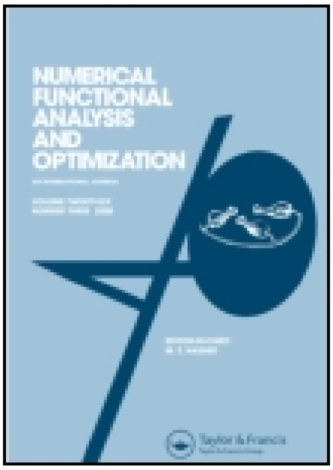

Numerical Functional Analysis and Optimization
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lnfa20

On Second-Order Properties of the Moreau-Yosida Regularization for Constrained Nonsmooth Convex Programs

Fanwen Meng [a] [b] [c] & Gongyun Zhao [b]
[a] Department of Mathematics, National University of Singapore, Singapore, Republic of Singapore
[b] School of Mathematics, University of Southampton, Highfield, Southampton, Great Britain
[c] School of Mathematics, University of Southampton, Highfield, S017 1BJ, Southampton, Great Britain
Published online: 31 Aug 2006.

To cite this article: Fanwen Meng & Gongyun Zhao (2004) On Second-Order Properties of the Moreau-Yosida Regularization for Constrained Nonsmooth Convex Programs, Numerical Functional Analysis and Optimization, 25:5-6, 515-529, DOI: 10.1081/NFA-200042235

To link to this article: http://dx.doi.org/10.1081/NFA-200042235

NUMERICAL FUNCTIONAL ANALYSIS AND OPTIMIZATION
Vol. 25, Nos. 5 & 6, pp. 515–529, 2004

On Second-Order Properties of the Moreau–Yosida Regularization for Constrained Nonsmooth Convex Programs

Fanwen Meng[1,2,*] and Gongyun Zhao[1]

[1]Department of Mathematics, National University of Singapore, Singapore, Republic of Singapore
[2]School of Mathematics, University of Southampton, Highfield, Southampton, Great Britain

ABSTRACT

In this paper, we attempt to investigate a class of constrained nonsmooth convex optimization problems, that is, piecewise C^2 convex objectives with smooth convex inequality constraints. By using the Moreau–Yosida regularization, we convert these problems into unconstrained smooth convex programs. Then, we investigate the second-order properties of the Moreau–Yosida regularization η. By introducing the (GAIPCQ) qualification, we show that the gradient of the regularized function η is piecewise smooth, thereby, semismooth.

Key Words: Moreau–Yosida regularization; Piecewise C^k functions; Semismooth functions.

AMS Subject Classification (1991): 90C25; 65K10; 52A41.

*Correspondence: Fanwen Meng, School of Mathematics, University of Southampton, Highfield, Southampton, S017 1BJ, Great Britain; E-mail: F.Meng@soton.ac.uk.

DOI: 10.1081/NFA-200042235
0163-0563 (Print); 1532-2467 (Online)

1. INTRODUCTION

In this paper, we consider the following constrained convex programming:

$$\min f_0(x) \quad \text{s.t. } f_i(x) \leq 0, \quad i \in I = \{1, 2, \ldots, m\}, \tag{1}$$

where f_0, f_i, $i \in I$, are finite, convex on R^n and f_0 is possibly nonsmooth.

For nonsmooth programs, many approaches have been presented so far and they are often restricted to the convex unconstrained case. In general, the various approaches are based on combinations of the following three methods: (i) subgradient methods (e.g., Ermoliev et al., 1997; Poljak, 1967); (ii) bundle techniques (e.g., Kiwiel, 1990; Lemaréchal et al., 1981, 1995); (iii) Moreau–Yosida regularization (e.g., Fukushima and Qi, 1996; Lemaréchal and Sagastizábal, 1997a,b; Qi and Chen, 1997). In this paper, we would like to investigate some properties of the Moreau–Yosida regularization. For any proper, closed convex function $f : R^n \to R \cup \{+\infty\}$, the Moreau (1965) and Yosida (1964) regularization of f, $\eta : R^n \to R$, is defined by

$$\eta(x) = \min_{y \in R^n} \left\{ f(y) + \frac{1}{2}\|y - x\|_M^2 \right\}, \tag{2}$$

where M is a symmetric positive definite $n \times n$ matrix and $\|z\|_M^2 = z^T M z$ for any $z \in R^n$. Let $p(x)$ denote the unique solution of (2). The motivation of the study of Moreau–Yosida regularization is due to the fact that

$$\min\{f(x) \mid x \in R^n\}, \tag{3}$$

is equivalent to

$$\min\{\eta(x) \mid x \in R^n\}. \tag{4}$$

More precisely, the following statements are equivalent:

(i) x minimizes f.
(ii) $p(x) = x$.
(iii) $\nabla\eta(x) = 0$.
(iv) x minimizes η.
(v) $f(p(x)) = f(x)$.
(vi) $\eta(x) = f(x)$.

It is easy to see that η is finite everywhere and convex. A nice property of the Moreau–Yosida regularization is that η is continuously differentiable and its gradient

$$g(x) \equiv \nabla\eta(x) = M(x - p(x)), \tag{5}$$

is Lipschitz continuous. For more details about Moreau–Yosida regularization, the readers may be referred to Hiriart-Urruty and Lemaréchal (1993) and Rockafellar (1970).

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

In many situations, besides the first order property mentioned above, we are also interested in the second order property of the Moreau–Yosida regularization, e.g., when we wish to use Newton's method to minimize $\eta(x)$. In most circumstances, due to the nondifferentiability of f, the second order derivative of η does not exist. Thus we need to investigate a weaker second order property, namely, the semismoothness of $g(=\nabla\eta)$. Under the condition of semismoothness, Qi and Sun (1993) derived the superlinear (quadratic) convergence of a nonsmooth version of Newton's method for solving a nonsmooth equation. Also, by virtue of the semismoothness of g, Fukushima and Qi (1996) have shown that the superlinear convergence can be guaranteed by using approximate solutions of the problem (2) to construct search directions for minimizing η. In addition, the concept of semismoothness plays a central role in the development of optimality conditions in nonsmooth analysis as well, e.g., Chaney (1988, 1989, 1990).

