus \bar{g}_\mathcal{V}(x) \in \mathrm{ri}\partial f(\bar{x})$, $\big(u(\bar{x}), \bar{g}_\mathcal{V}(\bar{x})\big) = (0, 0)$, and

$$\begin{bmatrix} Ju(x) \\ J\bar{g}_\mathcal{V}(x)\end{bmatrix} = \begin{bmatrix} \bar{U}^\top \mathcal{J}(x) \\ \mu[\bar{V}^\top\bar{V}]^{-1}\bar{V}^\top\big(I - \mathcal{J}(x)\big)\end{bmatrix}. \quad (24)$$

Proof. Note that (8) and (9) give, respectively, $v(0) = 0$ and $\nabla L_{\mathcal{U}}(0;0) = 0$. Therefore, the triple $(u, \bar{g}_{\mathcal{V}}, x) = (0, 0, \bar{x})$ satisfies the system equation. Note also that, by Definition 1(i) and Lemma 2(i), the system function is C^1 on a ball about $(0, 0, \bar{x})$. So, now we are in a position to apply the Implicit Function Theorem to this system, re-written in its $\mathcal{U}$ and $\mathcal{V}$ components, as follows:

$$\begin{array}{rcccccl} \mathbb{R}^{dim\,\mathcal{U}} \ni S_{\mathcal{U}}(u, \bar{g}_{\mathcal{V}}, x) & := & \frac{1}{\mu}\nabla L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) & + & u & - & (x-\bar{x})_{\mathcal{U}} = 0 \\ \mathbb{R}^{dim\,\mathcal{V}} \ni S_{\mathcal{V}}(u, \bar{g}_{\mathcal{V}}, x) & := & \frac{1}{\mu}\bar{g}_{\mathcal{V}} & + & v(u) & - & (x-\bar{x})_{\mathcal{V}} = 0\,. \end{array}$$

The corresponding Jacobian with respect to $(u, \bar{g}_{\mathcal{V}})$ is $\mathcal{S} := \begin{bmatrix} \mathcal{S}_{11} & \mathcal{S}_{12} \\ \mathcal{S}_{21} & \mathcal{S}_{22} \end{bmatrix}$, where

$$\begin{array}{rcll} \mathcal{S}_{11} := \dfrac{\partial S_{\mathcal{U}}}{\partial u} & = & \dfrac{1}{\mu}\nabla^2 L_{\mathcal{U}}(u; \bar{g}_{\mathcal{V}}) + I_{\mathcal{U}}\,, & \\ \mathcal{S}_{12} := \dfrac{\partial S_{\mathcal{U}}}{\partial \bar{g}_{\mathcal{V}}} & = & -\dfrac{1}{\mu}B_{\mathcal{U}}(u)^{\top}\bar{V}\,, & \text{[by Lemma 2}(i)] \\ \mathcal{S}_{21} := \dfrac{\partial S_{\mathcal{V}}}{\partial u} & = & Jv(u)\,,\ \text{and} & \\ \mathcal{S}_{22} := \dfrac{\partial S_{\mathcal{V}}}{\partial \bar{g}_{\mathcal{V}}} & = & \dfrac{1}{\mu}I_{\mathcal{V}}\,. & \end{array}$$

In particular, from (8) and the definition of $B_{\mathcal{U}}(u)$ and the fact that $\bar{U}^{\top}\bar{V} = 0$, this Jacobian at $(u, \bar{g}_{\mathcal{V}}, x) = (0, 0, \bar{x})$ equals

$$\begin{bmatrix} \frac{1}{\mu}\nabla^2 L_{\mathcal{U}}(0;0) + I_{\mathcal{U}} & 0 \\ 0 & \frac{1}{\mu}I_{\mathcal{V}} \end{bmatrix},$$

which, by the convexity of $L_{\mathcal{U}}$, is an invertible matrix. Thus, the Implicit Function Theorem gives the existence of the desired functions and, taking into account the definitions of $S_{\mathcal{U}}$ and $S_{\mathcal{V}}$, the following expressions for their respective Jacobians:

