Math. Program. 84: 269–281 (1999) / DOI 10.1007/s10107980029a

Robert Mifflin · Liqun Qi · Defeng Sun*

Properties of the Moreau-Yosida regularization of a piecewise C^2 convex function

Received December 14, 1995 / Revised version received May 7, 1998
Published online October 21, 1998

Abstract. In this paper we discuss second-order properties of the Moreau-Yosida regularization F of a piecewise twice continuously differentiable convex function f. We introduce a new constraint qualification in order to prove that the gradient of F is piecewise continuously differentiable. In addition, we discuss conditions, depending on the Hessians of the pieces, that guarantee positive definiteness of the generalized Jacobians of the gradient of F.

Key words. Moreau-Yosida regularization – piecewise smooth functions

1. Introduction

Consider the following minimization problem:

$$\min f(x), \tag{1}$$

where f is a finite-valued convex function defined on $\Re^n$.

Throughout this paper, we use $\|\cdot\|$ to denote the Euclidean norm on $\Re^n$. Let M be a symmetric positive definite $n \times n$ matrix. For any $x \in \Re^n$, let

$$\|x\|_M^2 = x^T Mx.$$

We let F be the Moreau [8]-Yosida [18] regularization of f, associated with M, defined by

$$F(u) = \min_{x \in \Re^n} \{ f(x) + \frac{1}{2}\|x-u\|_M^2\} \quad \text{for } u \in \Re^n. \tag{2}$$

It is well known that F is a continuously differentiable convex function defined on all of $\Re^n$ even though f may be nondifferentiable. The gradient of F at u is

$$G(u) \equiv \nabla F(u) = M(u - p(u)) \in \partial f(p(u)),$$

R. Mifflin: Department of Pure and Applied Mathematics, Washington State University, Pullman, WA 99164-3113, USA

L. Qi: School of Mathematics, The University of New South Wales, Sydney, New South Wales 2052, Australia

D. Sun: School of Mathematics, The University of New South Wales, Sydney, New South Wales 2052, Australia

Mathematics Subject Classification (1991): 90C25, 65K10, 52A41

* On leave from Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing 100080, P.R. China

where $p(u)$ is the unique solution of (2) and ∂f is the subdifferential mapping of f [14]. Here, $p(u)$ is called the proximal point of u. Furthermore, G is globally Lipschitz continuous with modulus $\|M\|$ and the set of minimizers of (1) is exactly the set of minimizers of

$$\min F(x). \tag{3}$$

See [5] for basic properties of F.

Recently, several authors have considered second-order properties of F, for example, see [4,5,12,15]. In the original version of [12], Qi conjectured that G is semismooth under a regularity condition if f is the maximum of several twice continuously differentiable convex functions. In [16], Sun and Han gave a proof of this conjecture under a constant rank constraint qualification for the case where $M = (1/\lambda)I$ and λ is a positive constant. In this paper we will consider the case where f is piecewise C^2 in the sense that for each $x \in \Re^n$

$$f(x) \in \{ f_j(x) : j \in J\}, \tag{4}$$

$J = \{1, ..., |J|\}$ is a finite index set and for each $j \in J$, f_j is a twice continuously differentiable function. Such a function is a generalization of a maximum of convex C^2 functions. To see this, let $f : \Re \to \Re$ be defined by

$$f(x) = \begin{cases} f_1(x) \text{ if } x \geq 2, \\ f_2(x) \text{ if } 0 \leq x \leq 2, \\ f_3(x) \text{ if } x \leq 0, \end{cases}$$

where $f_1(x) = x^2 - x$, $f_2(x) = x$ and $f_3(x) = -x^3$. Then f is of the form (4) with $J = \{1, 2, 3\}$ and f is convex, but not differentiable at the solution point $x = 0$. It is clear that $f_3(x)$ is not convex when $x > 0$, and $f(x) < \max\{ f_1(x), f_2(x), f_3(x)\}$ when $x < 0$. Under the so called affine independence preserving constraint qualification given below in Section 2, we prove that about u there is an open neighborhood $N(u)$ such that G is piecewise smooth on $N(u)$ [6,10,17], i.e., there exist a family of finitely many continuously differentiable vector-valued functions $G^1, ..., G^k$ defined on $N(u)$ such that for any $v \in N(u)$,

$$G(v) \in \{G^1(v), ..., G^k(v)\}.$$

Our constraint qualification is weaker than the constant rank constraint qualification used in [16]. It was proved in Qi [12] that all of the generalized Jacobians of G are positive definite at u if and only only if f is strongly convex about $p(u)$. Here we will discuss conditions depending on the Hessians of the f_j's which imply this positive definiteness result. To accomplish this we give an expression, as in [7], for a generalized Jacobian of p which depends on a basis matrix for a certain subspace and on a convex combination of $\nabla^2 f_j$ for j in a subset of J.