Recently, several authors have investigated the semismoothness of g for some classes of f, for example, see Lemaréchal and Sagastizábal (1994, 1997b), Meng and Hao (200l), Mifflin et al. (1999), Qi (1995), Rockafellar (1985), Sun and Han (1997), Zhao and Meng (2000). Qi (1995) first conjectured that g is semismooth under a regularity condition if f is the maximum of several twice continuously differentiable convex functions. Later, Sun and Han (1997) gave a proof to this conjecture under a constant rank constraint qualification (CRCQ) for the case where $M = (1/\lambda)I$ and λ is a positive constant. In Meng and Hao (200l), the authors investigated the unconstrained case of problem (1), where f_0 is assumed to be piecewise C^2. By introducing the (SCRCQ) qualification, which is weaker than (CRCQ), they derived the same result for g as Sun and Han (1997). In addition, Mifflin et al. (1999) investigated the same problem as in Meng and Hao (200l). By introducing the (AIPCQ) qualification, which is weaker than (CRCQ) and (SCRCQ), they derived the same result for g as Sun and Han (1997). Recently, Zhao and Meng (2000) studied the Lagrangian-dual function of certain problems. Specifically, there are two cases under consideration: (i) linear objectives with either affine inequality or strictly convex inequality constraints; (ii) strictly convex objectives with general convex inequality and affine equality constraints. After deriving the piecewise C^2-smoothness of the Lagrangian-dual function, they showed that the gradient of the Moreau–Yosida regularization of this function is piecewise smooth, thereby, semismooth, under the (AIPCQ) qualification.

Notice that all papers mentioned above mainly study the Moreau–Yosida regularization of a continuous (possibly nondifferentiable) function. And, their results are applicable to the unconstrained optimization problem. In this paper, we consider the constrained problem (1) instead. We should point out that the smooth problem

$$\min \eta_0(x) \quad \text{s.t. } f_i(x) \le 0, \quad i \in I, \tag{6}$$

where η_0 is the Moreau–Yosida regularization of f_0, is not equivalent to (1), because η_0 coincides with f_0 only on the set of their unconstrained minima, which may not be the solutions of the constrained minimization problems. Therefore, although the results derived in Mifflin et al. (1999) can be easily extended to (6), but they are not applicable to (1). Due to this observation, we consider the following generalized function:

$$f(x) = f_0(x) + \delta(x \mid D), \tag{7}$$

where $D := \{x \in R^n : f_i(x) \leq 0, i \in I\}$ and $\delta(x \mid D)$ is the usual indicator function on the convex set D (see, Rockafellar, 1970). Then, problem (1) can be written as the following unconstrained problem:

$$\min\{f(x) \mid x \in R^n\}. \tag{8}$$

Problem (8) then is equivalent to

$$\min\{\eta(x) \mid x \in R^n\}, \tag{9}$$

where

$$\eta(x) = \min_{y \in R^n} \left\{ f(y) + \frac{1}{2}\|y - x\|_M^2 \right\}. \tag{10}$$

Throughout this paper, we assume that for problem (1), f_0 is piecewise C^2 and f_i are twice continuously differentiable. We aim to investigate the second-order properties of the regularized function η in (10). The analysis presented in this paper is related to the results in Mifflin et al. (1999), however, there are several fundamental distinctions. Firstly, f defined in form of (7) is a generalized function (indeed, f is not even finite or continuous everywhere on R^n.), whereas, in Mifflin et al. (1999), it was required to be finite everywhere, convex and continuous on the whole space. Actually, the latter condition is a fundamental assumption in many publications regarding to the Moreau–Yosida regularization, e.g., see Lemaréchal and Sagastizábal (1997a), Mifflin et al. (1999), Qi (1995), Zhao and Meng (2000). Secondly, although the feasible region D is defined by some smooth inequalities, but the geometric figure of D is clearly nonsmooth, on which the indicator function is defined. In this paper, we attempt to cope with this two kinds of nonsmooth situations simultaneously. To this end, we introduce a new qualification, called (GAIPCQ). Under this condition, we show that the gradient of the Moreau–Yosida regularization of f is piecewise smooth. Thereby, the gradient of η is semismooth.

The rest of this paper is organized as follows. In Sec. 2, some basic notions and properties are collected. Section 3 studies the piecewise smoothness and semismoothness of the gradient g. Section 4 concludes.

2. DEFINITIONS AND BASIC PROPERTIES

In this section, we recall some concepts, such as semismoothness and piecewise C^k-ness, which are basic elements of this paper. In addition, we introduce a new constraint qualification, (GAIPCQ), which will be used in next section.

It is well known that η is a continuously differentiable convex function defined on R^n not only for the specific type function f defined in (7), but even for general

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

proper, closed, nondifferentiable convex functions as well. In order to use Newton's method or modified Newton's methods, it is important to study the Hessian of η, i.e., the Jacobian of g. It is also known that g is globally Lipschitz continuous with modulus $\|M\|$. However, g may not be differentiable. To extend the definition of Jacobian to certain classes of nonsmooth functions, Qi and Sun (1993) introduced the definition of semismoothness for vector-valued functions below:

Definition 2.1. Suppose $F : R^n \to R^m$ is a locally Lipschitzian vector-valued function, we say that F is *semismooth* at x if

$$\lim_{\substack{V\in\partial F(x+th'),\\ zh'\to h, t\downarrow 0}} \{Vh'\}, \tag{11}$$

exists for any $h \in R^n$, where $\partial F(y)$ denotes the generalized Jacobian of F at y defined by Clarke (1983).

The semismoothness of the gradient g is a key condition in the analysis of the superlinear convergence of an approximate Newton method proposed by Fukushima and Qi (1996). In fact, semismoothness is a local property for nonsmooth functions. A direct verification of semismoothness is difficult. Fortunately, according to Qi and Sun (1993), it is known that piecewise smooth functions are semismooth functions. Hence, one needs to study the piecewise smoothness of g for some class of convex problems. Now, we define the piecewise smoothness for vector-valued functions (see, Pang and Ralph, 1996).