$$\begin{bmatrix} Ju(x) \\ J\bar{g}_{\mathcal{V}}(x) \end{bmatrix} = -\mathcal{S}^{-1}\begin{bmatrix} \dfrac{\partial S_{\mathcal{U}}}{\partial x} \\ \dfrac{\partial S_{\mathcal{V}}}{\partial x} \end{bmatrix} = -\mathcal{S}^{-1}\begin{bmatrix} -\bar{U}^{\top} \\ -[\bar{V}^{\top}\bar{V}]^{-1}\bar{V}^{\top} \end{bmatrix}, \tag{25}$$

where $\mathcal{S}$ is evaluated at $(u; \bar{g}_{\mathcal{V}}) = (u(x); \bar{g}_{\mathcal{V}}(x))$. The following expression for $\mathcal{S}^{-1}$ can be verified by multiplication, or by using (A.8) and (A.9) in [Ber95, p. 543], together with (22), (23), and the fact that $\bar{V}Jv(u) = B_{\mathcal{U}}(u) - \bar{U}$ by (11):

$$\mathcal{S}^{-1} = \begin{bmatrix} \bar{U}^\top \mathcal{J}(x)\bar{U} & \bar{U}^\top \mathcal{J}(x)\bar{V} \\ -\mu[\bar{V}^\top \bar{V}]^{-1}\bar{V}^\top \mathcal{J}(x)\bar{U} & \mu[\bar{V}^\top \bar{V}]^{-1}\bar{V}^\top \big(I - \mathcal{J}(x)\big)\bar{V} \end{bmatrix}.$$

Substituting this expression into (25) and writing the identity in $\mathbb{R}^n$ as $I = \bar{U}\bar{U}^\top + \bar{V}[\bar{V}^\top\bar{V}]^{-1}\bar{V}^\top$ yields (24).
Finally, to see that $0 \oplus \bar{g}_{\mathcal{V}}(x) \in \mathrm{ri}\partial f(\bar{x})$ when x is close enough to $\bar{x}$, just recall that $\bar{g}_{\mathcal{V}}(\bar{x}) = 0 \in \mathcal{V}$ and $0 \in \mathrm{ri}\partial f(\bar{x})$. ☐

We mention in passing that the above implicit function result still holds if $1/\mu$ in the statement of Theorem 4 is replaced by the inverse of an arbitrary positive definite $n \times n$ matrix.

Relation with proximal points Given a positive scalar parameter μ, the proximal point function, corresponding to f and μ, at a given point $x \in \mathbb{R}^n$ is defined by

$$p(x) := \arg\min_{y \in \mathbb{R}^n} \{f(y) + \frac{1}{2}\mu\|y - x\|^2\}.$$

Observe that the optimality condition giving the proximal point $y = p(x)$ is the following:

$$\text{there exists} \quad g \in \partial f(p(x)) \quad \text{such that} \quad g = \mu\big(x - p(x)\big). \tag{26}$$

Moreover, letting $F(x) := f(p(x)) + \frac{1}{2}\mu\|p(x) - x\|^2$ be the Moreau-Yosida regularization of f at x, it can be shown that

$$g(x) := \mu(x - p(x)) = \nabla F(x) \tag{27}$$

is a Lipschitz continuous function of x, [Mor65], [Yos64], [Roc76].

We now state our main result, and show that for a point near a minimizer of f its corresponding proximal point is on the fast track.

