MATHEMATICS OF OPERATIONS RESEARCH
Vol. 28, No. 4, November 2003, pp. 677–692
Printed in U.S.A.

ON A CLASS OF NONSMOOTH COMPOSITE FUNCTIONS

ALEXANDER SHAPIRO

We discuss in this paper a class of nonsmooth functions which can be represented, in a neighborhood of a considered point, as a composition of a positively homogeneous convex function and a smooth mapping which maps the considered point into the null vector. We argue that this is a sufficiently rich class of functions and that such functions have various properties useful for purposes of optimization

1. Introduction. There are many instances where nonsmoothness appears naturally in optimization problems. Typically, the involved nonsmoothness is not arbitrary and is structured in some particular way. In the last three decades quite a number of theories were developed to deal with various aspects of nonsmooth problems, both theoretically and computationally.

In this paper we discuss a class of extended real-valued functions $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, which can be represented in an open neighborhood $\mathcal{O} \subset \mathbb{R}^n$ of a given point $\bar{x} \in \mathbb{R}^n$ in the form

$$f(x) := f(\bar{x}) + g(F(x)), \tag{1.1}$$

for some mapping $F\colon \mathcal{O} \to \mathbb{R}^m$ and function $g\colon \mathbb{R}^m \to \overline{\mathbb{R}}$.

DEFINITION 1.1. We say that the function $f(x)$ is $g \circ F$ *decomposable* at $\bar{x}$ if the representation (1.1) holds for all $x \in \mathcal{O}$ with $f(\bar{x})$ being finite, and F and g satisfying the following assumptions: (A1) The mapping $F\colon \mathcal{O} \to \mathbb{R}^m$ is smooth. (A2) $F(\bar{x}) = 0$. (A3) The function $g(\cdot)$ is positively homogeneous, proper, convex, and lower semicontinuous.

By smoothness of F we mean that $F(\cdot)$ is k-times continuously differentiable on $\mathcal{O}$, where k can be one, two, etc., depending on an application. Recall that it is said that the function $g(\cdot)$ is proper if $g(y) > -\infty$ for all $y \in \mathbb{R}^m$ and its domain $\operatorname{dom} g := \{y \in \mathbb{R}^m\colon g(y) < +\infty\}$ is nonempty. It follows from (A3) that $g(0) = 0$.

The class of decomposable functions is closely related to the class of amenable functions, in the sense of Rockafellar and Wets (1998, Definition 10.23), which assumes local representation (1.1) with the function $g(\cdot)$ being proper, convex, and lower semicontinuous and a certain regularity condition being satisfied. If in addition g is piecewise-linear quadratic, then it is said that f is fully amenable. The main difference between the classes of decomposable functions (in the sense of Definition 1.1) and (fully) amenable functions is that the function $g(\cdot)$ in (1.1) is assumed to be positively homogeneous, but not necessarily piecewise linear, and that F maps $\bar{x}$ into the null vector. The above class of decomposable functions may be viewed as being somewhat between the classes of amenable and fully amenable functions. An important example which can be handled in the framework of decomposable, but not fully amenable, functions is the example of eigenvalue functions (see Example 2.3 below).

Received July 26, 2002; revised December 31, 2002.
MSC 2000 subject classification. Primary: 90C30.
OR/MS subject classification. Primary: Programming/nondifferentiable.
Key words. Nonsmooth optimization, directional derivatives, semismooth functions, partially smooth functions, optimality conditions, sensitivity analysis.

0364-765X/03/2804/0677
1526-5471 electronic ISSN, © 2003, INFORMS

REMARK 1.1. In the above definition we treat cases where the function $f(\cdot)$ can be proper *extended* real valued, i.e., it can take value of $+\infty$. The theory will simplify considerably if we assume that the function $f(\cdot)$ is real valued on $\mathcal{O}$, and hence $g(\cdot)$ is real valued on $\mathbb{R}^m$. This will cover all examples of the next section except Example 2.4 of the indicator function. Adding the indicator function of the feasible set allows us to absorb set constraints into the objective function of the considered optimization problem. Alternatively, one can deal with set constraints by adding an exact penalty term. Therefore, we often assume that actually the function $f(\cdot)$ is real valued. Because any real-valued convex function is continuous, in fact even locally Lipschitz continuous, we have that if the function $g(\cdot)$ is real valued, then the assumption of lower semicontinuity in (A3) is superfluous.

The class of $g \circ F$ decomposable functions has a number of useful properties that we study in this paper. In particular, we show that such real-valued functions are semismooth in the sense of Mifflin (1977), have a $\mathcal{V}\mathcal{U}$-structure in the sense of Lemaréchal et al. (1999) and partly smooth in the sense of Lewis (forthcoming). We also demonstrate that this class is sufficiently rich. In particular, all examples given in Lemaréchal et al. (1999), Mifflin and Sagastizábel (2000), and Lewis (2002), except the example of spectral abcissa in Lewis (2002), can be treated in the framework of $g \circ F$ decomposable functions.

Recall that it is said that $f(\cdot)$ is directionally differentiable at a point $x \in \mathcal{O}$ if the limit

$$f'(x, h) := \lim_{t \downarrow 0} \frac{f(x+th) - f(x)}{t}$$

exists for every $h \in \mathbb{R}^n$. If, moreover, the limit

$$f''(x; h, w) := \lim_{t \downarrow 0} \frac{f(x+th+\frac{1}{2}t^2 w) - f(x) - tf'(x, h)}{\frac{1}{2}t^2}$$

exists for every $h, w \in \mathbb{R}^n$, then it is said that $f(\cdot)$ is parabolically second-order directionally differentiable at x. If $f(\cdot)$ is directionally differentiable at x and the directional derivative $f'(x, h)$ is linear in $h \in \mathbb{R}^n$, then it is said that f is (Gâteaux) differentiable at x and $Df(x)h = f'(x, h)$.

We use the following notation and terminology throughout the paper. By

$$\text{epi } g := \{(y, c) \in \mathbb{R}^m \times \mathbb{R}\colon c \geq g(y)\},$$

we denote the epigraph of the function $g(\cdot)$. Let C be a subset of $\mathbb{R}^m$. We denote by $\sigma_C(y)$ the support function of C, i.e., $\sigma_C(y) := \sup_{z \in C} z^T y$, and by $\text{dist}_C(y) := \inf_{z \in C} \|y - z\|$ the distance from $y \in \mathbb{R}^m$ to the set C with respect to a chosen norm $\|\cdot\|$ on $\mathbb{R}^m$. By $i_C(\cdot)$, we denote the indicator function of the set C; that is, $i_C(y) = 0$ if $y \in C$ and $i_C(y) = +\infty$ if $y \notin C$. By $\text{conv}(C)$, we denote the convex hull of C. In case the set C is convex, we denote by $\text{ri}(C)$ the relative interior of C, and by $T_C(x)$ and $N_C(x)$ the tangent and normal cones, respectively, to C at $x \in C$. For a convex cone C we denote by $\text{lin}(C) := C \cap (-C)$ its lineality space. For a linear space $L \subset \mathbb{R}^m$, we denote by

$$L^\perp := \{z \in \mathbb{R}^m\colon z^T y = 0,\ y \in L\}$$

its orthogonal space. For a linear mapping $A\colon \mathbb{R}^n \to \mathbb{R}^m$, we denote by $A^*\colon \mathbb{R}^m \to \mathbb{R}^n$ its adjoint mapping, defined by $y^T A x = (A^* y)^T x$ for any $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$.

2. Examples. In this section we discuss several examples of decomposable functions.

EXAMPLE 2.1. Let $f_i\colon \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, be smooth functions and consider

$$f(\cdot) := \max_{1 \leq i \leq m} f_i(\cdot). \tag{2.1}$$

Consider a point $\bar{x} \in \mathbb{R}^n$ and define

$$I(\bar{x}) := \{i\colon f(\bar{x}) = f_i(\bar{x}),\ i = 1, \dots, m\}.$$

Then for all x in a neighborhood of $\bar{x}$ we have that $f(x) = \max_{i\in I(\bar{x})} f_i(x)$. Therefore, the function $f(\cdot)$ can be represented near $\bar{x}$ in the form (1.1) with the mapping

$$F(\cdot) := (f_{i_1}(\cdot) - f(\bar{x}), \dots, f_{i_k}(\cdot) - f(\bar{x}))\colon \mathbb{R}^n \to \mathbb{R}^k,$$

and $g(y) := \max\{y_1, \dots, y_k\}$, where $i_1, \dots, i_k$ are the elements of the set $I(\bar{x})$. Clearly, $F(\bar{x}) = 0$ and the function $g(\cdot)$ is real-valued convex, positively homogeneous, and piecewise linear. We also have here that $g(\cdot)$ is the support function of the set $C := \operatorname{conv}\{e_1, \dots, e_k\}$, where e_i denotes the ith coordinate vector of $\mathbb{R}^k$.

The above example can be extended to the following more general case.