Definition 2.2. A continuous function $\psi : R^n \to R^m$ is said to be *piecewise* C^k on a set $W \subseteq R^n$ if there exist a finite number of functions $\psi_1, \psi_2, \ldots, \psi_q$ such that each $\psi_j \in C^k(U)$ for some open set U containing W, and $\psi(x) \in \{\psi_1(x), \ldots, \psi_q(x)\}$ for any $x \in W$.

We refer to $\{\psi_j\}_{j\in I_q}$ as a *representation* of ψ on W, where $I_q := \{1, 2, \ldots, q\}$.

Remark 2.1. If $m = 1$, then Definition 2.2 is referred to the notion of piecewise C^k-ness for real-valued functions. In addition, the domain W in above definition is a general set in R^n. In particular, when W is open, then all pieces ψ_j above could be simply required to be C^k on $W = U$. Actually, in the following context, we mainly pay attention to the case where ψ is piecewise C^k on some open set in R^n.

In the following context, f_0 is always assumed to be convex. Let f_0 be piecewise C^1 on $W \subseteq R^n$ with a representation $\{h_j\}_{j\in\widehat{J}}$. For any $j \in \widehat{J}$, define $D_j := \{x \in W : h_j(x) = f_0(x)\}$. We define the active index set at x by

$$J(x) = \{j \in \widehat{J} : x \in D_j\}, \quad x \in W. \tag{12}$$

And, let

$$K(x) = \{j \in \widehat{J} : x \in \mathrm{cl}\,(\mathrm{int}\, D_j)\}, \quad x \in W. \tag{13}$$

We then obtain the following result regarding to the subdifferential of f_0.

Proposition 2.1 (Pang and Ralph, 1996). *Let f_0 be a convex real-valued function on R^n. Suppose that f_0 is piecewise C^1 on an open set $W \subseteq R^n$ with a representation $\{h_j\}_{j \in \hat{J}}$. Then, for every $x \in W$,*

$$\partial f_0(x) = \operatorname{conv}\{\nabla h_j(x) : j \in K(x)\}. \tag{14}$$

where conv *S denotes the convex hull of the set S.*

For the piecewise C^1 function f_0 above, it is easy to see that $K(x) \subseteq J(x)$ for any $x \in W$. Hence,

$$\partial f_0(x) = \operatorname{conv}\{\nabla h_j(x) : j \in K(x)\} \subseteq \operatorname{conv}\{\nabla h_j(x) : j \in J(x)\}. \tag{15}$$

Moreover, if f_0 above is piecewise C^2 on W, by virtue of Lemma 2 in Mifflin et al. (1999), we then have

$$\nabla^2 h_j(x) \succeq 0, \quad j \in K(x), \ \ x \in W, \tag{16}$$

where $A \succeq 0$ means the matrix A is positive semidefinite.

Now, let's define the set of indices of active constraints at any point $x \in D$ by

$$I(x) := \{i \in I \mid f_i(x) = 0\}. \tag{17}$$

In order to study the piecewise smoothness of g in (5), we introduce the following constraint qualification.

Definition 2.3. *Generalized Affine Independence Preserving Constraint Qualification* (GAIPCQ) is said to hold for f_0 and f_i, $i \in I$, at x, if for every subset $J \times K \subseteq J(x) \times I(x)$ for which there exists a sequence $\{x^k\}$ with $\{x^k\} \to x$, $J \times K \subseteq J(x^k) \times I(x^k)$ and the vectors

$$\left\{ \begin{pmatrix} \nabla h_j(x^k) \\ 1 \end{pmatrix}, \begin{pmatrix} \nabla f_i(x^k) \\ 0 \end{pmatrix} : j \in J, \ \ i \in K \right\}, \tag{18}$$

being linearly independent, it follows that the vectors

$$\left\{ \begin{pmatrix} \nabla h_j(x) \\ 1 \end{pmatrix}, \begin{pmatrix} \nabla f_i(x) \\ 0 \end{pmatrix} : j \in J, \ \ i \in K \right\}, \tag{19}$$

are linearly independent.

Remark 2.2. For problem (8) when f_0 is piecewise C^2, in order to investigate the smoothness of g, obviously, we shall need the information about the constraints f_i together with pieces h_j of f_0. Thus, (GAIPCQ) above involves both pieces h_j (i.e., f_0) and f_i. On the other hand, if problem (1) is unconstrained and f_0 is piecewise C^2 on R^n, then $f = f_0$, which implies that f is piecewise C^2 on the whole R^n as well. Evidently, in this case problem (8) becomes to problem (1). This is the exact situation studied in Mifflin et al. (1999). In this case, the terms concerning I and $I(x)$ in Definition 2.3 will vanish automatically, and (GAIPCQ) is reduced to the (AIPCQ) qualification introduced in Mifflin et al. (1999).

3. SMOOTHNESS OF g FOR PIECEWISE C^2 OBJECTIVES

In this section, we shall investigate the piecewise smoothness of the gradient of the Moreau–Yosida regularized function η under the (GAIPCQ) qualification. Then, according to Qi and Sun (1993), we can derive the semismoothness of $g(=\nabla\eta)$. In the rest of this paper, we make the following assumptions:

Assumption 3.1. $\emptyset \neq D \subseteq R^n$.

Assumption 3.2. f_0 is convex, piecewise C^2 on R^n with a representation $\{h_j\}_{j\in\hat{J}}$.

Assumption 3.3. f_i is convex, twice continuously differentiable on R^n for all $i \in I$.

For the generalized function f in (7), we can easily see that its effective domain is exactly the feasible region D. Then, for any $x \in R^n$, due to the closeness of D, it follows that the proximal solution $p(x)$ lies in D, hence, $\eta(x)$ is finite. Thereby, the effective domain of η is the whole n-dimensional space.