Theorem 5 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function with minimizer $\bar{x} \in \mathbb{R}^n$. Suppose that $0 \in \mathrm{ri}\partial f(\bar{x})$ and let $\bar{x} + u \oplus v(u)$ be a fast track as described in Definition 1. Given a positive parameter μ, for all x close enough to $\bar{x}$, there exists a $u(x)$ from Theorem 4 such that $p(x) = \bar{x} + u(x) \oplus v(u(x))$, $g(x) = \mu(x - p(x)) \in \partial f(p(x))$ and $Jp(x) = \mathcal{J}(x)$, as defined in (22).*
In particular, $p(\bar{x}) = \bar{x}$, $g(\bar{x}) = 0$ and $Jp(\bar{x}) = \bar{U}\left[\frac{1}{\mu}\nabla^2 L_{\mathcal{U}}(0;0) + I_{\mathcal{U}}\right]^{-1}\bar{U}^\top$.
If, in addition, any one of the three statements in Theorem 1 holds, then

$$B_{\mathcal{U}}(u(x))^\top g(x) = \nabla L_{\mathcal{U}}(u(x);0) \quad \textit{and} \quad g(x) \in \mathrm{ri}\partial f(p(x)).$$

Proof. Given x, the optimality condition (26) characterizes $p(x)$ as the unique $y \in \mathbb{R}^n$ for which there exists $g \in \partial f(y)$ such that $g = \mu(x-y)$. For x close enough to $\bar{x}$, Theorem 4 shows the existence of $(u(x), \bar{g}_{\mathcal{V}}(x))$ such that

$$\frac{1}{\mu}\Big(\nabla L_{\mathcal{U}}(u(x); \bar{g}_{\mathcal{V}}(x)) \oplus \bar{g}_{\mathcal{V}}(x)\Big) - \Big(x - (\bar{x} + u(x) \oplus v(u(x))\Big) = 0\,,$$

and $0 \oplus \bar{g}_{\mathcal{V}}(x) \in \mathrm{ri}\partial f(\bar{x})$. When x is close to $\bar{x}$, $u(x)$ is close to 0, so Lemma 2 (ii) with $\bar{g} := 0 \oplus \bar{g}_{\mathcal{V}}(x)$ implies that the vector

$$g := \nabla L_{\mathcal{U}}(u(x); \bar{g}_{\mathcal{V}}(x)) \oplus \bar{g}_{\mathcal{V}}(x) = \mu\Big(x - \Big(\bar{x} + u(x) \oplus v(u(x))\Big)\Big)$$

is in $\partial f(\bar{x} + u(x) \oplus v(u(x)))$, so $p(x) = \bar{x} + u(x) \oplus v(u(x))$ and $\mu(x - p(x)) \in \partial f(p(x))$. The expression for $Jp(x)$ is obtained by differentiating $p(x) = \bar{x} + \bar{U}u(x) + \bar{V}v(u(x))$ and performing the following steps:

$$\begin{array}{rcll} Jp(x) & = & \bar{U}Ju(x) + \bar{V}Jv(u(x))Ju(x) & \\ & = & \Big[\bar{U} + \bar{V}Jv(u(x))\Big]\bar{U}^{\top}\mathcal{J}(x) & \text{[by (24)]} \\ & = & B_{\mathcal{U}}(u(x))\bar{U}^{\top}\mathcal{J}(x) & \text{[by (11)]} \\ & = & \mathcal{J}(x)\,. & \text{[by (23)]} \end{array}$$

If Theorem 1 applies, then $B_{\mathcal{V}}(u(x))$ is a basis for $\mathcal{V}(u(x))$, $B_{\mathcal{U}}(u(x))$ is a basis for $\mathcal{V}(u(x))^{\perp}$, and, since

$$g(x) \in \partial f(p(x)) = \partial f(\bar{x} + u(x) \oplus v(u(x)))\,,$$

$B_{\mathcal{U}}(u(x))^{\top} g(x) = \nabla L_{\mathcal{U}}(u(x); 0)$, so

$$g(x) = B_{\mathcal{U}}(u(x))[B_{\mathcal{U}}(u(x))^{\top} B_{\mathcal{U}}(u(x))]^{-1}\nabla L_{\mathcal{U}}(u(x); 0) + B_{\mathcal{V}}(u(x))z(x)\,, \qquad (28)$$

where $z(x) \in \mathbb{R}^{dim\,\mathcal{V}}$. Since $g(x) \to 0$ as $x \to \bar{x}$, $z(x) \to 0$ also, because on the right hand side of (28) $B_{\mathcal{U}}(u(x)) \to \bar{U}$, $\nabla L_{\mathcal{U}}(u(x); 0) \to 0$ and $B_{\mathcal{V}}(u(x)) \to \bar{V}$, a matrix with linearly independent columns. So, from Theorem 2(ii) with $u = u(x)$, we conclude that $g(x) \in \mathrm{ri}\partial f(p(x))$, because $\|z(x)\| \le \rho/2$ for x close enough to $\bar{x}$. $\square$