EXAMPLE 2.2. Let $F\colon \mathbb{R}^n \to \mathbb{R}^m$ be a smooth mapping and $f(\cdot) := \sigma_C(F(\cdot))$, where C is a nonempty convex closed subset of $\mathbb{R}^m$. Consider a point $\bar{x} \in \mathbb{R}^n$. We have that the support function $g(\cdot) := \sigma_C(\cdot)$ is proper, convex, lower semicontinuous, and positively homogeneous. Therefore, if $F(\bar{x}) = 0$, then Assumptions (A1)–(A3) hold. So, let us consider $\bar{y} := F(\bar{x}) \neq 0$. Suppose further that the set C is *polyhedral*, and hence $\sigma_C(\cdot)$ is piecewise linear, and that $F(\bar{x}) \in \operatorname{dom} g$; i.e., $\sigma_C(\bar{y})$ is finite. Define

$$\overline{C}(\bar{y}) := \left\{z \in C\colon z^T\bar{y} = \sigma_C(\bar{y})\right\}.$$

Because C is polyhedral and $\sigma_C(\bar{y})$ is finite, the set $\overline{C}(\bar{y})$ is nonempty and forms a face of the set C, and moreover

$$\sigma_C(y) = \sup_{z\in\overline{C}(\bar{y})} z^T y = \sigma_{\overline{C}(\bar{y})}(\bar{y}) + \sigma_{\overline{C}(\bar{y})}(y - \bar{y})$$

for all y sufficiently close to $\bar{y}$. Therefore, for all x in a neighborhood of $\bar{x}$ we have that

$$f(x) = f(\bar{x}) + \sigma_{\overline{C}(\bar{y})}(F(x) - F(\bar{x})). \tag{2.2}$$

That is, $f(\cdot)$ is representable near $\bar{x}$ in the form (1.1) with $g(\cdot) := \sigma_{\overline{C}(\bar{y})}(\cdot)$, and clearly the mapping $x \mapsto F(x) - F(\bar{x})$ maps $\bar{x}$ into $0 \in \mathbb{R}^m$.

In particular, any norm $\|\cdot\|$ on $\mathbb{R}^m$ is the support function of the unit ball with respect to its dual norm. Therefore, the above construction can be applied to functions of the form $f(\cdot) := \|F(\cdot)\|$ at a point $\bar{x}$ such that $F(\bar{x}) = 0$. If $F(\bar{x}) \neq 0$, then the above construction can be applied to any polyhedral norm, e.g., l_1 or l_∞ norms. If $g(\cdot) := \|\cdot\|$ is the l_p norm with $p \in (1, \infty)$, then it is smooth in a neighborhood of any nonnull point, and hence, again, the function $f(\cdot) := \|F(\cdot)\|$ is $g \circ F$ decomposable at any point.

The following example is related to the eigenvalue optimization and is more sophisticated.

EXAMPLE 2.3. Consider the linear space $\mathcal{S}^p$ of $p \times p$ symmetric matrices. For a matrix $X \in \mathcal{S}^p$ denote by $\lambda_1(X) \geq \cdots \geq \lambda_p(X)$ its eigenvalues arranged in the decreasing order. We say that $\lambda_k(X)$ is a *leading* eigenvalue of X, of multiplicity r, if $\lambda_{k-1}(X) > \lambda_k(X) = \cdots = \lambda_{k+r-1}(X) > \lambda_{k+r}(X)$. By the definition of the largest eigenvalue, $\lambda_1(X)$ is a leading eigenvalue of X. Let $\overline{X} \in \mathcal{S}^p$ and $\lambda_k(\overline{X})$ be a leading eigenvalue of $\overline{X}$ of multiplicity r, for some $k, r \in \{1, \dots, p\}$. Then it is possible to construct a mapping $\Xi\colon \mathcal{S}^p \to \mathcal{S}^r$ with the following properties (a detail construction of this mapping is given in Shapiro and Fan 1995 and Bonnans and Shapiro 2000, Example 3.98, p. 211).

(i) The mapping $\Xi(\cdot)$ is analytic (and hence is infinitely differentiable) in a neighborhood of $\overline{X}$,

(ii) $\Xi(\overline{X}) = \alpha I_r$, where $\alpha := \lambda_k(\overline{X})$ and I_r is the $r \times r$ identity matrix;

(iii) $D\Xi(\overline{X})\colon \mathcal{S}^p \to \mathcal{S}^r$ is onto;

(iv) For all $X \in \mathcal{S}^p$ in a neighborhood of $\overline{X}$,

$$\lambda_{k+j-1}(X) = \nu_j(\Xi(X)), \qquad j = 1, \ldots, r,$$

where $\nu_1(Y) \geq \cdots \geq \nu_r(Y)$ denote the eigenvalues of matrix $Y \in \mathcal{S}^r$.

Consider the function $f(X) := \lambda_k(X) + \cdots + \lambda_{k+l-1}(X)$ for some $l \in \{1, \ldots, r\}$. We then have that for all $X \in \mathcal{S}^p$ in a neighborhood of $\overline{X}$, the function $f(X)$ can be represented in the form

$$f(X) = f(\overline{X}) + g(F(X)), \tag{2.3}$$

where $F(\cdot) := \Xi(\cdot) - \alpha I_r$ and $g(\cdot) := \nu_1(\cdot) + \cdots + \nu_l(\cdot)$. The function g: $\mathcal{S}^r \to \mathbb{R}$ is convex and positively homogeneous, and by (ii) the mapping $X \mapsto F(X)$ maps $\overline{X}$ into the null matrix of $\mathcal{S}^r$. Consequently, the function $f(X)$ is $g \circ F$ decomposable at $\overline{X}$. In particular, the largest eigenvalue function $f(\cdot) := \lambda_1(\cdot)$ is decomposable at any $X \in \mathcal{S}^p$.

Let us make the following observations.

Remark 2.1. Let $f(x)$ be a $g \circ F$ decomposable at $\bar{x}$ function, and H: $\mathbb{R}^k \to \mathbb{R}^n$ be a smooth mapping with $H(\bar{z}) = \bar{x}$. Then the composite function $\phi(z) := f(H(z))$ can be written as $\phi(z) = \phi(\bar{z}) + g(Q(z))$, where $Q(\cdot) := F(H(\cdot))$, and hence is also representable in the form (1.1) near $\bar{z}$. Moreover, $Q(\bar{z}) = F(\bar{x}) = 0$, and hence the function $\phi(\cdot)$ is $g \circ Q$ decomposable at $\bar{z}$.

Remark 2.2. Let $f(x)$ be a $g \circ F$ decomposable at $\bar{x}$ function and $\phi(x) := f(x) + \psi(x)$, where ψ: $\mathbb{R}^n \to \mathbb{R}$ is a smooth function. We can also represent $\phi(x)$ in the form (1.1) by changing the mapping $x \mapsto F(x)$ to the mapping $x \mapsto (F(x), \psi(x)) \in \mathbb{R}^{m+1}$, and the function $y \mapsto g(y)$ to the function $\mathbb{R}^{m+1} \ni (y, t) \mapsto g(y) + t$. Therefore, the function $\phi(x)$ is also decomposable at $\bar{x}$.

Remark 2.3. Let $f_i(\cdot)$, $i = 1, \ldots, s$, be $g_i \circ F_i$ decomposable at $\bar{x} \in \mathbb{R}^n$ functions. That is, on an open neighborhood $\mathcal{O}$ of $\bar{x} \in \mathbb{R}^n$ each function f_i has Representation (1.1), the functions F_i: $\mathcal{O} \to \mathbb{R}^{m_i}$ are smooth, $F_i(\bar{x}) = 0$, $i = 1, \ldots, s$, and the functions g_i: $\mathbb{R}^{m_i} \to \overline{\mathbb{R}}$ are positively homogeneous, proper, convex, and lower semicontinuous. Consider the function $f(\cdot) := \alpha_1 f_1(\cdot) + \cdots + \alpha_s f_s(\cdot)$, where $\alpha_1, \ldots, \alpha_s$ are some positive constants. We then have that f is $g \circ F$ decomposable at $\bar{x}$ with

$$F(\cdot) := (F_1(\cdot), \ldots, F_s(\cdot)) \colon \mathcal{O} \to \mathbb{R}^{m_1 + \cdots + m_s},$$

and $g(y_1, \ldots, y_s) := \alpha_1 g_1(y_1) + \cdots + \alpha_s g_s(y_s)$. Of course, the construction of Remark 2.2 is a particular case of the above construction.

Consider now the max-function $\phi(\cdot) := \max\{f_1(\cdot), \ldots, f_s(\cdot)\}$. Define

$$I(\bar{x}) := \big\{i\colon \phi(\bar{x}) = f_i(\bar{x}), i = 1, \ldots, s\big\}.$$

Further, suppose that for all $i \in \{1, \ldots, s\} \setminus I(\bar{x})$ the functions $g_i(\cdot)$ are real valued, and hence the corresponding functions $f_i(\cdot)$ are continuous near $\bar{x}$. Then for all x in a neighborhood of $\bar{x}$ we have that $\phi(x) = \max_{i \in I(\bar{x})} f_i(x)$. It then follows that the max-function ϕ is $g \circ F$ decomposable at $\bar{x}$ with

$$F(\cdot) := (F_{i_1}(\cdot), \ldots, F_{i_k}(\cdot)) \colon \mathcal{O} \to \mathbb{R}^{m_{i_1} + \cdots + m_{i_k}},$$

and $g(y_{i_1}, \ldots, y_{i_k}) := \max\{g_{i_1}(y_{i_1}), \ldots, g_{i_k}(y_{i_k})\}$, where $i_1, \ldots, i_k$ are the elements of the set $I(\bar{x})$. Construction of Example 2.1 is a particular case of the above construction.