Now, for $x \in R^n$, since $p(x)$ minimizes $f(y) + \frac{1}{2}\|y - x\|_M^2$ on R^n, equivalently, $p(x)$ minimizes $f_0(y) + \frac{1}{2}\|y - x\|_M^2$ over D, then there exists a multiplier $(\lambda(x), \mu(x)) \in R^{|J(p(x))|\times|I(p(x))|}$ such that

$$
\begin{aligned}
& g(x) = M(x - p(x)) = \sum_{j\in J(p(x))} \lambda_j(x)\nabla h_j(p(x)) + \sum_{i\in I(p(x))} \mu_i(x)\nabla f_i(p(x)),\\
& \lambda_j(x) \geq 0, \ \ \forall j \in J(p(x)),\\
& \sum_{j\in J(p(x))} \lambda_j(x) = 1, \quad \mu_i(x) \geq 0, \ \ \forall i \in I(p(x)).
\end{aligned}
\tag{20}
$$

For a nonnegative vector $a \in R^m$, we shall let $\operatorname{supp}(a)$, called the *support* of a, be the subset of $\{1, \ldots, m\}$ consisting of the indices i for which $a_i > 0$. We define the following sets:

$$
\mathcal{M}(x) = \{(\lambda(x), \mu(x)) \in R^{|J(p(x))|\times|I(p(x))|} \mid (\lambda(x), \mu(x)) \text{ satisfies (20)}\}, \tag{21}
$$

$$
\begin{aligned}
\mathcal{B}(x) = \Bigg\{ & J \times K \subseteq J(p(x)) \times I(p(x)) \mid \text{there exists } (\lambda(x), \mu(x)) \in \mathcal{M}(x) \text{ such}\\
& \text{that } \operatorname{supp}(\lambda(x)) \subseteq J \subseteq J(p(x)),\ \operatorname{supp}(\mu(x)) \subseteq K \subseteq I(p(x)),\ K \neq \emptyset,\\
& \text{and the vectors } \left\{ \begin{pmatrix} \nabla h_j(p(x)) \\ 1 \end{pmatrix}, \begin{pmatrix} \nabla f_i(p(x)) \\ 0 \end{pmatrix} : j \in J, i \in K \right\}\\
& \text{are linearly independent.} \Bigg\},
\end{aligned}
\tag{22}
$$

and

$$
\widehat{\mathcal{B}}(x) = \{J \times K \in \mathcal{B}(x) \mid J \subseteq K(p(x))\}. \tag{23}
$$

If $I(p(x)) = \emptyset$, i.e., $|I(p(x))| = 0$, then all terms with respect to $I(p(x))$, $\mu(x)$ and K in (20)–(23) would vanish automatically.

If $I(p(x)) \neq \emptyset$, namely, the optimal solution $p(x)$ lies on the boundary of D, as to the corresponding Lagrangian multipliers $\mu(x)$ in (20), we make the following non-degenerate assumption:

Assumption 3.4. For any $(\lambda(x), \mu(x)) \in \mathscr{M}(x)$, $\mu(x) \neq 0$.

Lemma 3.1. *Let $\mathscr{B}(x)$ be defined in* (22), *then $\mathscr{B}(x) \neq \emptyset$ for any $x \in R^n$.*

Proof. It is easy to verify that $\mathscr{M}(x)$ in (21) is a convex set. It's obvious that the result is true if $|I(p(x))| = 0$.

Next, we consider the case where $|I(p(x))| \geq 1$. According to the convexity of $\mathscr{M}(x)$ and Assumption 3.4, it follows that there exists a vertex of $\mathscr{M}(x)$, say, $\alpha(x) = (\alpha_\lambda(x), \alpha_\mu(x))$, such that $\alpha_\mu(x) \neq 0$. Let $J = \mathrm{supp}(\alpha_\lambda(x))$ and $K = \mathrm{supp}(\alpha_\mu(x))$, then we have $J \subseteq J(p(x))$ and $K \subseteq I(p(x))$. Without loss of generality, we assume $J = \{1, 2, \ldots, s\}$, $K = \{1, \ldots, t\}$. In what follows, we need to show the vectors

$$\left\{\begin{pmatrix}\nabla h_1(p(x))\\ 1\end{pmatrix}, \ldots, \begin{pmatrix}\nabla h_s(p(x))\\ 1\end{pmatrix}, \begin{pmatrix}\nabla f_1(p(x))\\ 0\end{pmatrix}, \ldots, \begin{pmatrix}\nabla f_t(p(x))\\ 0\end{pmatrix}\right\}, \tag{24}$$

are linearly independent. If this desired result is not true, then there exist scalars $k_1, k_2, \ldots, k_s$ and $l_1, l_2, \ldots, l_t$ which are not all zero satisfying

$$\sum_{j=1}^{s} k_j \begin{pmatrix}\nabla h_j(p(x))\\ 1\end{pmatrix} + \sum_{i=1}^{t} l_i \begin{pmatrix}\nabla f_i(p(x))\\ 0\end{pmatrix} = 0, \tag{25}$$

namely,

$$\begin{cases} \sum_{j=1}^{s} k_j \nabla h_j(p(x)) + \sum_{i=1}^{t} l_t \nabla f_i(p(x)) = 0, \\ k_1 + k_2 + \cdots + k_s = 0. \end{cases} \tag{26}$$

Let $\beta = (\beta_\lambda, \beta_\mu)$ with $\beta_\lambda = (k_1, \ldots, k_s, 0, \ldots, 0) \in R^{|J(p(x))|}$, $\beta_\mu = (l_1, \ldots, l_t, 0, \ldots, 0) \in R^{|I(p(x))|}$, then for any sufficiently small $\varepsilon > 0$, e.g., taking $\varepsilon = \min\{\alpha_{\lambda_j}(x)/|k_j|, \alpha_{\mu_i}(x)/|l_i| : k_j \neq 0, l_i \neq 0, j \in J, i \in K.\}$, we have

$$\alpha(x) \pm \varepsilon\beta \geq 0, \tag{27}$$

and

$$\alpha_k(v) \pm \varepsilon\beta_k = 0, \quad \forall k \in (J(p(x))\backslash J) \cup (I(p(x))\backslash K). \tag{28}$$

Hence, it follows from (26) that

$$\begin{aligned}\sum_{j\in J(p(x))} (\alpha_j(x) \pm \varepsilon\beta_j) = 1 \pm \varepsilon \sum_{j\in J(p(x))} \beta_j = 1 \pm \varepsilon \sum_{j\in J} \beta_j \\ = 1 \pm \varepsilon(k_1 + \cdots + k_s) = 1 \pm 0 = 1.\end{aligned} \tag{29}$$