For completeness we include a result concerning the extreme case where $dim\,\mathcal{U} = 0$. A similar result is used in [Roc76, Proposition 8] to prove finite convergence of the proximal point algorithm in the polyhedral function case.

Theorem 6 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function with minimizer $\bar{x} \in \mathbb{R}^n$. Suppose that $dim\,\mathcal{V} = n$ and $0 \in \mathrm{ri}\partial f(\bar{x})$. Then 0 is in the interior of $\partial f(\bar{x})$ and $p(x) = \bar{x}$ for all x close enough to $\bar{x}$.*

Proof. Since $dim\,\mathcal{V} = n$, the relative interior and the interior of $\partial f(\bar{x})$ are the same, so 0 is in the interior of $\partial f(\bar{x})$. This implies that for all x sufficiently close to $\bar{x}$ (depending on μ and the size of $\partial f(\bar{x})$)

$$0 + \mu(x - \bar{x}) \in \partial f(\bar{x})\,,$$

so $g = \mu(x - \bar{x})$ and $p(x) = \bar{x}$ satisfy (26). □

We conclude with a result showing that a Newton step based on the Moreau-Yosida regularization F is equivalent to a proximal step plus a $\mathcal{U}$-Newton step.

Theorem 7 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function with minimizer $\bar{x} \in \mathbb{R}^n$. Suppose that $0 \in \mathrm{ri}\partial f(\bar{x})$ and let $\bar{x} + u \oplus v(u)$ be a fast track as described in Definition 1. Also suppose that $\nabla^2 L_{\mathcal{U}}(0;0)$ is positive definite. Then for all x close enough to $\bar{x}$ the inverse of $\nabla^2 F(x)$ exists and*

$$x - \left[\nabla^2 F(x)\right]^{-1} \nabla F(x) = p(x) - B_{\mathcal{U}}(u(x)) \left[\nabla^2 L_{\mathcal{U}}(u(x); \bar{g}_{\mathcal{V}}(x))\right]^{-1} B_{\mathcal{U}}(u(x))^\top g(x)\,,$$

where $u(x)$ and $\bar{g}_{\mathcal{V}}(x)$ are given by Theorem 4.

Proof. Differentiating (27) and using Theorem 5 gives the relation

$$\nabla^2 F(x) = \mu(I - Jp(x)) = \mu(I - \mathcal{J}(x))\,.$$

Since $\nabla^2 L_{\mathcal{U}}(0;0)$ is positive definite and $(u(x); \bar{g}_{\mathcal{V}}(x)) \to (0;0)$ as $x \to \bar{x}$, the $\mathcal{U}$-Hessian $\nabla^2 L_{\mathcal{U}}(u(x); \bar{g}_{\mathcal{V}}(x))$ has an inverse for x close to $\bar{x}$. Using the definition of $\mathcal{J}(x)$ in (22) it can be seen either by multiplication or by applying (A.7) in [Ber95, p. 543] that

$$[\nabla^2 F(x)]^{-1} \;=\; = \frac{1}{\mu}\left(I + B_{\mathcal{U}}(u(x)) \left[\frac{1}{\mu}\nabla^2 L_{\mathcal{U}}(u(x); \bar{g}_{\mathcal{V}}(x))\right]^{-1} B_{\mathcal{U}}(u(x))^\top\right)\,.$$

Therefore, recalling from (27) that $\nabla F(x) = g(x)$, we have

$$\left[\nabla^2 F(x)\right]^{-1} \nabla F(x) = \frac{1}{\mu} g(x) + B_{\mathcal{U}}(u(x)) \left[\nabla^2 L_{\mathcal{U}}(u(x); \bar{g}_{\mathcal{V}}(x))\right]^{-1} B_{\mathcal{U}}(u(x))^\top g(x)\,.$$

The result then follows from the fact that $g(x) = \mu(x - p(x))$. □

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. By Theorem 2(iii), making a $\mathcal{V}$-step essentially amounts to finding a fast track point. We showed in Theorem 5 that, at least locally, $\mathcal{V}$-steps can be replaced by proximal steps.