The class of $g \circ F$ decomposable functions can be viewed as a functional analogue of cone-reducible sets (see Definition 2.1) discussed in Bonnans and Shapiro (2000, §§3.4.4 and 4.6.1). In the remainder of this section we discuss a connection between these two concepts. Let K be a nonempty closed subset of $\mathbb{R}^n$.

DEFINITION 2.1. It is said that the set K is *cone reducible* at a point $\bar{x} \in K$ if there exists a neighborhood $\mathcal{O}$ of $\bar{x}$, a convex closed pointed cone C in a finite dimensional space $\mathbb{R}^m$, and a smooth mapping Φ: $\mathcal{O} \to \mathbb{R}^m$ such that: (i) $\Phi(\bar{x}) = 0$, (ii) $D\Phi(\bar{x})$: $\mathbb{R}^n \to \mathbb{R}^m$ is onto, and (iii) $K \cap \mathcal{O} = \{x \in \mathcal{O}\colon \Phi(x) \in C\}$.

The following example shows that for indicator functions the concepts of $g \circ F$ decomposability and cone reducibility are equivalent except for the additional requirement (ii) in the definition of cone reducibility.

EXAMPLE 2.4. Consider the indicator function $f(\cdot) := i_K(\cdot)$ and a point $\bar{x} \in K$. Suppose that the set K is cone reducible at $\bar{x}$; then $f(x) = i_C(\Phi(x))$ for all $x \in \mathcal{O}$. Because C is a convex closed cone, we have that the indicator function $i_C(\cdot)$ is convex, lower semicontinuous, and positively homogeneous. It follows that the indicator function $f(\cdot)$ is $g \circ \Phi$ decomposable at $\bar{x}$ with $g(\cdot) := i_C(\cdot)$.

Conversely, suppose that the indicator function $f(\cdot) := i_K(\cdot)$ is $g \circ F$ decomposable at $\bar{x}$ and $DF(\bar{x})$: $\mathbb{R}^n \to \mathbb{R}^m$ is onto. It follows from the last assumption that F maps $\mathcal{O}$ onto a neighborhood of $0 \in \mathbb{R}^m$, and hence $g(y)$ can take only two values, zero or $+\infty$. Consequently, $g(\cdot)$ is the indicator function of the set $C := \{y\colon g(y) = 0\}$, which is a convex closed cone. We then obtain that K is cone reducible at $\bar{x}$ to the cone C by the mapping $F(\cdot)$.

EXAMPLE 2.5. Suppose that the set K is cone reducible at $\bar{x} \in K$ and consider the function $f(x) := \operatorname{dist}_C(\Phi(x))$. Since C is a convex cone, we have that the function $g(\cdot) := \operatorname{dist}_C(\cdot)$ is convex, continuous, and positively homogeneous. Consequently, f is $g \circ \Phi$ decomposable at $\bar{x}$. Moreover, the function $f(\cdot)$ is equivalent to the distance function $\operatorname{dist}_K(\cdot)$ in the sense that there exist positive constants α and β such that for all x in a neighborhood of $\bar{x}$ the following inequalities hold:

$$\alpha \operatorname{dist}_C(\Phi(x)) \le \operatorname{dist}_K(x) \quad \text{and} \quad \operatorname{dist}_K(x) \le \beta \operatorname{dist}_C(\Phi(x)). \tag{2.4}$$

Indeed, because the mapping $\Phi(\cdot)$ is locally Lipschitz continuous, there is $\gamma > 0$ such that $\|\Phi(x) - \Phi(y)\| \le \gamma \|x - y\|$ for all x, y near $\bar{x}$. Because by Assumption (ii) of Definition 2.1 Φ maps $\mathcal{O}$ onto a neighborhood of $0 \in \mathbb{R}^m$, it follows that for all x sufficiently close to $\bar{x}$ and $\alpha := \gamma^{-1}$,

$$\operatorname{dist}_K(x) = \inf_{y \in K} \|x - y\| = \inf_{\Phi(y) \in C} \|x - y\| \ge \alpha \inf_{\Phi(y) \in C} \|\Phi(x) - \Phi(y)\| = \alpha \operatorname{dist}_C(\Phi(x)).$$

Under Assumption (ii) of Definition 2.1, the right-hand-side inequality of (2.4) follows by metric regularity.

For example, let $M \subset \mathbb{R}^n$ be a smooth manifold. Then, in a neighborhood $\mathcal{O}$ of a point $\bar{x} \in M$, the manifold M can be represented in the form $M \cap \mathcal{O} = \{x \in \mathcal{O}\colon F(x) = 0\}$, where F: $\mathcal{O} \to \mathbb{R}^m$ is a smooth mapping such that $DF(\bar{x})$: $\mathbb{R}^n \to \mathbb{R}^m$ is onto and $m = n - \dim M$. In this example the set M is reducible to the cone $C := \{0\}$. Clearly, $\operatorname{dist}_C(y) = \|y\|$, and hence $\operatorname{dist}_C(F(x)) = \|F(x)\|$. We have here that $\operatorname{dist}_M(\cdot)$ and $\|F(\cdot)\|$ are equivalent near $\bar{x}$ and can be used as exact penalty functions (see §4).

3. Properties. In this section we study various useful properties of $g \circ F$ decomposable functions. Unless stated otherwise, we assume in this section that the function $f(\cdot)$ is $g \circ F$ decomposable at $\bar{x}$ and *real valued* on $\mathcal{O}$, and hence $g(\cdot)$ is real valued on $\mathbb{R}^m$ (see Remark 1.1).

Because $g(\cdot)$ is convex real valued, it is locally Lipschitz continuous and directionally differentiable. It follows that $f(\cdot)$ is locally Lipschitz continuous on $\mathcal{O}$ and, by the chain rule, that $f(\cdot)$ is directionally differentiable at every point $x \in \mathcal{O}$ with

$$f'(x, h) = g'(F(x), DF(x)h). \tag{3.1}$$

In particular because $g(\cdot)$ is positively homogeneous, we have that $g'(0,\cdot)=g(\cdot)$ and hence $f'(\bar{x},h)=g(DF(\bar{x})h)$. Moreover, because $f(\cdot)$ is locally Lipschitz continuous, we have that, for any $x\in\mathcal{O}$,

$$f(x+h)=f(x)+f'(x,h)+o(\|h\|). \tag{3.2}$$

Also, because $g(\cdot)$ is convex real valued and positively homogeneous, we have that $g(\cdot)$ is the support function of a convex compact set $C\subset\mathbb{R}^m$; i.e., $g(\cdot)=\sigma_C(\cdot)$. The set C coincides with the subdifferential $\partial g(0)$ of $g(\cdot)$, at $x=0$, in the sense of convex analysis. Moreover, we have (e.g., Bonnans and Shapiro 2000, Proposition 2.121) that $g'(y,\cdot)$ is the support function of the set

$$\overline{C}(y):=\{z\in C\colon y^Tz=g(y)\},$$

and hence,

$$\partial g(y)=\overline{C}(y)=\arg\max_{z\in\partial g(0)} y^Tz. \tag{3.3}$$

Although the function $f(\cdot)$ is not necessarily convex, its directional derivative function $\psi(\cdot):=f'(x,\cdot)$ is convex for any $x\in\mathcal{O}$. We define $\partial f(x):=\partial\psi(0)$. In view of (3.2), this definition coincides with the set of regular subgradients, of f at x, in the sense of Rockafellar and Wets (1998, Definition 8.3). We also have here that $\partial f(x)$ is equal to Clarke's generalized gradient, and hence the function f is regular in the sense of Clarke (1983, Theorem 2.3.9).