In addition, by (21) and (26), we have

$$\sum_{j\in J(p(x))}(\alpha_j(x)\pm\varepsilon\beta_j)\nabla h_j(p(x))+\sum_{i\in I(p(x))}(\alpha_i(x)\pm\varepsilon\beta_i)\nabla f_i(p(x))$$
$$=g(x)\pm\varepsilon\left(\sum_{j=1}^{s}k_j\nabla h_j(p(x))+\sum_{i=1}^{t}l_i\nabla f_i(p(x))\right)=g(x). \tag{30}$$

Therefore, by (27)–(30), it follows that $\alpha(x)\pm\varepsilon\beta\in\mathscr{M}(x)$. Furthermore, we have $\alpha(x)=(\alpha(x)+\varepsilon\beta)/2+(\alpha(x)-\varepsilon\beta)/2$, which leads to a contradiction with that $\alpha(x)$ is the vertex of $\mathscr{M}(x)$. Hence, (24) are linearly independent, thereby, $J\times K\in\mathscr{B}(x)$. Thus, $\mathscr{B}(x)\neq\emptyset$. □

Now, we establish the following lemma, which is very important in analysis.

Lemma 3.2. *Let $\mathscr{B}(x)$ be defined in (22). Suppose that (GAIPCQ) is satisfied at $p(x)$. Then, there exists an open neighborhood $\overline{\mathscr{N}}(x)$ such that $\mathscr{B}(y)\subseteq\mathscr{B}(x)$ for any $y\in\overline{\mathscr{N}}(x)$.*

Proof. In what follows, we consider the following two cases: $|I(p(x))|=0$, and $|I(p(x))|\geq 1$, respectively.

Case i $|I(p(x))|=0$. In this case, it is easy to see that $p(x)$ lies in the relative interior of D. By Theorem XV 4.1.4 of Hiriart-Urruty and Lemaréchal (1993), we have $p(\cdot)$ is continuous on R^n. So, it follows that $p(y)\in \operatorname{ri} D$ for y close to x, thereby, $I(p(y))=\emptyset$. Thus, for any y close to x, the terms regarding the index set K in $\mathscr{B}(y)$ are automatically vacuous. Hence, we only need to verify that the desired result concerning J is true. However, in this case, according to Remark 2.2, (GAIPCQ) is reduced to (AIPCQ). So, the desired result is valid with the help of Lemma 3 of Mifflin et al. (1999).

Case ii $|I(p(x))|\geq 1$. By the continuity of $p(\cdot)$ and (12), for any y close to x, we have

$$J(p(y))\subseteq J(p(x)). \tag{31}$$

Since $f_i(p(x))<0$ for any $i\in I\backslash I(p(x))$, then by virtue of the continuity of $p(\cdot)$ again, there exists an open neighborhood $\mathscr{N}_1(x)$ around x such that for any $y\in\mathscr{N}_1(x)$, $f_i(p(y))<0$, $i\in I\backslash I(p(x))$. Hence, $I\backslash I(p(x))\subseteq I\backslash I(p(y))$. So, it follows that

$$I(p(y))\subseteq I(p(x)) \tag{32}$$

for any $y\in\mathscr{N}_1(x)$. Now, suppose on the contrary that the proposed result is not true. Hence, noticing that $J(p(x))$ and $I(p(x))$ are finite index sets, then there exists a sequence $\{x^k\}$ tending to x such that for all k, there is an index set $J\times K\in\mathscr{B}(x^k)\backslash\mathscr{B}(x)$. So, for each k, the vectors

$$\left\{\begin{pmatrix}\nabla h_j(p(x^k))\\ 1\end{pmatrix},\begin{pmatrix}\nabla f_i(p(x^k))\\ 0\end{pmatrix}: j\in J, i\in K\right\} \tag{33}$$

are linearly independent and there exists $(\lambda(x^k), \mu(x^k)) \in \mathscr{M}(x^k)$ such that $\mathrm{supp}(\lambda(x^k)) \subseteq J \subseteq J(p(x^k))$, $\mathrm{supp}(\mu(x^k)) \subseteq K \subseteq I(p(x^k))$, but $J \times K \notin \mathscr{B}(x)$. According to the (GAIPCQ), the vectors

$$\left\{ \begin{pmatrix} \nabla h_j(p(x)) \\ 1 \end{pmatrix}, \begin{pmatrix} \nabla f_i(p(x)) \\ 0 \end{pmatrix} : j \in J, i \in K \right\} \tag{34}$$

must be linearly independent. In addition, by (31) and (32), it follows that $J \times K \subseteq J(p(x)) \times I(p(x))$. Then, we get the following statement:

There doesn't exist $(\lambda(x), \mu(x)) \in \mathscr{M}(x)$ *such that* $\mathrm{supp}(\lambda(x)) \subseteq J \subseteq J(p(x))$,

and $\mathrm{supp}(\mu(x)) \subseteq K \subseteq I(p(x))$. (35)

Due to the boundedness of $\lambda_J(x^k)$, there exists a convergent subsequence of $\{\lambda_J(x^k)\}_{k=1}^{\infty}$. For convenience in description, we still use this same notation. Let's assume $\{\lambda_J(x^k)\}$ tends to $\overline{\lambda}_J$. Then, regarding to this convergent subsequence, it follows from (20) that

$$g(x^k) - \sum_{j \in J} \lambda_j(x^k) \nabla h_j(p(x^k)) = \sum_{i \in K} \mu_i(x^k) \nabla f_i(p(x^k)), \tag{36}$$

According to (33), we have $\nabla f_K(p(x^k))$ has full column rank. So, it follows from (36) that

$$\begin{aligned} \mu_K(x^k) = {} & \left[(\nabla f_K(p(x^k)))^T \nabla f_K(p(x^k)) \right]^{-1} (\nabla f_K(p(x^k)))^T \\ & \times \left[g(x^k) - \sum_{j \in J} \lambda_j(x^k) \nabla h_j(p(x^k)) \right]. \end{aligned} \tag{37}$$