Bundle methods can be used to approximate proximal steps and other $\mathcal{VU}$-related quantities. The properties of fast tracks from this paper will be used in a subsequent paper to give conditions on how well quantities such as $p(x)$, $B_{\mathcal{U}}(u(x))$, and $\nabla^2 L_{\mathcal{U}}(u(x); \bar{g}_{\mathcal{V}}(x))$ need to be approximated in order to make a bundle-based $\mathcal{VU}$-algorithm that is globally and superlinearly convergent.

References

[Aus87] A. Auslender. Numerical methods for nondifferentiable convex optimization. *Mathematical Programming Study*, 30:102–126, 1987.

[Ber95] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1995.

[Fuk84] M. Fukushima. A descent algorithm for nonsmooth convex optimization. *Mathematical Programming*, 30:163–175, 1984.

[HUL93] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms*. Number 305-306 in Grund. der math. Wiss. Springer-Verlag, 1993. (two volumes).

[LO96] A.S. Lewis and M.L. Overton. Eigenvalue optimization. *Acta Numerica*, pages 149–190, 1996.

[LOS00] C. Lemaréchal, F. Oustry, and C. Sagastizábal. The $\mathcal{U}$-Lagrangian of a convex function. *Trans. Amer. Math. Soc.*, 352(2):711–729, 2000.

[Mar70] B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. *Revue Française d'Informatique et Recherche Opérationnelle*, R-3:154–179, 1970.

[Mif96] R. Mifflin. A quasi-second-order proximal bundle algorithm. *Mathematical Programming*, 73(1):51–72, 1996.

[Mor65] J.J. Moreau. Proximité et dualité dans un espace Hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299, 1965.

[MS99] R. Mifflin and C. Sagastizábal. $\mathcal{VU}$-decomposition derivatives for convex max-functions. In R. Tichatschke and M. Théra, editors, *Ill-posed Variational Problems and Regularization Techniques*, number 477 in Lecture Notes in Economics and Mathematical Systems, pages 167–186. Springer-Verlag Berlin Heidelberg, 1999.

[MS00a] R. Mifflin and C. Sagastizábal. Functions with primal-dual gradient structure and $\mathcal{U}$-Hessians. In G. Di Pillo and F. Giannessi, editors, *Nonlinear Optimization and Related Topics*, number 36 in Applied Optimization, pages 219–233. Kluwer Academic Publishers B.V., 2000.

[MS00b] R. Mifflin and C. Sagastizábal. On $\mathcal{VU}$-theory for functions with primal-dual gradient structure. *SIAM Journal on Optimization*, 11(2):547–571, 2000.

[Ous99] F. Oustry. The $\mathcal{U}$-Lagrangian of the maximum eigenvalue function. *SIAM Journal on Optimization*, 9:526–549, 1999.

[Ous00] F. Oustry. A second-order bundle method to minimize the maximum eigenvalue function. *Mathematical Programming*, 89(1, Ser. A):1–33, 2000.

[OW95] M. L. Overton and R.S. Womersley. Second derivatives for optimizing eigenvalues of symmetric matrices. *Journal of Mathematical Analysis and Applications*, 3:667–718, July 1995.

[Roc76] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14:877–898, 1976.

[SF95] A. Shapiro and M.K.H. Fan. On eigenvalue optimization. *SIAM Journal on Optimization*, 5:552–569, 1995.

[Yos64] K. Yosida. *Functional Analysis*. Springer Verlag, 1964.