2. Constraint qualifications

Let

$$J(x) = \{j \in J : f_j(x) = f(x)\}.$$

Definition 1. *Affine Independence Constraint Qualification (AICQ): The affine independence qualification is said to hold at x if the vectors*

$$\left\{ \begin{pmatrix} \nabla f_j(x) \\ 1 \end{pmatrix} : j \in J(x) \right\}$$

are linearly independent.

Definition 2. [3] *Constant Rank Constraint Qualification (CRCQ): The constant rank constraint qualification is said to hold at x if there exists a neighbourhood V of x such that for every subset $K \subseteq J(x)$, the family of the vectors*

$$\left\{ \begin{pmatrix} \nabla f_j(z) \\ 1 \end{pmatrix} : j \in K \right\}$$

has the same rank (which depends on K) for all vectors $z \in V$.

Definition 3. *Affine Independence Preserving Constraint Qualification (AIPCQ): The affine independence preserving constraint qualification is said to hold at x if for every subset $K \subseteq J(x)$ for which there exists a sequence $\{x^k\}$ with $\{x^k\} \to x$, $K \subseteq J(x^k)$ and the vectors*

$$\left\{ \begin{pmatrix} \nabla f_j(x^k) \\ 1 \end{pmatrix} : j \in K \right\}$$

being linearly independent, it follows that the vectors

$$\left\{ \begin{pmatrix} \nabla f_j(x) \\ 1 \end{pmatrix} : j \in K \right\}$$

are linearly independent.

From the above definitions it can be shown that AICQ implies CRCQ and CRCQ implies AIPCQ, but not vice versa. In fact it is easy to give an example to show that the CRCQ holds, but AICQ does not hold. For an example where AIPCQ holds, but CRCQ does not hold, let $f : \Re \to \Re$ be defined by

$$f(x) = \max\{f_1(x), f_2(x)\},$$

where $f_i(x) = ix^2$ for $i = 1, 2$. Note that, $J(0) = \{1, 2\}$, and for any $y \neq 0$, $J(y) = \{2\}$. So, clearly, AIPCQ holds at $x = 0$. However, $(\nabla f_1(0), 1)$ and $(\nabla f_2(0), 1)$ are linearly dependent, and for any $y \neq 0$, $(\nabla f_1(y), 1)$ and $(\nabla f_2(y), 1)$ are linearly independent, so the CRCQ does not hold at $x = 0$. So even for this simple maximum function, the fact that ∇F is piecewise smooth does not follow from the result of Sun and Han [16]. However, F is actually twice continuously differentiable.

Throughout this paper we will use cl S, int S and conv S to denote the closure, interior and convex hull of a set S, respectively. For any locally Lipschitz continuous function $H : \Re^n \rightarrow \Re^m$, we denote the generalized Jacobian of H by ∂H as defined in [1]. $\partial H(x)$ is the convex hull of $\partial_B H(x)$, where as in [11],

$$\partial_B H(x) = \{\lim \nabla H(x^k) : \{x^k\} \rightarrow x \text{ and } H \text{ is differentiable at } x^k\}.$$

If $m = 1$ and H is convex, ∂H is the subdifferential of H.

For each $j \in J$ let

$$D_j \equiv \{y \in \Re^n : j \in J(y)\}.$$

By definition, f is finite-valued everywhere. Since f is convex, f is also continuous. Hence, D_j is closed for each j. Let

$$\tilde{D}_j \equiv \text{cl int } D_j.$$

Then

$$\tilde{D}_j \subseteq D_j.$$

For any $x \in \Re^n$, let

$$I(x) = \{j \in J : x \in \tilde{D}_j\}.$$

Then, we have $I(x) \subseteq J(x)$. But equality does not hold in some cases. For the above two-piece example, $I(0) = \{2\} \subset \{1, 2\} = J(0)$.

Lemma 1. [10] *If each f_j for $j \in J$ is continuously differentiable, then*

$$\partial_B f(x) = \{\nabla f_j(x) : j \in I(x)\}.$$

Then, since $I(x) \subseteq J(x)$, we have

$$\partial f(x) = \text{conv}\,\{\nabla f_j(x) : j \in I(x)\} \subseteq \text{conv}\,\{\nabla f_j(x) : j \in J(x)\}. \tag{5}$$

The latter set can be regarded as an overestimation of $\partial f(x)$, because, in general, $I(x)$ may not equal $J(x)$. In fact, the example in the introduction has $\partial f(0) = [0, 1] \neq [-1, 1] = \text{conv}\{\nabla f_j(0) : j \in J(0) = \{1, 2, 3\}\}$.