It follows from the above that

$$\partial f(x)=[DF(x)]^*\overline{C}(F(x)). \tag{3.4}$$

In particular,

$$\partial f(\bar{x})=[DF(\bar{x})]^*\partial g(0). \tag{3.5}$$

Suppose now that the mapping $F(\cdot)$ is *twice* continuously differentiable on $\mathcal{O}$. Consider a path $x(\cdot)\colon\mathbb{R}_+\to\mathbb{R}^n$ of the form

$$x(t):=\bar{x}+th+\tfrac{1}{2}t^2w, \tag{3.6}$$

where $h,w\in\mathbb{R}^n$. Because $F(\bar{x})=0$, by using the second-order Taylor expansion of $F(\cdot)$ at $\bar{x}$ we can write

$$F(x(t))=tDF(\bar{x})h+\tfrac{1}{2}t^2\big[DF(\bar{x})w+D^2F(\bar{x})(h,h)\big]+o(t^2). \tag{3.7}$$

We then have that

$$\begin{aligned}f(x(t))-f(\bar{x})&=g\big(tDF(\bar{x})h+\tfrac{1}{2}t^2[DF(\bar{x})w+D^2F(\bar{x})(h,h)]\big)+o(t^2)\\&=tg\big(DF(\bar{x})h+\tfrac{1}{2}t[DF(\bar{x})w+D^2F(\bar{x})(h,h)]\big)+o(t^2)\\&=tg(DF(\bar{x})h)+\tfrac{1}{2}t^2g'\big(DF(\bar{x})h,DF(\bar{x})w+D^2F(\bar{x})(h,h)\big)+o(t^2).\end{aligned} \tag{3.8}$$

We obtain that $f(\cdot)$ is parabolically second-order directionally differentiable at $\bar{x}$ and

$$f''(\bar{x};h,w)=g'\big(DF(\bar{x})h,DF(\bar{x})w+D^2F(\bar{x})(h,h)\big). \tag{3.9}$$

Moreover, consider a mapping (path) $t\mapsto w_t$, $t\ge 0$, such that $tw_t\to 0$ as $t\downarrow 0$. We then have that the second-order expansion (3.7) with w replaced by w_t still holds. Also, since

$g(\cdot)$ is convex, we have that for any $y, z \in \mathbb{R}^m$ and $t \geq 0$ the following inequality holds $g(y+tz) \geq g(y) + tg'(y, z)$, and hence

$$g\big(DF(\bar{x})h + \tfrac{1}{2}t[DF(\bar{x})w_t + D^2F(\bar{x})(h,h)]\big) \geq g(DF(\bar{x})h) + \tfrac{1}{2}tg'\big(DF(\bar{x})h, DF(\bar{x})w_t + D^2F(\bar{x})(h,h)\big). \tag{3.10}$$

By (3.9) and (3.10) we obtain

$$f(\bar{x} + th + \tfrac{1}{2}t^2 w_t) \geq f(\bar{x}) + t f'(\bar{x}, h) + \tfrac{1}{2}t^2 f''(\bar{x}; h, w_t) + o(t^2). \tag{3.11}$$

Functions $f(\cdot)$ satisfying Condition (3.11) for any path $t \mapsto w_t$ such that $tw_t \to 0$ as $t \downarrow 0$, were called *second-order epiregular*, at $\bar{x}$, in Bonnans and Shapiro (2000, §3.3.4). Second-order epiregularity is a useful property in a second-order analysis of nonsmooth functions. In particular, second-order epiregularity of $f(\cdot)$ at $\bar{x}$ implies that $f(\cdot)$ is *parabolically regular* at $\bar{x}$ in the sense of Rockafellar and Wets (1998; Definition 13.65) (see Bonnans and Shapiro 2000; Proposition 3.103). By the above discussion we have the following results.

PROPOSITION 3.1. *Suppose that the function $f(\cdot)$ is real valued and $g \circ F$ decomposable at $\bar{x}$, with $F(\cdot)$ being twice continuously differentiable. Then Formulas* (3.1) *and* (3.9) *hold, and $f(\cdot)$ is second-order epiregular and parabolically regular at $\bar{x}$.*

Next we study semismooth properties of $f(\cdot)$ in the sense of Mifflin (1977). There are several equivalent ways to define semismoothness. We use the following definition (cf. Qi and Sun 1993). Consider $r(h) := f'(\bar{x}+h, h) - f'(\bar{x}, h)$. It is said that $f(\cdot)$ is *semismooth* at the point $\bar{x}$ if $r(h) = o(\|h\|)$. It is said that $f(\cdot)$ is *strongly semismooth* at $\bar{x}$ if $r(h) = O(\|h\|^2)$.

PROPOSITION 3.2. *Suppose that the function $f(\cdot)$ is real valued and $g \circ F$ decomposable at $\bar{x}$. Then $f(\cdot)$ is semismooth at $\bar{x}$. If $F(\cdot)$ is twice continuously differentiable, then $f(\cdot)$ is strongly semismooth at $\bar{x}$.*

PROOF. Recall that because $f(\cdot)$ is real valued and $g \circ F$ decomposable at $\bar{x}$, it follows that $f(\cdot)$ is locally Lipschitz continuous and directionally differentiable near $\bar{x}$. Also, because $g(\cdot)$ is positively homogeneous, we have that $g'(0, z) = g(z)$ and $g'(z, z) = g(z)$. It follows that $g(\cdot)$ is strongly semismooth at 0. It is known that composition of a (strongly) semismooth function with a continuously (twice continuously) differentiable mapping is (strongly) semismooth (see, e.g., Qi and Sun 1993). This completes the proof. □

Let us show now that, under a regularity condition, $g \circ F$ decomposable functions are partially smooth in the sense of Lewis (2002). Consider $C := \partial g(0)$, i.e., $g(\cdot) = \sigma_C(\cdot)$. Let a be a point in the relative interior of the set C and consider the set $S := C - a$. Then $0 \in \text{ri}(S)$ and $g(y) = \sigma_S(y) + a^T y$. Let us denote by L the linear subspace of $\mathbb{R}^m$ orthogonal to the relative interior of the set S (or, equivalently, to the relative interior of C), and by $L^\perp$ its orthogonal complement (in the present case $L^\perp$ coincides with the linear space generated by $\text{ri}(S)$). Note that $\sigma_S(y) = 0$ for all $y \in L$, and $\sigma_S(y) > 0$ for any $y \in \mathbb{R}^m \setminus L$.

Consider the set

$$\mathcal{M} := \{x \in \mathcal{O} \colon F(x) \in L\}, \tag{3.12}$$

and suppose that the following regularity condition holds:

$$DF(\bar{x})\mathbb{R}^n + L = \mathbb{R}^m. \tag{3.13}$$

The above regularity condition means that F intersects L *transversally* at the point $\bar{x}$. We refer to (3.13) as the *nondegeneracy* condition. We have that $\bar{x} \in \mathcal{M}$ and, under the nondegeneracy Condition (3.13), the set $\mathcal{M}$ forms a smooth manifold in a neighborhood of the point $\bar{x}$ with the tangent space $T_{\mathcal{M}}(x)$, at a point $x \in \mathcal{M}$ sufficiently close to $\bar{x}$, is given by

$$T_{\mathcal{M}}(x) = \{h \in \mathbb{R}^n \colon DF(x)h \in L\} = [DF(x)^* L^{\perp}]^{\perp}. \tag{3.14}$$

Moreover, because $\sigma_S(y) = 0$ for all $y \in L$, we have that $\sigma_S(F(x)) = 0$, and hence $f(x) = a^T F(x)$ for all $x \in \mathcal{M}$. It follows that the function $f(\cdot)$ is smooth on $\mathcal{M}$. This is the "restricted smoothness" property of Lewis (2002).

For points $x \in \mathcal{M}$ sufficiently close to $\bar{x}$, consider $N_{\mathcal{M}}(x) := T_{\mathcal{M}}(x)^{\perp}$, i.e., $N_{\mathcal{M}}(x)$ is the linear subspace of $\mathbb{R}^n$ orthogonal to $\mathcal{M}$ at x. Note that $N_{\mathcal{M}}(x) = [DF(x)]^* L^{\perp}$. We have that the linear space $DF(x)N_{\mathcal{M}}(x)$ has only the null vector in common with the linear space L. Therefore, there exists a constant $\gamma > 0$ such that for all $x \in \mathcal{M}$ sufficiently close to $\bar{x}$ the following holds:

$$\sigma_S(DF(x)h) \geq \gamma \|h\|, \qquad \forall\, h \in N_{\mathcal{M}}(x). \tag{3.15}$$

Moreover, for any $y \in \mathbb{R}^m$ and $z \in L$,

$$\sigma_S(y+z) \leq \sigma_S(y) + \sigma_S(z) = \sigma_S(y),$$

and

$$\sigma_S(y) = \sigma_S(y+z-z) \leq \sigma_S(y+z) + \sigma_S(-z) = \sigma_S(y+z),$$

and hence $\sigma_S(y) = \sigma_S(y+z)$. Because $g'(y, d) = \sigma'_S(y, d) + a^T d$, it follows that for any $y, d \in \mathbb{R}^m$ and $z \in L$, $g'(y, d) = g'(y+z, d)$. In particular, this implies that for any $d \in \mathbb{R}^m$ and $z \in L$, $g'(z, d) = g'(0, d) = g(d)$. Consequently, for any $h \in \mathbb{R}^n$ and $x \in \mathcal{M}$ we have by (3.1) that $f'(x, h) = g(DF(x)h)$, and hence

$$f'(x, h) = \sigma_S(DF(x)h) + [DF(x)^* a]^T h. \tag{3.16}$$

It follows that there exists $\gamma > 0$ such that for all $x \in \mathcal{M}$ sufficiently close to $\bar{x}$ and $h \in N_{\mathcal{M}}(x)$, the following holds:

$$f'(x, h) \geq b^T h + \gamma \|h\|, \qquad \forall h \in N_{\mathcal{M}}(x), \tag{3.17}$$

where $b := DF(\bar{x})^* a$. Inequality (3.17) implies the "normal sharpness" property of Lewis (2002).

Because for any $h \in \mathbb{R}^n$ and $x \in \mathcal{M}$, $f'(x, h) = g(DF(x)h)$, we have by continuity of $DF(\cdot)h$ and $g(\cdot)$ that $f'(\cdot, h)$ is continuous on $\mathcal{M}$. This implies the "regularity" and "subderivative continuity" properties of Lewis (2002) and completes the arguments showing that under the nondegeneracy Condition (3.13), the function f is partially smooth.