By the continuity of $p(\cdot)$ and $\nabla f_i(\cdot)$, as $k \to \infty$, we have $(\nabla f_K(p(x^k)))^T \nabla f_K(p(x^k)) \to (\nabla f_K(p(x)))^T \nabla f_K(p(x))$. In addition, from (34) we have $\nabla f_K(p(x))$ has full column rank. Thereby, the inverse of $(\nabla f_K(p(x)))^T \nabla f_K(p(x))$ does exist. Hence, as $k \to \infty$, the right hand side of (37) must tend to

$$\left[(\nabla f_K(p(x)))^T \nabla f_K(p(x)) \right]^{-1} (\nabla f_K(p(x)))^T \left[g(x) - \sum_{j \in J} \bar{\lambda}_j \nabla h_j(p(x)) \right], \tag{38}$$

which is denoted by $\bar{\mu}_K$. So, we have

$$g(x) - \sum_{j \in J} \bar{\lambda}_j \nabla h_j(p(x)) = \sum_{i \in K} \bar{\mu}_i \nabla f_i(p(x)), \tag{39}$$

Now, by setting $\lambda_i(x) = \bar{\lambda}_i(x)$ if $i \in J$, $\lambda_i(x) = 0$ for $i \in J(p(x)) \backslash J$, and, $\mu_i(x) = \bar{\mu}_i(x)$ if $i \in K$, $\mu_i(x) = 0$ for $i \in I(p(x)) \backslash K$, we have $(\lambda(x), \mu(x)) \in \mathscr{M}(x)$ satisfying $\mathrm{supp}(\lambda(x)) \subseteq J$, $\mathrm{supp}(\mu(x)) \subseteq K$, which leads to a contradiction with (35).

Therefore, by Case i and Case ii, we have shown the proposed conclusion. □

Corollary 3.1. *Let $\widehat{\mathscr{B}}(x)$ be defined in* (23). *Suppose that* (*GAIPCQ*) *is satisfied at* $p(x)$. *Then, there exists an open neighborhood $\widehat{\mathscr{N}}(x)$ such that $\widehat{\mathscr{B}}(y) \subseteq \widehat{\mathscr{B}}(x)$ for any $y \in \widehat{\mathscr{N}}(x)$.*

Proof. By virtue of Proposition 2.1, Lemma 3.1 and (15), it is clear that $\widehat{\mathscr{B}}(x) \neq \emptyset$ for any $x \in R^n$. Noticing the continuity of $p(\cdot)$, it follows that $K(p(y)) \subseteq K(p(x))$ for y close to x. Hence, the result is valid by virtue of Lemma 3.2. $\square$

Now, for $x \in R^n$ and any $J \times K \subseteq \widehat{\mathscr{B}}(x)$, choose some $k \in J$ and define $\overline{J} = J \backslash \{k\}$. We consider the following mapping:

$$H_{J\times K}(y, z, q, w) = \begin{pmatrix} \left(h_j(y) - h_k(y)\right)_{j\in\overline{J}} \\ (f_i(y))_{i\in K} \\ M^{-1}\left(\sum_{j\in\overline{J}} z_j \nabla h_j(y) + \left(1 - \sum_{j\in\overline{J}} z_j\right)\nabla h_k(y) + \sum_{i\in K} q_i \nabla f_i(y)\right) + y - w \end{pmatrix}, \tag{40}$$

where $w \in R^n$, $(y, z, q) \in R^n \times R^{|\overline{J}|} \times R^{|K|}$. Let $(y^0, z^0, q^0, w^0) = (p(x), \lambda_{\overline{J}}(x), \mu_K(x), x)$, then we have $H_{J\times K}(y^0, z^0, q^0, w^0) = 0$. For convenience, set $\bar{h}_j(y) = h_j(y) - h_k(y)$, $j \in \overline{J}$.

Then, the matrix of partial derivatives of $H_{J\times K}(y, z, q, w)$ with respect to (y, z, q) is

$$J_{(y,z,q)} H_{J\times K}(y, z, q, w) = \begin{pmatrix} \nabla \bar{h}_{\overline{J}}(y)^T & 0 & 0 \\ \nabla f_K(y)^T & 0 & 0 \\ B_{J\times K}(y, z, q) & M^{-1}\nabla \bar{h}_{\overline{J}}(y) & M^{-1}\nabla f_K(y) \end{pmatrix}, \tag{41}$$

where for $i \in K$, $\nabla f_i(y)$ is the ith column of $\nabla f_K(y)$ and $\nabla \bar{h}_j(y)$ is the jth column of $\nabla \bar{h}_{\overline{J}}(y)$ for any $j \in \overline{J}$, and

$$B_{J\times K}(y, z, q) = I + M^{-1}\left(\sum_{j\in\overline{J}} z_j \nabla^2 h_j(y) + \left(1 - \sum_{j\in\overline{J}} z_j\right)\nabla^2 h_k(y) + \sum_{i\in K} q_i \nabla^2 f_i(y)\right). \tag{42}$$

When $|J| = 1$, then $\overline{J}$, the variable z, and the terms concerning them in above formulae are vacuous. Similarly, all terms regarding to K above will vanish when $|I(p(x))| = 0$. In this two cases, the mapping H in (40) and its corresponding Jacobian matrix would be described by simpler forms in some lower dimensional spaces, and the analysis becomes much easier. Now, we get the following result:

Lemma 3.3. *For any $J \times K \in \widehat{\mathscr{B}}(x)$, there exists an open neighborhood $S_{J\times K}$ of $w^0(= x)$ and an open neighborhood $Q_{J\times K}$ of $(y^0, z^0, q^0)(= (p(x), \lambda_J(x), \mu_K(x)))$ such that when $w \in \operatorname{cl} S_{J\times K}$, the equations $H_{J\times K}(y, z, q, w) = 0$ have a unique solution $(y(w), z(w), q(w))_{J\times K} \in \operatorname{cl} Q_{J\times K}$ where $z_{J\times K}(w)$ or $q_{J\times K}(w)$ is vacuous if $|J| = 1$ or $|K| = 0$, respectively. Furthermore, $(y(w), z(w), q(w))_{J\times K}$ is continuously differentiable on $S_{J\times K}$.*

Proof. When $|J| = 1$ or $|I(p(x))| = 0$, it is easy to see that the desired results are true. Next, we consider the case when $|J| > 1$ and $|I(p(x))| \neq 0$.