Lemma 2. *If f is a convex function and each f_j for $j \in J$ is twice continuously differentiable, then all the matrices $\nabla^2 f_j(x)$ for $j \in I(x)$ are positive semidefinite.*

Proof. Since $\cup_{j\in J} D_j = \Re^n$ and J is finite, $\cup_{j\in J}\text{int } D_j$ is a dense, open subset of $\Re^n$. Note that

$$\cup_{j\in J} \tilde{D}_j = \Re^n.$$

So, for any $x \in \Re^n$, $I(x)$ is nonempty. Suppose $j \in I(x)$. Then $x \in \tilde{D}_j$. Let N_j be an arbitrary open ball contained in int D_j. Then we have

$$f(y) = f_j(y) \text{ for } y \in N_j. \tag{6}$$

The convexity of f and (6) imply that f_j is convex on N_j, so $\nabla^2 f_j(y)$ is positive semidefinite for any $y \in N_j$. This means that $\nabla^2 f_j(y)$ is positive definite for any y in int D_j. Since $x \in \tilde{D}_j$, it follows that $\nabla^2 f_j(x)$ is positive semidefinite.

□

Let $\mathcal{M}(u)$ denote the set of multiplier vectors $\alpha(u)$ such that

$$\begin{cases} G(u) = M(u - p(u)) = \sum_{j \in J(p(u))} \alpha_j(u) \nabla f_j(p(u)), \\ \alpha_j(u) \geq 0 \text{ for } j \in J(p(u)) \quad \text{and} \sum_{j \in J(p(u))} \alpha_j(u) = 1. \end{cases} \tag{7}$$

The nonemptiness of $\mathcal{M}(u)$ follows from (2), the definitions of $p(u)$ and $G(u)$ and Lemma 1.

For a nonnegative vector $d \in \Re^{|J|}$, we let supp(d), called the support of d, be the subset of J consisting of all the indices i for which $d_i > 0$. Define $\mathcal{B}(u)$ as a family of subsets of $J(p(u))$ as follows: $K \in \mathcal{B}(u)$ if and only if supp$(\alpha(u)) \subseteq K \subseteq J(p(u))$ for some $\alpha(u) \in \mathcal{M}(u)$ and the vectors

$$\left\{ \begin{pmatrix} \nabla f_j(p(u)) \\ 1 \end{pmatrix} : j \in K \right\}$$

are linearly independent. This family $\mathcal{B}(u)$ is nonempty, because $\mathcal{M}(u)$ has an extreme point which easily yields a desired index set K with the stated properties. For $K \in \mathcal{B}(u)$ let $\alpha_K(u)$ be the corresponding $|K|$-dimensional subvector of $\alpha(u)$ obtained by deleting from $\alpha(u)$ elements $\alpha_j(u)$ for $j \in J(p(u))\backslash K$. Due to the linear independence assumption $\alpha_K(u)$ is uniquely determined by

$$\alpha_K(u) = \left(\begin{pmatrix} \nabla f_K(p(u)) \\ e^K \end{pmatrix}^T \begin{pmatrix} \nabla f_K(p(u)) \\ e^K \end{pmatrix} \right)^{-1} \begin{pmatrix} \nabla f_K(p(u)) \\ e^K \end{pmatrix}^T \begin{pmatrix} G(u) \\ 1 \end{pmatrix} \tag{8}$$

where $\nabla f_K(p(u))$ is an $n \times |K|$ matrix with columns $\nabla f_j(p(u))$ for $j \in K$ and e^K is a $|K|$-dimensional row vector of ones. Note that it is possible that $\alpha_j(u) = 0$, for some, but not all, $j \in K$.

The following lemma is similar to a result of [10] for normal maps.

Lemma 3. *Suppose that each f_j for $j \in J$ is a continuously differentiable function and the affine independence preserving constraint qualification holds at the proximal point $p(u)$. Then about u there exists an open neighbourhood $N(u)$ such that* $\mathcal{B}(v) \subseteq \mathcal{B}(u)$ for all $v \in N(u)$.

Proof. If the conclusion of this lemma does not hold, then there exists a sequence $\{u^k\}$ converging to u such that for all k, there is an index set $K^k \in \mathcal{B}(u^k)\backslash\mathcal{B}(u)$. Since there are only finitely many such index sets, by taking a subsequence if necessary, we may assume that these index sets K^k are the same for all k. By letting K be this common index set, we have that for each k the vectors

$$\left\{ \begin{pmatrix} \nabla f_j(p(u^k)) \\ 1 \end{pmatrix} : j \in K \right\} \tag{9}$$

are linearly independent and there exists $\alpha(u^k) \in \mathcal{M}(u^k)$ such that $\text{supp}(\alpha(u^k)) \subseteq K \subseteq J(p(u^k))$, but $K \notin \mathcal{B}(u)$. Clearly $K \subseteq J(p(u))$. By AIPCQ, the vectors