The above analysis also shows that under (3.13) the function $f(\cdot)$ has the $\mathcal{V}\mathcal{U}$-structure at $\bar{x}$ in the sense of Lemaréchal et al. (1999).

Consider, for instance, the setting of Example 2.1. Denote $e := (1, \ldots, 1) \in \mathbb{R}^k$. Then the point $a := k^{-1}e$ belongs to the relative interior of the set $C = \operatorname{conv}\{e_1, \ldots, e_k\}$. The corresponding linear space L, orthogonal to $\operatorname{ri}(C)$, is the one-dimensional space generated by vector e. The nondegeneracy Condition (3.13) can be formulated here in the following form: vectors

$$\left(\frac{\partial f_{i_s}(\bar{x})}{\partial x_1}, \ldots, \frac{\partial f_{i_s}(\bar{x})}{\partial x_n}, 1\right) \in \mathbb{R}^{n+1}, \qquad s = 1, \ldots, k,$$

are linearly independent.

As another example, consider the setting of Example 2.3. Let $\overline{X} \in \mathcal{S}^p$ and $\lambda_k(\overline{X})$ be a leading eigenvalue of $\overline{X}$ of multiplicity r. Let $\Xi(\cdot)$ be a mapping satisfying Properties (i)–(iv) specified in Example 2.3, and $F(\cdot) := \Xi(\cdot) - \alpha I_r$. Consider the function $f(\cdot) := \lambda_k(\cdot) + \cdots + \lambda_{k+l-1}(\cdot)$ for some $l \in \{1, \ldots, r\}$. We have here that f is $g \circ F$ decomposable at $\overline{X}$ and $DF(\overline{X})$ is onto. The corresponding space L is the one-dimensional space generated by the identity matrix I_r, and the set $\mathcal{M}$ is given by

$$\mathcal{M} = \left\{X \in \mathcal{O}\colon \lambda_{k-1}(X) > \lambda_k(X) = \cdots = \lambda_{k+r-1}(X) > \lambda_{k+r}(X)\right\}, \tag{3.18}$$

where $\mathcal{O} \subset \mathcal{S}^p$ is a neighborhood of $\overline{X}$. Because $DF(\overline{X})$ is onto, the above set $\mathcal{M}$ is a smooth manifold.

Now let $A\colon \mathbb{R}^n \to \mathcal{S}^p$ be a smooth mapping such that $A(\bar{x}) = \overline{X}$ for some $\bar{x} \in \mathbb{R}^n$. Then the function $f(A(\cdot))$ is $g \circ H$ decomposable at $\bar{x}$, with $H(\cdot) := F(A(\cdot))$. The corresponding nondegeneracy Condition (3.13) takes the form

$$DA(\bar{x})\mathbb{R}^n + T_{\overline{X}}\mathcal{M} = \mathcal{S}^p, \tag{3.19}$$

and means that the mapping A intersects $\mathcal{M}$ transversally at $\bar{x}$. In the eigenvalue optimization, transversality Condition (3.19) was introduced in Shapiro and Fan (1995).

4. Optimality conditions and locally convergent algorithms. In this section we consider the optimization problem

$$\operatorname*{Min}_{x \in \mathcal{O}} f(x), \tag{4.1}$$

where, as before, $\mathcal{O}$ is an open neighborhood of a point $\bar{x} \in \mathbb{R}^n$. We assume that f is $g \circ F$ decomposable at $\bar{x}$ and is *real valued* on $\mathcal{O}$. Let us observe that set constraints of the form $x \in K$, where K is a closed subset of $\mathbb{R}^n$, can be absorbed into the objective function. One obvious way is to add the indicator function $i_K(\cdot)$ to $f(\cdot)$. This, however, may destroy the real valuedness of the objective function. Alternatively, we may add the penalty function $\gamma \operatorname{dist}_K(\cdot)$ to $f(\cdot)$. Because locally $f(\cdot)$ is Lipschitz continuous, by taking γ bigger than the corresponding Lipschitz constant of $f(\cdot)$, we obtain that in a neighborhood of $\bar{x}$ the problem of minimizing $f(\cdot) + \gamma \operatorname{dist}_K(\cdot)$ is equivalent to (4.1); i.e., $\gamma \operatorname{dist}_K(\cdot)$ is an exact penalty term. Further, suppose that K is cone reducible at $\bar{x}$ in the sense of Definition 2.1, and consider the penalty term $\gamma \operatorname{dist}_C(\Phi(\cdot))$. Because of the equivalence relations (2.4), we have that for γ large enough this is also an exact penalty term. Recall that the sum of two decomposable functions is also decomposable (see Remark 2.3), and hence this penalty term preserves decomposability of the objective function.

We can formulate the optimization problem (4.1) in the following equivalent form

$$\operatorname*{Min}_{(x,c) \in \mathcal{O} \times \mathbb{R}} c \quad \text{subject to} \quad (F(x), c) \in Q, \tag{4.2}$$

where $Q := \operatorname{epi} g$. Because g is real valued, convex, and positively homogeneous, the set Q is a closed convex cone in $\mathbb{R}^{m+1}$ with a nonempty interior. Therefore, the Problem (4.2) can be treated in the framework of cone constraint optimization (see, e.g., Bonnans and Shapiro 2000, §3.4.1). General results of that theory can be applied to the Problem (4.2).

It is not difficult to write first-order optimality conditions for the Problem (4.1) directly. We have that if $\bar{x}$ is an optimal solution of (4.1), then $f'(\bar{x}, h) \geq 0$ for all $h \in \mathbb{R}^n$. This is equivalent to the condition $0 \in \partial f(\bar{x})$. By (3.5) we then obtain the following first-order necessary condition for $\bar{x}$ to be an optimal solution of (4.1):

$$0 \in [DF(\bar{x})]^* \partial g(0). \tag{4.3}$$

Consider the Lagrangian $L(x, \lambda) := \lambda^T F(x)$, where $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$, of the Problem (4.1), and the set of Lagrange multipliers

$$\Lambda(\bar{x}) := \big\{\lambda \in \partial g(0)\colon D_x L(\bar{x}, \lambda) = 0\big\} = \big\{\lambda \in \partial g(0)\colon [DF(\bar{x})]^* \lambda = 0\big\}. \tag{4.4}$$

Condition (4.3) means, of course, that the set $\Lambda(\bar{x})$ is nonempty. It could be observed that the point $(F(\bar{x}), 1)$ belongs to the interior of the cone Q, and hence the Slater condition always holds for the Problem (4.2). We say that the point $\bar{x}$ is *stationary* if Condition (4.3) is satisfied or, equivalently, the set $\Lambda(\bar{x})$ is nonempty.

Let us observe that under the nondegeneracy Condition (3.13) the following holds:

$$[DF(\bar{x})]^*(\operatorname{ri} \partial g(0)) = \operatorname{ri}\big([DF(\bar{x})]^* \partial g(0)\big) = \operatorname{ri} \partial f(\bar{x}). \tag{4.5}$$

Indeed, we have that $[DF(\bar{x})]^* y = 0$ iff $y \in [DF(\bar{x})\mathbb{R}^n]^\perp$ and the nondegeneracy Condition (3.13) is equivalent to the condition

$$[DF(\bar{x})\mathbb{R}^n]^\perp \cap L^\perp = \{0\}. \tag{4.6}$$

It follows that $[DF(\bar{x})]^*$ restricted to $L^\perp$ is one-to-one, and because $L^\perp$ is parallel to the affine space generated by $\partial g(0)$, $[DF(\bar{x})]^*$ restricted to $\partial g(0)$ is also one-to-one. The first equality in (4.5) then follows. The second equality in (4.5) simply follows from (3.5).

The leneality space of the cone Q is formed by points $(y, g(y))$, $y \in L$. Therefore, the *nondegeneracy* condition discussed in Bonnans and Shapiro (2000, §4.6.1) coincides here with the nondegeneracy Condition (3.13). Suppose also that $\bar{x}$ is stationary. Under the nondegeneracy Condition (3.13), the *strict complementarity* condition (cf. Bonnans and Shapiro 2000, Definition 4.74) means here that

$$0 \in \operatorname{ri} \partial f(\bar{x}), \tag{4.7}$$

or, equivalently, that $f'(\bar{x}, h) > 0$ for all $h \in N_{\mathscr{M}}(\bar{x}) \setminus \{0\}$. By the general theory we have that if the nondegeneracy Condition (3.13) holds, then $\Lambda(\bar{x}) = \{\bar{\lambda}\}$ is a singleton. Together with (4.5), this implies that the strict complementarity Condition (4.7) is equivalent to the Condition

$$\bar{\lambda} \in \operatorname{ri} \partial g(0). \tag{4.8}$$

Conversely, if $\Lambda(\bar{x})$ is a singleton and the strict complementarity condition holds, then the nondegeneracy condition follows (Bonnans and Shapiro 2000, Proposition 4.75).