It follows from (16) that $\nabla^2 h_j(x) \succeq 0$ for any $j \in J$. Also, noticing $q^0 \geq 0$, $0 \leq z_j^0 \leq 1$, $0 \leq 1 - \sum_{j\in\overline{J}} z_j^0 \leq 1$ for any $j \in \overline{J}$, then there exists an open neighborhood of (y^0, z^0, q^0, w^0), say $N_{J\times K}(y^0, z^0, q^0, w^0)$, such that $B_{J\times K}(y, z, q)$ is positive definite on this neighborhood.

On the other hand, by the definition of $\widehat{\mathscr{B}}(x)$, we can easily know that the vectors

$$\{\nabla\bar{h}_j(y^0), \nabla f_i(y^0) : j \in \overline{J}, i \in K\} \tag{43}$$

are linearly independent. So, by the continuity of ∇h_j and ∇f_i, $j \in J$, $i \in K$, there exists an open neighborhood of (y^0, z^0, q^0, w^0), for convenience say $N_{J\times K}(y^0, z^0, q^0, w^0)$ as above, such that for any $(y, z, q, w) \in N_{J\times K}(y^0, z^0, q^0, w^0)$

$$\{\nabla\bar{h}_j(y), \nabla f_i(y) : j \in \overline{J}, i \in K\} \tag{44}$$

are linearly independent. Then, it follows that the matrix $\left(\nabla\bar{h}_{\overline{J}}(y), \nabla f_K(y)\right)$ has full column rank. Hence, we have $J_{(y,z,q)}H_{J\times K}(y, z, q, w)$ is nonsingular on $N_{J\times K}(y^0, z^0, q^0, w^0)$. Noticing the fact that $H_{J\times K}(y^0, z^0, q^0, w^0) = 0$ and by virtue of the Implicit Function Theorem, we derive the desired conclusion. $\square$

Since $J(p(x))$ and $I(p(x))$ are both finite index sets, according to Lemma 3.3, we derive the following finitely many functions, $g_{J\times K} : S_{J\times K} \to R^n$ defined by

$$g_{J\times K}(w) = M(w - y_{J\times K}(w)), \quad J \times K \in \widehat{\mathscr{B}}(x). \tag{45}$$

where $S_{J\times K}$, $y_{J\times K}(w)$ are defined in Lemma 3.3. Obviously, $g_{J\times K}(w)$ is continuously differentiable on $S_{J\times K}$. By using these functions and the (GAIPCQ) qualification, we derive the piecewise smoothness of g.

Proposition 3.1. *Suppose that (GAIPCQ) holds at $p(x)$. Then, there exists an open neighborhood $\mathscr{N}(x)$ around x such that g is piecewise smooth on $\mathscr{N}(x)$. In particular, g is semismooth at x.*

Proof. From (20) and Corollary 3.1, for any $w \in \widehat{\mathscr{N}}(x)$, it follows that there exists $J \times K \in \widehat{\mathscr{B}}(w) \subseteq \widehat{\mathscr{B}}(x)$ such that

$$\begin{aligned} &g(w) = M(w - p(w)) = \sum_{j\in J}\lambda_j(w)\nabla h_j(p(w)) + \sum_{i\in K}\mu_i(w)\nabla f_i(p(w)),\\ &\quad\lambda_j(w) \geq 0,\ \ \forall j \in J(p(w)),\\ &\sum_{j\in J(p(w))}\lambda_j(w) = 1, \quad \mu_i(w) \geq 0,\ \ i \in I(p(w)). \end{aligned} \tag{46}$$

Let

$$A(p(w)) = \begin{pmatrix} \nabla h_J(p(w)) & \nabla f_K(p(w)) \\ 1_J & 0_K \end{pmatrix}, \tag{47}$$

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

where 1_J, 0_K, denotes the $|J|$-dimensional row vector with all ones components, and the $|K|$-dimensional zero row vector, respectively. According to the definition of $\widehat{\mathcal{B}}(x)$, it follows that $A(p(w))$ has full column rank. Then, from (46), we get

$$\begin{pmatrix} \lambda_J(w) \\ \mu_K(w) \end{pmatrix} = \left(A(p(w))^T A(p(w))\right)^{-1} A(p(w))^T \begin{pmatrix} g(w) \\ 1 \end{pmatrix}, \quad w \in \widehat{\mathcal{N}}(x). \tag{48}$$

Since $p(\cdot)$, $\nabla h_j(\cdot)$ and $\nabla f_i(\cdot)$ are continuous, so we have $\lambda_J(w) \to \lambda_J(x)$ and $\mu_K(w) \to \mu_K(x)$ as $w \to x$. Thus, we can choose $\mathcal{N}(x) \subseteq \{\bigcap_{J\times K \in \widehat{\mathcal{B}}(x)} S_{J\times K}\} \cap \widehat{\mathcal{N}}(x)$ to be an open neighborhood of $w^0(=x)$ such that for any $w \in \mathcal{N}(x)$ and $J \times K \in \widehat{\mathcal{B}}(w)$, $(p(w), \lambda_J(w), \mu_K(w)) \in \mathrm{cl}\, Q_{J\times K}$, where $\lambda_J(w)$, $\mu_K(w)$ is vacuous when $|J| = 1$ or $|I(p(x))| = 0$, respectively. For any $w \in \mathcal{N}(x)$, there exists $J \times K \in \widehat{\mathcal{B}}(x)$ such that

$$H_{J\times K}(p(w), \lambda_J(w), \mu_K(w), w) = 0, \tag{49}$$

and $(p(w), \lambda_J(w), \mu_K(w)) \in \mathrm{cl} Q_{J\times K}$. So, by virtue of the Implicit Function Theorem, we have $(p(w), \lambda_J(w), \mu_K(w)) = (y_{J\times K}(w), z_J(w), q_K(w))$. Thereby, it follows that

$$g(w) = M(w - p(w)) = M(w - y_{J\times K}(w)) = g_{J\times K}(w), \tag{50}$$

where $g_{J\times K}(w)$ is defined in (45). Thus, for any $w \in \mathcal{N}(x)$, there exists at least one continuously differentiable function $g_{J\times K} : S_{J\times K} \to R^n$ such that $g(w) = g_{J\times K}(w)$ with $J \times K \in \widehat{\mathcal{B}}(x)$. Noticing that $\widehat{\mathcal{B}}(x)$ is a finite set, we have g is piecewise smooth on $\mathcal{N}(x)$ by Definition 2.2. Moreover, g is semismooth at x according to Qi and Sun (1993). □