$$\left\{ \begin{pmatrix} \nabla f_j(p(u)) \\ 1 \end{pmatrix} : j \in K \right\}$$

must be linearly independent. So the only way for $K \notin \mathcal{B}(u)$ is that there does not exist $\alpha(u) \in \mathcal{M}(u)$ such that $\text{supp}(\alpha(u)) \subseteq K$. Since $\{\alpha_K(u^k)\}$ is bounded $(0 \le \alpha_j(u^k) \le 1$, $j \in K)$, it produces at least one accumulation point, say, $\alpha'_K(u)$. Define $\alpha(u)$ by

$$\alpha_i(u) = \begin{cases} \alpha'_i(u) & \text{if } i \in K, \\ 0 & \text{if } i \in J(p(u)) \setminus K. \end{cases}$$

Clearly, $\alpha(u) \in \mathcal{M}(u)$ and $\text{supp}(\alpha(u)) \subseteq K$. This is a contradiction.

□

3. The piecewise smoothness of G

In this section we will discuss the piecewise smoothness of G at u under the assumption that the affine independence preserving constraint qualification holds at $p(u)$.

From Lemma 3 we know that there exists a neighbourhood of u, denoted by $N_1(u)$, such that

$$\mathcal{B}(v) \subseteq \mathcal{B}(u), \ \forall v \in N_1(u). \tag{10}$$

For every $v \in \Re^n$, from Lemma 1 and the definition of $\mathcal{B}(v)$ we know that there exists $K \in \mathcal{B}(v)$ such that $K \subseteq I(p(v))$. Define

$$\mathcal{B}'(v) = \{K : K \in \mathcal{B}(v) \text{ and } K \subseteq I(p(v))\}.$$

For v close to u, $I(p(v)) \subseteq I(p(u))$, so from (10) we know that there exists a neighbourhood of u, denoted by $N_2(u)$, such that

$$\mathcal{B}'(v) \subseteq \mathcal{B}'(u), \ \forall v \in N_2(u). \tag{11}$$

In the following discussion we set up certain definitions in order to obtain relevant consequences of the Implicit Function Theorem under the assumptions of Lemmas 2 and 3.

Suppose $K \in \mathcal{B}'(u)$. Choose some $i \in K$ and let

$$\bar{K} = K \setminus \{i\}$$

with $\bar{K}$ being empty if $|K| = 1$. Also, let $\alpha_{\bar{K}}(u)$ be the (possibly vacuous) subvector of $\alpha_K(u)$ obtained by deleting $\alpha_i(u)$ from $\alpha_K(u)$. Relative to these definitions consider the following vector function and corresponding system of equations

$$H^K(x, q, v) := \begin{pmatrix} \left(f_j(x) - f_i(x)\right)_{j\in\bar{K}} \\ M^{-1}(\sum_{j\in\bar{K}} q_j \nabla f_j(x) + (1 - \sum_{j\in\bar{K}} q_j)\nabla f_i(x)) + x - v \end{pmatrix} = 0, \tag{12}$$

where $v \in \Re^n$ is a parameter vector and $(x, q) \in \Re^n \times \Re^{|\bar{K}|}$ are vectors of variables with q vacuous if $|K| = 1$. If we set $x^0 = p(u)$, $q^0 = \alpha_{\bar{K}}(u)$, and $v^0 = u$, then, by (7) and the definition of $J(p(u))$, we have

$$H^K(x^0, q^0, v^0) = 0.$$

The matrix of partial derivatives of $H^K(x, q, v)$ with respect to (x, q) is

$$A^K(x, q) := \partial_{(x,q)} H^K(x, q, v) = \begin{pmatrix} \nabla \bar{f}_{\bar{K}}(x)^T & 0 \\ B^K(x, q) & M^{-1}\nabla \bar{f}_{\bar{K}}(x) \end{pmatrix}, \tag{13a}$$

where for $j \in \bar{K}$, $\nabla \bar{f}_j(x) := \nabla f_j(x) - \nabla f_i(x)$ is the j-th column of $\nabla \bar{f}_{\bar{K}}(x)$ and

$$B^K(x, q) = I + M^{-1}(\sum_{j\in\bar{K}} q_j \nabla^2 f_j(x) + (1 - \sum_{j\in\bar{K}} q_j)\nabla^2 f_i(x)). \tag{13b}$$