We say that the *quadratic growth* condition holds at $\bar{x}$ if for some $\kappa > 0$ and all x in a neighborhood of $\bar{x}$ the following holds:

$$f(x) \geq f(\bar{x}) + \kappa \|x - \bar{x}\|^2.$$

Clearly, the above quadratic growth condition implies that $\bar{x}$ is a locally optimal solution of (4.1). Suppose that $\bar{x}$ is stationary, and consider the so-called *critical cone*

$$\mathscr{C}(\bar{x}) := \{h \in \mathbb{R}^n\colon f'(\bar{x}, h) = 0\}. \tag{4.9}$$

Because $f'(\bar{x}, h) = g(DF(\bar{x})h)$ and $g(\cdot)$ is the support function of the set $\partial g(0)$, we have

$$\mathscr{C}(\bar{x}) = \{h \in \mathbb{R}^n\colon g(DF(\bar{x})h) = 0\} = \{h \in \mathbb{R}^n\colon \sup\nolimits_{\lambda \in \partial g(0)} [DF(\bar{x})^* \lambda]^T h = 0\}.$$

Because $\bar{x}$ is stationary, it follows that for any $\bar{\lambda} \in \Lambda(\bar{x})$,

$$\mathscr{C}(\bar{x}) = \{h \in \mathbb{R}^n\colon DF(\bar{x})h \in N_{\partial g(0)}(\bar{\lambda})\}. \tag{4.10}$$

We assume in the remainder of this section that $F(\cdot)$ is twice continuously differentiable. We have that if $\bar{x}$ is an optimal solution of (4.1), then

$$\inf_{w\in\mathbb{R}^n} f''(\bar{x}; h, w) \geq 0, \qquad \forall\, h \in \mathscr{C}(\bar{x}). \tag{4.11}$$

Moreover, suppose that the point $\bar{x}$ is stationary. Then, because of the parabolic regularity of f (see Proposition 3.1), we have that the following condition is necessary and sufficient for the quadratic growth condition to hold at $\bar{x}$:

$$\inf_{w\in\mathbb{R}^n} f''(\bar{x}; h, w) > 0, \qquad \forall\, h \in \mathscr{C}(\bar{x}), \tag{4.12}$$

(Bonnans and Shapiro 2000, Proposition 3.105).

Recall Formula (3.9) for $f''(\bar{x}; h, w)$. Consider a point $h \in \mathscr{C}(\bar{x})$. By (4.9) we have that $g(DF(\bar{x})h) = 0$, and hence $g'(DF(\bar{x})h, \cdot)$ is the support function of the set

$$\mathscr{Z}_h := \{z \in \partial g(0)\colon z^T DF(\bar{x})h = 0\}.$$

It follows that the infimum in the left-hand side of (4.11) and (4.12) is equal to the optimal value of the the following problem:

$$\begin{array}{ll} \underset{(w,\,c)\in\mathbb{R}^{n+1}}{\text{Min}} & c \\ \text{subject to} & \big(DF(\bar{x})w + D^2F(\bar{x})(h, h), c\big) \in Q_h^*, \end{array} \tag{4.13}$$

where Q_h^* is the epigraph of the support function of the set $\mathscr{Z}_h$. Moreover, the optimal value of the above problem is equal to the optimal value of its dual:

$$\underset{\lambda\in\Lambda(\bar{x})}{\text{Max}}\ \lambda^T D^2F(\bar{x})(h, h) \tag{4.14}$$

(cf. Bonnans and Shapiro 2000, p. 175). Note that

$$\lambda^T D^2F(\bar{x})(h, h) = D^2_{xx}L(\bar{x}, \lambda)(h, h),$$

and that the additional, so-called sigma term vanishes here. This is because the mapping $(x, c) \mapsto (F(x), c)$ in Problem (4.2) maps $(\bar{x}, 0)$ into the vertex (null vector) of the cone Q. We obtain the following second-order optimality conditions (cf. Bonnans and Shapiro 2000, Theorems 3.108 and 3.109).

PROPOSITION 4.1. *If $\bar{x}$ is an optimal solution of Problem* (4.1)*, then the following holds*

$$\sup_{\lambda\in\Lambda(\bar{x})} D^2_{xx}L(\bar{x}, \lambda)(h, h) \geq 0, \qquad \forall\, h \in \mathscr{C}(\bar{x}). \tag{4.15}$$

Suppose that $\bar{x}$ is a stationary point of Problem (4.1)*. Then the quadratic growth condition holds at $\bar{x}$ iff*

$$\sup_{\lambda\in\Lambda(\bar{x})} D^2_{xx}L(\bar{x}, \lambda)(h, h) > 0, \qquad \forall\, h \in \mathscr{C}(\bar{x})\backslash\{0\}. \tag{4.16}$$

Note again that Conditions (4.15) and (4.16) are equivalent to the respective Conditions (4.11) and (4.12).

Consider the problem

$$\underset{x\in\mathscr{M}}{\text{Min}}\, f(x), \tag{4.17}$$

where the set $\mathscr{M}$ is defined in (3.12). Suppose that the nondegeneracy condition holds, and hence $\mathscr{M}$ is a smooth manifold. Because $f(\cdot)$ coincides with $a^T F(\cdot)$ on $\mathscr{M}$ and $\mathscr{M}$ is defined by smooth constraints (3.12), we have by the standard first-order necessary conditions for smooth problems that if $\bar{x}$ is an optimal solution of (4.17), then there exists $z \in L^{\perp}$ such that $[DF(\bar{x})]^*(a+z) = 0$. Because $a \in \partial g(0)$, this means that there exists λ in the affine space generated by $\partial g(0)$ such that $[DF(\bar{x})]^*\lambda = 0$. This is equivalent to the condition that the affine space generated by $\partial f(\bar{x})$ includes 0. This is also equivalent to the condition that $f'(\bar{x}, h) \geq 0$ for all $h \in T_{\mathscr{M}}(\bar{x})$. Of course, the above conditions are weaker than the condition $0 \in \partial f(\bar{x})$; that is, a stationary point of (4.17) may be not a stationary point of (4.1).

Suppose now that $\bar{x}$ is a stationary point of the Problem (4.1) and the nondegeneracy and strict complementarity conditions hold. It follows, then, that $\Lambda(\bar{x}) = \{\bar{\lambda}\}$ is a singleton, and the critical cone $\mathscr{C}(\bar{x})$ coincides with the tangent space $T_{\mathscr{M}}(\bar{x})$. The second-order optimality Condition (4.16) then takes the form

$$D^2_{xx}L(\bar{x}, \bar{\lambda})(h, h) > 0, \qquad \forall\, h \in T_{\mathscr{M}}(\bar{x})\backslash\{0\}. \tag{4.18}$$

Under the above assumptions, (4.18) is a necessary and sufficient condition for the quadratic growth condition to hold at $\bar{x}$ for Problem (4.17) as well as for Problem (4.1).

The above analysis suggests that in order to solve Problem (4.1), one may try to solve the restricted Problem (4.17). Yet the restricted Problem (4.17) cannot be solved by algorithms designed for smooth problems because the function $f(\cdot)$ is not smooth on $\mathscr{O}$. So, suppose that we can construct a *smooth* function ψ: $\mathscr{O} \to \mathbb{R}$ such that $f(x) = \psi(x)$ for all $x \in \mathscr{M}$. Often such a function ψ can be constructed in a natural way. For instance, in the setting of Example 2.1 we can take ψ to be any convex combination of the functions $f_{i_1}, \ldots, f_{i_k}$. In the setting of Example 2.3 we can take $\psi(\cdot) := (l/r)[\lambda_k(\cdot) + \cdots + \lambda_{k+r-1}(\cdot)]$.

Clearly, Problem (4.17) is equivalent to the problem

$$\operatorname*{Min}_{x\in\mathscr{M}} \psi(x). \tag{4.19}$$

Problem (4.19) is smooth and can be solved, at least locally, by Newton-type methods. Under the nondegeneracy, strict complementarity, and second-order optimality Conditions (4.18), a locally optimal solution of (4.19) provides a locally optimal solution for the original Problem (4.1). Typically, Newton-type algorithms are locally convergent at a quadratic rate. In the eigenvalue optimization, such an algorithm was suggested in Overton (1988).

To apply a Newton-type algorithm to (4.19), one also needs a constructive way of defining the smooth manifold $\mathscr{M}$. In the setting of Example 2.1 the corresponding mapping $F(\cdot)$ is defined explicitly provided that the set $I(\bar{x})$, of active at the optimal solution functions, is known. In practice, of course, one estimates the set $I(\bar{x})$ by introducing a certain tolerance parameter. In the eigenvalue optimization describing the manifold $\mathscr{M}$ by smooth constraints is a more delicate problem (see, e.g., Shapiro and Fan 1995 for such a construction).

5. Perturbation analysis. In this section we discuss how optimal solutions of the Problem (4.1) behave under small perturbations of the objective function. As in the previous sections, we assume that f is $g \circ F$ decomposable at $\bar{x}$ and is real valued on $\mathscr{O}$. We say that a mapping $\mathscr{F}$: $\mathscr{O} \times \mathscr{U} \to \mathbb{R}^m$, where $\mathscr{U}$ is an open subset of a finite dimensional linear space, is a *smooth parameterization* of the mapping F, if $\mathscr{F}(\cdot, \cdot)$ is smooth (at least twice continuously differentiable) on $\mathscr{O} \times \mathscr{U}$ and $F(x) = \mathscr{F}(x, u_0)$ for some $u_0 \in \mathscr{U}$ and all $x \in \mathscr{O}$. Of course, the parameterization $\mathscr{F}(x, u)$ defines the corresponding parameterization $\varphi(x, u) := g(\mathscr{F}(x, u))$ of the objective function f. Consider the parameterized optimization problem

$$\operatorname*{Min}_{x\in\mathscr{O}} \varphi(x, u). \tag{5.1}$$

We have that for $u = u_0$, Problem (5.1) coincides with the (unperturbed) Problem (4.1).

Let us observe that Problem (5.1) can be formulated in the form of a cone-constrained problem similar to Problem (4.2). Perturbation (sensitivity) analysis of such parameterized problems is discussed extensively in Bonnans and Shapiro (2000). That theory can be applied to the considered case. What is specific about the considered problem is that the constraint mapping $(x, c) \mapsto (\mathscr{F}(x, u_0), c)$ maps $(\bar{x}, 0)$ into the vertex of the cone $Q := \operatorname{epi} g$. The cone Q has no "curvature" at its vertex point and consequently the additional, so-called sigma, term vanishes in the corresponding formulas. We already observed that in the previous section while discussing second-order optimality conditions. Note also that the cone Q is second-order regular, in the sense of Bonnans and Shapiro (2000, Definition 3.85), at its vertex point. Let us discuss some of the implications of the general theory to the parameterized Problem (5.1).

The nondegeneracy Condition (3.13) is stable under small (smooth) perturbations and, in a certain sense, is generic. That is, if the parameterization $\mathscr{F}(x, u)$ is sufficiently rich, then the nondegeneracy holds for almost every value of the parameter vector $u \in \mathscr{U}$. The "sufficiently rich" means here that the mapping $\mathscr{F}$ intersects L transversally, and the above is a classical result in differential geometry (see, e.g., Golubitsky and Guillemin 1973). For the eigenvalue optimization this generic property of the nondegeneracy was observed in Shapiro and Fan (1995, §2).

Let $\hat{x}(u) \in \mathscr{O}$ be a stationary point of Problem (5.1); i.e., first-order necessary optimality conditions hold at $\hat{x}(u)$. For the composite function $\varphi(x, u)$ these optimality conditions can be formulated in the following form:

$$\exists\, \hat{\lambda} \in \partial g(\hat{y}) \quad \text{such that} \quad D_x \mathscr{L}(\hat{x}, \hat{\lambda}, u) = 0, \tag{5.2}$$

where $\mathscr{L}(x, \lambda, u) := \lambda^T \mathscr{F}(x, u)$ is the Lagrangian of Problem (5.1), $\hat{x} = \hat{x}(u)$, $\hat{y} := \mathscr{F}(\hat{x}, u)$, and $\hat{\lambda} = \hat{\lambda}(u)$ is the corresponding Lagrange multipliers vector (see, e.g., Bonnans and Shapiro 2000, p. 219). Of course, for $u = u_0$ and $\hat{x} = \bar{x}$ we have that $\mathscr{F}(\bar{x}, u_0) = 0$, and Condition (5.2) is equivalent to Condition (4.3) for the unperturbed problem. Also, recall Formula (3.3) for $\partial g(y)$.

Suppose that the point $\bar{x}$ is a stationary point of the unperturbed problem and the nondegeneracy and strict complementarity conditions hold at $\bar{x}$. There then exists a unique vector $\bar{\lambda}$ of Lagrange multipliers and $\bar{\lambda} \in \operatorname{ri} \partial g(0)$. Moreover, for $u \in \mathscr{U}$ sufficiently close to u_0 and $\hat{x}(u)$ sufficiently close to $\bar{x}$, we have that $\hat{\lambda}(u)$ is close to $\bar{\lambda}$, and hence it follows by (5.2) and (3.3) that $\hat{\lambda}(u) \in \operatorname{ri} \partial g(0)$. That is, strict complementarity is also stable under small perturbations. Suppose also that the second-order Condition (4.18) holds. Then $\bar{x}$ is a locally optimal solution of the restricted Problem (4.17) as well as of the (unperturbed) Problem (4.1), and moreover $\bar{x}$ is a locally unique stationary point of (4.1). By compactness arguments it follows then that there exists a locally optimal solution of $\hat{x}(u) \in \mathscr{O}$ of (5.1) converging to $\bar{x}$ as u tends to u_0. It follows that $\hat{x}(u)$ is a stationary point of (5.1) for u sufficiently close to u_0, and $\hat{x}(u) \in \mathscr{M}(u)$, where

$$\mathscr{M}(u) := \left\{ x \in \mathscr{O} : \mathscr{F}(x, u) \in L \right\}. \tag{5.3}$$

Under the above conditions, local behavior of $\hat{x}(u)$ is explained by the classical Implicit Function Theorem. That is, for all u in a neighborhood of u_0, $\hat{x}(u)$ is a unique stationary and optimal solution of (5.1) in a neighborhood of $\bar{x}$, $\hat{x}(u)$ is continuously differentiable at $u = u_0$, and its differential $D\hat{x}(u_0)d$ is given by the optimal solution $\hat{h} = \hat{h}(d)$ of the linearized system:

$$\begin{aligned} &\operatorname*{Min}_{h \in \mathbb{R}^n} \; D^2_{xx} \mathscr{L}(\bar{x}, \bar{\lambda}, u_0)(h, h) + D^2_{xu} \mathscr{L}(\bar{x}, \bar{\lambda}, u_0)(h, d), \\ &\text{subject to } \; D_x \mathscr{F}(\bar{x}, u_0)h + D_u \mathscr{F}(\bar{x}, u_0)d \in N_{\partial g(0)}(\bar{\lambda}). \end{aligned} \tag{5.4}$$

Recall that the strict complementarity condition means that $\bar\lambda \in \operatorname{ri} \partial g(0)$. Therefore, it implies that $N_{\partial g(0)}(\bar\lambda)$ coincides with the linear space L. That is, under the strict complementarity condition, the constraint of the Problem (5.4) is just the linearization, at the point $(\bar x, u_0)$, of the constraint $\mathscr{F}(x, u) \in L$.

The above is quite a standard result in sensitivity analysis of optimization problems (cf. Bonnans and Shapiro 2000; Shapiro and Fan 1995; §5). Note that (5.4) is a quadratic programming problem subject to linear (or rather affine) constraints, the quadratic form $D^2_{xx}\mathscr{L}(\bar x, \bar\lambda, u_0)(\cdot,\cdot)$ coincides with $D^2_{xx}L(\bar x, \bar\lambda)(\cdot,\cdot)$, and that the affine space defined by the constraints of (5.4) is parallel to the tangent space $T_{\mathscr{M}}(\bar x)$. Therefore, the second-order Condition (4.18) ensures that Problem (5.4) possesses unique optimal solution $\hat h = \hat h(d)$, which is a linear function of d. One can also write (first-order) optimality conditions for (5.4), reducing it to a system of linear equations involving h and vector α of Lagrange multipliers of (5.4). Under the above assumptions, that system has a unique solution $(\hat h(d), \hat\alpha(d))$, which is a linear function of d. As was stated above, $D\hat x(u_0)d = \hat h(d)$, and moreover, it holds that $D\hat\lambda(u_0)d = \hat\alpha(d)$.

We have here that under the nondegeneracy, strict complementarity, and second-order Condition (4.18), the locally optimal solution $\hat x(u)$ is a smooth function of u and does not leave the manifold $\mathscr{M}(u)$ for all u sufficiently close to u_0. While relaxing these conditions, sensitivity analysis becomes considerably more involved. Yet it is possible to show that, under quite weak assumptions, $\hat x(u)$ is directionally differentiable with the directional derivative given as an optimal solution of an auxiliary problem (see Bonnans and Shapiro 2000). We give below a simplified version of a more general result.

THEOREM 5.1. *Suppose that the Lagrange multipliers set $\Lambda(\bar x) = \{\bar\lambda\}$ is a singleton and the second-order condition* (4.16) *is satisfied. Then the following holds*:

(i) *For all $u \in \mathscr{U}$ in a neighborhood of u_0, Problem* (5.1) *has a locally optimal solution $\hat x(u)$ converging to $\bar x$ as $u \to u_0$.*

(ii) *There exists a constant $\gamma > 0$ such that for all $u \in \mathscr{U}$ in a neighborhood of u_0 and a locally optimal solution $\hat x(u)$ of* (5.1) *in a neighborhood of $\bar x$, the following holds*

$$\|\hat x(u) - \bar x\| \le \gamma \|u - u_0\|. \tag{5.5}$$

(iii) *If, for $u(t) := u_0 + td$, $\hat x(u(t))$ is a locally optimal solution of* (5.1) *converging to $\bar x$ as $t \downarrow 0$, then any accumulation point of $[\hat x(u(t)) - \bar x]/t$, as $t \downarrow 0$, is an optimal solution of* (5.4).