If $|K| = 1$ then q is vacuous, the top $|\bar{K}|$ rows of (12) are vacuous and the summations over $\bar{K}$ in (12) and (13b) are vacuous, so that

$$A^K(x, q) = B^K(x, q) = I + M^{-1}\nabla^2 f_i(x) \text{ with } K = \{i\}.$$

Lemma 4. *Suppose that each f_j for $j \in J$ is twice continuously differentiable, f is convex, and the affine independence preserving constraint qualification holds at the proximal point $p(u)$. Then for each $K \in \mathcal{B}'(u)$, there exist an open neighbourhood U^K of $v^0(= u)$ and an open neighbourhood W^K of (x^0, q^0) such that when $v \in \mathrm{cl}\ U^K$, the equations $H^K(x, q, v) = 0$ have a unique solution $(x^K(v), q^K(v)) \in \mathrm{cl}\ W^K$ where $q^K(v)$ is vacuous if $|K| = 1$. Moreover, $(x^K(v), q^K(v))$ is continuously differentiable on U^K and*

$$\partial_{(x,q)} H^K(x^K(v), q^K(v), v) \begin{pmatrix} \nabla x^K(v) \\ \nabla q^K(v) \end{pmatrix} = -\partial_v H^K(x^K(v), q^K(v), v)\,,$$

i.e.,

$$A^K(x^K(v), q^K(v)) \begin{pmatrix} \nabla x^K(v) \\ \nabla q^K(v) \end{pmatrix} = -\begin{pmatrix} 0 \\ -I \end{pmatrix}. \tag{14}$$

Proof. We consider the case where $|K| > 1$ as the case where $|K| = 1$ is similar, but simpler, since $\alpha_i(u) = 1$ and q is vacuous when $\bar{K}$ is empty.

From Lemma 2 and the fact that $q^0 = \alpha_{\bar{K}}(u) \geq 0$ and $1 - \sum_{j \in \bar{K}} q_j^0 = \alpha_i(u) \geq 0$, there exists a neighbourhood of (x^0, q^0, v^0), denoted by $N^K(x^0, q^0, v^0)$, such that $B^K(x, q)$ is a positive definite matrix when $(x, q, v) \in N^K(x^0, q^0, v^0)$. Since $K \in \mathcal{B}'(u) \subseteq \mathcal{B}(u)$, the vectors

$$\left\{ \begin{pmatrix} \nabla f_j(x^0) \\ 1 \end{pmatrix} : j \in K \right\}$$

are linearly independent and, by continuity, there exists a neighbourhood of (x^0, q^0, v^0) (which we also denote by $N^K(x^0, q^0, v^0)$) such that the vectors

$$\left\{ \begin{pmatrix} \nabla f_j(x) \\ 1 \end{pmatrix} : j \in K \right\} \tag{15}$$

are linearly independent when $(x, q, v) \in N^K(x^0, q^0, v^0)$. Since the vectors in (15) are linearly independent, it follows that the vectors

$$\{\nabla \bar{f}_j(x) : j \in \bar{K}\} \tag{16}$$

are linearly independent. Thus, $A^K(x, q)$ is nonsingular when $(x, q, v) \in N^K(x^0, q^0, v^0)$, and the desired results follow from the Implicit Function Theorem [9].

□

For $K \in \mathcal{B}'(u)$ and $x^K : U^k \to \Re^n$ as defined in Lemma 4, define $G^K : U^K \to \Re^n$ by

$$G^K(v) = M(v - x^K(v)) \text{ for } v \in U^K. \tag{17}$$

Then $G^K(v)$ is continuously differentiable on U^K. So far, we have obtained a family of finitely many continuously differentiable functions:

$$G^K : U^K \to \Re^n, \quad K \in \mathcal{B}'(u).$$

Remark 1. If K is in the less restrictive set $\mathcal{B}(u)$ similar corresponding functions $x^K(v)$ and $G^K(v)$ can be shown to exist provided $\nabla^2 f_j(p(u))$ is positive semidefinite for each $j \in J(p(u)) \backslash I(p(u))$. This extra condition is required to imply that $B^K(x^0, q^0)$ is positive definite when $K \in \mathcal{B}(u) \supseteq \mathcal{B}'(u)$, because Lemma 2 only covers the indices in $I(p(u))$.

The following theorem on the piecewise smoothness of G is the main result of this section.

Theorem 1. *Suppose that each f_j for $j \in J$ is twice continuously differentiable, f is convex, and the affine independence preserving constraint qualification holds at the proximal point $p(u)$. Then about u there exists an open neighbourhood $N(u)$ such that G, the gradient function of the Moreau-Yosida regularization of f, satisfies*

$$G(v) \in \{G^K(v) : K \in \mathcal{B}'(u)\}, \quad v \in N(u),$$

i.e., G is piecewise smooth on $N(u)$.