(iv) *If Problem* (5.4) *has a unique optimal solution $\hat h = \hat h(d)$, then the directional derivative $\hat x'(u_0, d)$ exists and is equal to $\hat h(d)$.*

It follows that, under the assumptions of the above theorem, the locally optimal solution $\hat x(u)$ leaves the manifold $\mathscr{M}(u)$ under small perturbations in the direction d if every optimal solution $\hat h = \hat h(d)$ of the Problem (5.4) is such that $D_x\mathscr{F}(\bar x, u_0)\hat h + D_u\mathscr{F}(\bar x, u_0)d \notin L$.

Uniqueness of the optimal solution $\hat h(d)$ can be ensured by a sufficiently strong second-order sufficient condition. For example, if $D^2_{xx}L(\bar x, \bar\lambda)$ is positive definite over the linear space generated by the feasible set of (5.4), then $\hat h(d)$ is unique. If this holds for all d, then it follows that $\hat x(u)$ is directionally differentiable at u_0.

As was mentioned before, the nondegeneracy is a sufficient condition for uniqueness of the vector $\bar\lambda$ of Lagrange multipliers. A more general sufficient condition for uniqueness of $\bar\lambda$ is the following (see Bonnans and Shapiro 2000, Proposition 4.47):

$$DF(\bar x)\mathbb{R}^n + N_{\partial g(0)}(\bar\lambda) = \mathbb{R}^m. \tag{5.6}$$

Moreover, if the radial cone to $\partial g(0)$ at $\bar\lambda$ is closed, then (5.6) is also necessary for uniqueness of $\bar\lambda$. In particular, (5.6) is a necessary and sufficient condition for uniqueness of $\bar\lambda$

if the set $\partial g(0)$ is polyhedral, or in other words, if $g(\cdot)$ is piecewise linear. Under the strict complementarity condition we have that $N_{\partial g(0)}(\bar{\lambda}) = L$, and hence in that case (5.6) coincides with the nondegeneracy condition.

Now consider the following so-called uniform quadratic growth condition introduced in Bonnans and Shapiro (2000, Definition 5.16).

DEFINITION 5.1. Let $\bar{x}$ be a stationary point of Problem (4.1). It is said that the *uniform quadratic growth* condition holds at $\bar{x}$ if for any smooth parameterization $\mathcal{F}(x, u)$ of $F(x)$ there exist $c > 0$ and neighborhoods $\mathcal{X}$ and $\mathcal{U}$ of $\bar{x}$ and u_0, respectively, such that for any $u \in \mathcal{U}$ and any stationary point $\hat{x}(u) \in \mathcal{X}$ of the corresponding parameterized problem, the following holds:

$$\varphi(x, u) \geq \varphi(\hat{x}(u), u) + c\|x - \hat{x}(u)\|^2, \qquad \forall\, x \in \mathcal{X}. \tag{5.7}$$

The uniform quadratic growth has important implications for the behavior of stationary and locally optimal solutions. In particular, it implies that for all $u \in \mathcal{U}$ the parameterized problem has a unique, and continuous in u, stationary point $\hat{x}(u) \in \mathcal{X}$, and that $\hat{x}(u)$ is the minimizer of $\varphi(\cdot, u)$ over $\mathcal{X}$ (see Bonnans and Shapiro 2000, Theorem 5.17, Remark 5.18).

THEOREM 5.2. *Let $\bar{x}$ be a stationary point of Problem* (4.1). *Suppose that the nondegeneracy and strict complementarity conditions are satisfied at $\bar{x}$. Then the second-order Condition* (4.18) *is necessary and sufficient for the uniform quadratic growth condition to hold at $\bar{x}$.*

PROOF. It is clear that the uniform quadratic growth condition implies the quadratic growth condition. Because under the present assumptions Condition (4.18) is necessary (and sufficient) for the quadratic growth at $\bar{x}$, necessity of this condition for the uniform quadratic growth follows.

So let us prove the sufficiency. The following proof is similar to the proof of Theorem 5.27 in Bonnans and Shapiro (2000). We argue by contradiction. Suppose that the uniform quadratic growth condition does not hold for a smooth parameterization $\mathcal{F}(x, u)$. Then there exist sequences $u_n \to u_0$, $x_n \to \bar{x}$, and $h_n \to 0$ ($h_n \neq 0$) such that with x_n is associated a Lagrange multiplier λ_n, of the parameterized problem, and

$$\varphi(x_n + h_n, u_n) \leq \varphi(x_n, u_n) + o(\|h_n\|^2). \tag{5.8}$$

Because by the nondegeneracy assumption $\Lambda(\bar{x}) = \{\bar{\lambda}\}$ is a singleton, we have that $\lambda_n \to \bar{\lambda}$. Consider $\bar{h}_n := h_n/\|h_n\|$. By passing to a subsequence if necessary, we can assume that $\bar{h}_n$ converges to a vector $\bar{h}$. Clearly $\|\bar{h}\| = 1$, and hence $\bar{h} \neq 0$.

By (5.2) and (3.3) we have that $\varphi(x_n, u_n) = \mathcal{L}(x_n, \lambda_n, u_n)$. Because $\varphi(x, u) = \sigma_{\partial g(0)}(\mathcal{F}(x, u))$ and $\lambda_n \in \partial g(0)$, we also have that $\varphi(x_n + h_n, u_n) \geq \mathcal{L}(x_n + h_n, \lambda_n, u_n)$. Consequently, it follows by (5.8) that

$$\mathcal{L}(x_n + h_n, \lambda_n, u_n) - \mathcal{L}(x_n, \lambda_n, u_n) \leq o(\|h_n\|^2). \tag{5.9}$$

Because $D_x\mathcal{L}(x_n, \lambda_n, u_n) = 0$, by using a second-order Taylor expansion of the left-hand side of (5.9), with respect to h_n, and passing to the limit, we obtain that

$$D^2_{xx}L(\bar{x}, \bar{\lambda})(\bar{h}, \bar{h}) \leq 0. \tag{5.10}$$

Because $g(\cdot)$ is Lipschitz continuous and convex, we have

$$\begin{aligned} \varphi(x_n + h_n, u_n) &= g\big(\mathcal{F}(x_n, u_n) + D_x\mathcal{F}(x_n, u_n)h_n\big) + o(\|h_n\|) \\ &\geq g\big(\mathcal{F}(x_n, u_n)\big) + \lambda_n^T D_x\mathcal{F}(x_n, u_n)h_n + o(\|h_n\|). \end{aligned} \tag{5.11}$$

Moreover, by (5.2) we have that $\lambda_n^T D_x \mathcal{F}(x_n, u_n) h_n = 0$, and hence it follows from (5.8) and (5.11) that

$$\varphi(x_n + h_n, u_n) - \varphi(x_n, u_n) = o(\|h_n\|). \tag{5.12}$$

So far, we did not use the strict complementarity condition.

As was discussed earlier, the strict complementarity condition implies that $\mathcal{F}(x_n, u_n) \in L$ for sufficiently large n. Consequently, by passing to the limit we obtain by (5.12) that $g\big(DF(\bar{x})\bar{h}\big) = 0$, which means that $\bar{h} \in \mathscr{C}(\bar{x})$. Together with (5.10), this contradicts (4.18), and therefore the proof is complete. □

The above proof shows that without the assumption of strict complementarity, the condition for the Hessian matrix $D^2_{xx} L(\bar{x}, \bar{\lambda})$ to be positive definite is sufficient for the uniform quadratic growth, at $\bar{x}$, to hold. It seems that such a condition is too strong. General necessary and sufficient second-order conditions for the uniform quadratic growth are not known.

Acknowledgment. This work was supported by the National Science Foundation under Grant DMS-0073770.

References

Bonnans, J. F., A. Shapiro. 2000. *Perturbation Analysis of Optimization Problems*. Springer, New York.

Clarke, F. H. 1983. *Optimization and Nonsmooth Analysis*. Wiley, New York.

Golubitsky, M., V. Guillemin. 1973. *Stable Mappings and Their Singularities*. Springer-Verlag, New York.

Lemaréchal, C., F. Oustry, C. Sagastizábal. 1999. The $\mathcal{U}$-Lagrangian of a convex function. *Trans Amer. Math. Soc.* **352** 711–729.

Lewis, A. S. 2002. Active sets, nonsmoothness and sensitivity. *SIAM J. Optim.* **13** 702–705.

Mifflin, R. 1977. Semismooth and semiconvex functions in constrained optimization. *SIAM J. Control Optim.* **15** 957–972.

———, C. Sagastizábal. 2000. On $\mathcal{VU}$-theory for functions with primal-dual gradient structure. *SIAM J. Optim.* **11** 547–571.

Overton, M. L. 1988. On minimizing the maximum eigenvalue of a symmetric matrix. *SIAM J. Matrix Anal. Appl.* **9** 256–268.

Qi, L., J. Sun. 1993. A nonsmooth version of Newton's method. *Math. Programming* **58** 353–367.

Rockafellar, R. T., R. J-B. Wets. 1998. *Variational Analysis*. Springer-Verlag, New York.

Shapiro, A., M. K. H. Fan. 1995. On eigenvalue optimization. *SIAM J. Optim.* **5** 552–569.

A. Shapiro: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205; e-mail: ashapiro@isye.gatech.edu