Proof. By (7) and (11), for any $v \in N_2(u)$, there exists $K \in \mathcal{B}'(v) \subseteq \mathcal{B}'(u)$ such that

$$\begin{cases} G(v) = M(v - p(v)) = \sum_{j \in K} \alpha_j(v) \nabla f_j(p(v)), \\ \sum_{j \in K} \alpha_j(v) = 1 \end{cases}$$

and for all $j \in K$ we have

$$f(p(v)) = f_j(p(v)).$$

Let $N(u) \subseteq \{\cap_{K \in \mathcal{B}'(u)} U^K\} \cap N_2(u)$ be an open neighbourhood of $v^0 (= u)$ such that for any $v \in N(u)$ and $K \in \mathcal{B}'(v)$

$$(p(v), \alpha_{\bar{K}}(v)) \in \text{cl } W^K \tag{18}$$

where $\alpha_{\bar{K}}(v)$ is vacuous if $|K|$=1. Relation (18) can be satisfied due to the facts that as $v \to u$

$$p(v) \to p(u) \text{ and } \alpha_K(v) \to \alpha_K(u)$$

where $\alpha_K(\cdot)$ is defined by (8).

For $K \in \mathcal{B}'(u)$, let $V^K(u) = \{v \in N(u) : K \in \mathcal{B}'(v)\}$. Then

$$N(u) = \cup_{K \in \mathcal{B}'(u)} V^K(u). \tag{19}$$

So, for any $v \in N(u)$ there exists $K \in \mathcal{B}'(u)$ such that $v \in V^K(u)$. But in this case we know that

$$H^K(p(v), \alpha_{\bar{K}}(v), v) = 0 \tag{20}$$

and

$$(p(v), \alpha_{\bar{K}}(v)) \in \text{cl } W^K,$$

so, it follows that

$$(p(v), \alpha_{\bar{K}}(v)) = (x^K(v), q^K(v)) \tag{21}$$

from the uniqueness of the solution in cl W^K of the equations $H^K(x, q, v) = 0$ for $v \in V^K(u) \subseteq$ cl U^K. So, by (17), for $v \in V^K(u)$

$$\begin{aligned} G(v) &= M(v - p(v)) \\ &= M(v - x^K(v)) \\ &= G^K(v). \end{aligned}$$

This means that for any $v \in N(u)$, there exists at least one continuously differentiable function $G^K : U^K \supseteq N(u) \to \Re^n$ such that

$$G(v) = G^K(v).$$

This shows that in a neighbourhood of u, G is piecewise smooth, and completes the proof.

□

When each f_j for $j \in J$ is an affine function, AIPCQ holds automatically, so we have the following:

Corollary 1. *Suppose that the functions f_j for $j \in J$ are affine and f is convex. Then G, the gradient function of the Moreau-Yosida regularization of f, is a piecewise affine function on the whole space $\Re^n$.*

Proof. Since in this case the AIPCQ holds everywhere, $A^K(x, q)$ is independent of x and q, and there are only finitely many choices for K from $\mathcal{B}(v)$, $v \in \Re^n$, it follows from (13a), (13b), (14) and (17) that G is a piecewise affine function on the whole space $\Re^n$.

□

Corollary 2. *Suppose that each f_j for $j \in J$ is twice continuously differentiable, f is convex, and the affine independence preserving constraint qualification holds at the proximal point $p(u)$. If $\mathcal{B}'(u)$ contains only one index set (in particular, the AICQ holds at $p(u)$, and for all $\alpha(u) \in \mathcal{M}(u), \alpha_j(u) > 0$, for each $j \in I(p(u))$), then G is continuously differentiable in a neighbourhood of u.*

Corollary 2 may be regarded as a generalization of the result obtained in [7] where f is the maximum of two C^2 convex functions f_1 and f_2 with the assumptions that $\alpha_1(u), \alpha_2(u) > 0$ and $\nabla f_1(p(u)) \neq \nabla f_2(p(u))$.

4. Conditions for positive definiteness

From Theorem 1 we know that if each f_j for $j \in J$ is twice continuously differentiable, f is convex, and the affine independence preserving constraint qualification holds at the proximal point $p(u)$, then about u there exist an open neighbourhood $N(u)$ and a family of finitely many continuously differentiable functions G^K, $K \in \mathcal{B}'(u)$ defined on $N(u)$ such that for each $v \in N(u)$

$$G(v) \in \{G^K(v) : \ K \in \mathcal{B}'(u)\}.$$

Hence, G is also semismooth at u [13]. Recall that given an initial point u^0, an approximate Newton method for solving a nonsmooth equation $G(u) = 0$ is:

$$u^{k+1} = u^k - V_k^{-1}\tilde{G}(u^k), \quad k = 0, 1, ...,$$

where V_k is an approximation of an element in $\partial G(u^k)$ (or $\partial_B G(u^k)$) and $\tilde{G}(u^k)$ is an approximation of $G(u^k)$. In order to obtain superlinear (quadratic) convergence of the approximate Newton method for minimizing F, as in [11] and [2], we need all the matrices in ∂G (or $\partial_B G$) to be positive definite. It was proved in [12] that all such matrices are positive definite if and only if f is strongly convex on a ball about $p(u)$. Here we give conditions depending on the Hessians of the functions f_j for $j \in J$, that guarantee that the elements of the generalized Jacobian of G are positive definite. In order to do this let

$$D(u) = \{\nabla G^K(u) : K \in \mathcal{B}(u)\}. \tag{22}$$

and

$$D'(u) = \{\nabla G^K(u) : K \in \mathcal{B}'(u)\}. \tag{23}$$

Theorem 2. *Suppose that each f_j for $j \in J$ is twice continuously differentiable, f is convex, and the affine independence preserving constraint qualification holds at the proximal point $p(u)$. If for each $K \in \mathcal{B}'(u)$, $C^K(u) := \sum_{j \in K} \alpha_j(u) \nabla^2 f_j(p(u))$ is positive definite on the subspace*

$$L^K(u) := \{d : \nabla \bar{f}_j(p(u))^T d = 0,\ j \in \bar{K}\},$$

where $\bar{K} = K \backslash \{i\}$ for some $i \in K$ and $L^K(u) = \Re^n$ if $|K| = 1$, then all $V \in D'(u)$ are positive definite. As a consequence, all matrices $V \in \partial G(u)$ are positive definite.

Proof. Since G is piecewise smooth in a neighbourhood of u, we can easily show that

$$\partial G(u) \subseteq \text{conv } D'(u). \tag{24}$$

So we only need to prove that all matrices $V \in D'(u)$ are positive definite. For each $V \in D'(u)$, there exists $K \in \mathcal{B}'(u)$ such that $V = \nabla G^K(u)$. From (17), we know that

$$\begin{aligned} V &= \nabla G^K(u) \\ &= M(I - \nabla x^K(u)). \end{aligned} \tag{25}$$

Let $\mathcal{V}$ be the (possibly vacuous) matrix with linearly independent columns $\nabla \bar{f}_j(p(u))$ for $j \in \bar{K}$. Let $\mathcal{U}$ be a matrix with linearly independent columns spanning the subspace $L^K(u)$. Then

$$\mathcal{V}^T \mathcal{U} = 0. \tag{26}$$

We will use B^K and C^K to represent $B^K(x^K(u), q^K(u))$ and $C^K(u)$, respectively. From (13a), (13b) and (16) we have

$$\begin{pmatrix} \mathcal{V}^T & 0 \\ B^K & M^{-1}\mathcal{V} \end{pmatrix} \begin{pmatrix} \nabla x^K(u) \\ \nabla q^K(u) \end{pmatrix} = \begin{pmatrix} 0 \\ I \end{pmatrix},$$

i.e.,

$$\mathcal{V}^T \nabla x^K(u) = 0 \tag{27}$$

and

$$B^K \nabla x^K(u) + M^{-1} \mathcal{V} \nabla q^K(u) = I. \tag{28}$$

Now, by the $\Re^n$ spanning property of $\mathcal{U}$ and $\mathcal{V}$, $\nabla x^K(u)$ can be decomposed into

$$\nabla x^K(u) = \mathcal{U} \nabla x^K_{\mathcal{U}}(u) + \mathcal{V} \nabla x^K_{\mathcal{V}}(u),$$

where $\nabla x_{\mathcal{U}}^K(u)$ is an $(n-|\bar{K}|)\times n$ matrix and $\nabla x_{\mathcal{V}}^K(u)$ is a $|\bar{K}|\times n$ matrix. Multiplying this expression on the left by $\mathcal{V}^T$ and using (26) and (27) gives

$$0=\mathcal{V}^T\mathcal{V}\nabla x_{\mathcal{V}}^K(u),$$

which implies

$$\nabla x_{\mathcal{V}}^K(u)=0.$$

So,

$$\nabla x^K(u)=\mathcal{U}\nabla x_{\mathcal{U}}^K(u). \tag{29}$$

Multiplying (28) on the left by M and using the definition of B^K in (13b), gives

$$(M+C^K)\nabla x^K(u)+\mathcal{V}\nabla q^K(u)=M.$$

Multiplying this expression on the left by $\mathcal{U}^T$ and using (26) and (29) gives

$$\mathcal{U}^T(M+C^K)\mathcal{U}\nabla x_{\mathcal{U}}^K(u)+0=\mathcal{U}^TM.$$

Therefore,

$$\nabla x^K(u)=\mathcal{U}\nabla x_{\mathcal{U}}^K(u)=\mathcal{U}\left(\mathcal{U}^T(M+C^K)\mathcal{U}\right)^{-1}\mathcal{U}^TM. \tag{30}$$

From Lemma 1 of [7], the positive definiteness of C^K implies that $\nabla x^K(u)$ has all of its eigenvalues in the interval $[0,1)$. So the positive definiteness of V follows easily from (25).

□

In practice, $I(p(u))$ may not be known. So next, in view of Remark 1, we consider a stronger condition, depending on $K\in\mathcal{B}(u)\supseteq\mathcal{B}'(u)$, that guarantees the nonsingularity of $\partial G(u)$ without assuming knowledge of $I(p(u))$.

Corollary 3. *Suppose that each f_j for $j\in J$ is twice continuously differentiable, f is convex, and the affine independence preserving constraint qualification holds at the proximal point $p(u)$. If all matrices $\nabla^2 f_j(p(u))$ for $j\in J(p(u))$ are positive semidefinite, and for each $K\in\mathcal{B}(u)$, $C^K(u):=\sum_{j\in K}\alpha_j(u)\nabla^2 f_j(p(u))$ is positive definite on the subspace*

$$L^K(u):=\{d:\nabla\bar{f}_j{}^T(p(u))d=0,\ j\in\bar{K}\},$$

where $\bar{K}=K\backslash\{i\}$ for some $i\in K$ and $L^K(u)=\Re^n$ if $|K|=1$, then all $V\in D(u)$ are positive definite. As a consequence, all $V\in\partial G(u)$ are positive definite.

5. Conclusions

In this paper we have discussed second-order properties of the Moreau-Yosida regularization of a piecewise C^2 convex function. This function is of a special form, but it is useful for gaining insight into what is needed for attempting to design better than linearly convergent algorithms for minimizing f based on approximate Newton and quasi-Newton methods for minimizing F. We believe that the results given here can lead to a deeper understanding of the Moreau-Yosida regularization of a general convex function.

Acknowledgements. This research was supported by the Australian Research Council. R. Mifflin was also supported by the National Science Foundation under grants DMS-9402018 and DMS-9703952.

The authors would like to thank two referees for their helpful comments.

References

1. Clarke, F.H. (1983): Optimization and nonsmooth analysis. Wiley, New York
2. Fukushima, M., Qi, L. (1996): A globally and superlinearly convergent algorithm for nonsmooth convex minimization. SIAM J. Optim. **6**, 1106–1120
3. Janin, R. (1984): Directional derivative of the marginal function in nonlinear programming. Math. Program. Study **21**, 110–126
4. Lemaréchal, C., Sagastizábal, C. (1994): An approach to variable metric methods. In: Henry, J., Yvon, J.-P., eds., Systems modelling and optimization, pp. 144-162. Lecture notes in control and information sciences 197. Springer, Berlin
5. Lemaréchal, C., Sagastizábal, C. (1997): Practical aspects of the Moreau-Yosida regularization I: Theoretical Preliminaries. SIAM J. Optim. **7**, 367–385
6. Madsen, K., Mifflin, R. (1984): Generalized gradients of piecewise smooth functions. Numerisk Institut Report No. 84-04. Technical University of Denmark, Lyngby, Denmark
7. Mifflin, R. (1996): A quasi-second-order proximal bundle algorithm. Math. Program. **73**, 51–72
8. Moreau, J.J. (1965): Proximite et dualite dans un espace hilbertien. Bull. Soc. Math. Fr. **93**, 273–299
9. Ortega, J.M., Rheinboldt, W.C. (1970): Iterative solutions of nonlinear equations in several variables. Academic Press, New York
10. Pang, J.S., Ralph, D. (1996): Piecewise smoothness, local invertibility, and parametric analysis of normal maps. Math. Oper. Res. **21**, 401–426
11. Qi, L. (1993): Convergence analysis of some algorithms for solving nonsmooth equations. Math. Oper. Res. **18**, 227–244
12. Qi, L. (1995): Second-order analysis of the Moreau-Yosida regularization of a convex function. Preprint, Revised version. School of Mathematics, the University of New South Wales, Sydney, Australia
13. Qi, L., Sun, J. (1993): A nonsmooth version of Newton's method. Math. Program. **58**, 353–367
14. Rockafellar, R.T. (1970): Convex analysis. Princeton University Press, Princeton
15. Rockafellar, R.T. (1985): Maximal monotone relations and the second derivatives on nonsmooth functions. Ann. Inst. Henri Poincaré, Anal. Non Linéaire **2**, 167–186
16. Sun, D., Han, J. (1997): On a conjecture in Moreau-Yosida regularization of a nonsmooth convex function. Chin. Sci. Bull. **42**, forthcoming (in Chinese)
17. Womersley, R.S. (1982): Optimality conditions for piecewise smooth functions. Math. Program. Study **17**, 13–27
18. Yosida, K. (1964): Functional analysis. Springer, Berlin