Remark 3.1. In this section, we show that g is piecewise smooth around x, thereby, semismooth at x, for the problems with C^2 inequality constraints. Actually, if constraint functions f_i in (1) are piecewise C^2 convex on R^n, we can also analogously derive the same results without difficulty by following the idea of this paper. The description in this case should be much more complicated. However, it is not straightforward to extend these results to the general constraint problems with both equality and inequality constraints. In fact, the analysis would become quite difficult and complex in this situation since the methods used in this paper are perhaps not valid. One of the main obstacles is that at the KKT points the Lagrangian multipliers corresponding to the equality constraints are very hard to deal with.

4. CONCLUSION

The Moreau–Yosida regularization is a powerful tool for smoothing nondifferentiable functions, which is widely used in nonsmooth optimization. A significant feature is the semismoothness of the gradient of the Moreau–Yosida regularization. In this paper, we investigate the semismoothness of the gradient g of the Moreau–Yosida regularization of f, which plays a key role in the superlinear convergence analysis of approximate Newton methods. The function f is closely related to a constrained program with a convex, piecewise C^2 objective f_0 and inequality constraints. By introducing a new qualification, we derive the piecewise smoothness of the gradient g. We also believe that the results established in this paper could enrich the theory in nonsmooth optimization greatly. Actually, there are still a lot of questions left open. For instance, for general convex problems with both equality and inequality constraints, we have not a confirmative result for the semismoothness of the gradient g so far. It is a great challenge to answer these questions.

ACKNOWLEDGMENT

The research of the first author was supported by Grant R-146-000-036-592 of IHPC-CIM, National University of Singapore.

REFERENCES

Chaney, R. W. (1988). Second-order necessary conditions in semismooth optimization. *Math. Programming* 40(1):95–109.

Chaney, R. W. (1989). Optimality conditions for piecewise C^2 nonlinear programming. *J. Optim. Theory Anal.* 61(2):179–202.

Chaney, R. W. (1990). Piecewise C^k functions in nonsmooth analysis. *Nonlinear Anal. Theory, Methods Appl.* 15(7):649–660.

Clarke, F. H. (1983). *Optimization and Nonsmooth Analysis*. New York: John Wiley and Sons.

Ermoliev, Y. M., Kryazhimskii, A. V., Ruszczyński, A. (1997). Constraint aggregation principle in convex optimization. *Math. Programming* 76(3): 353–372.

Fukushima, M., Qi, L. (1996). A Global and superlinear convergent algorithm for nonsmooth convex minimization. *SIAM J. Optim.* 6(4):1106–1120.

Hiriart-Urruty, J. B., Lemaréchal, C. (1993). *Convex Analysis and Minimization Algorithms*. Berlin: Springer Verlag.

Kiwiel, K. (1990). Proximity control in bundle methods for convex nondifferentiable minimization. *Math. Programming* 46(1):105–122.

Lemaréchal, C., Sagastizábal, C. (1994). An approach to variable metric methods. In: Henry, J., Yvon, J.-P., eds. *Systems Modelling and Optimization*. Lecture notes in Control and Information Sciences 197. Berlin: Springer-Verlag, pp. 144–162.

Lemaréchal, C., Sagastizábal, C. (1997a). Variable metric bundle methods: From conceptual to implementable forms. *Math. Programming* 76(3):393–410.

Lemaréchal, C., Sagastizábal, C. (1997b). Practical aspects of the Moreau–Yosida regularization I: Theoretical preliminaries. *SIAM J. Optim.* 7(2):367–385.

Lemaréchal, C., Strodiot, J. J., Bihain, A. (1981). On a bundle method for nonsmooth optimization. In: Mangasarian, O. L., Meyer, R. R., Robinson, S. M., eds. *Nonlinear Programming* 4. New York: Academic Press, pp. 245–282.

Lemaréchal, C., Nemirovskii, A., Nesterov, Y. (1995). New variants of bundle methods. *Math. Programming* 69(1):111–147.

ORDER REPRINTS

Meng F. W., Hao Y. (2001). The property of piecewise smoothness of Moreau–Yosida approximation for a piecewise C^2 convex function. *Adv. Math. (Chinese)* 30(4):354–358.

Mifflin, R., Qi, L., Sun, D. (1999). Properties of the Moreau–Yosida regularization of a piecewise C^2 convex function. *Math. Programming* 84(2):269–281.

Moreau, J. J. (1965). Proximite et Dualite Dans un Espace Hilbertien. *Bulletin de la Societe Mathematique de France (French)* 93:273–299.

Pang, J. S., Ralph, D. (1996). Piecewise smoothness, local invertibility, and parametric analysis of normal maps. *Math. Oper. Res.* 21(2):401–426.

Poljak, B. T. (1967). A general method of solving extremum problems. *Dokl. Akad. Nauk SSSR.* 174:33–36.

Qi, L. (1995). Second-order analysis of the Moreau-Yosida regularization of a convex function. Preprint, Revised version, School of Mathematics, Sydney, Australia: The University of New South Wales.

Qi, L., Sun, J. (1993). A nonsmooth version of Newton's method. *Math. Programming* 58(3):353–367.

Qi, L., Chen, X. (1997). A preconditioning proximal Newton method for nondifferentiable convex optimization. *Math. Programming* 76(3):411–429.

Rockafellar, R. T. (1970). *Convex Analysis*. New Jersey: Princeton.

Rockafellar, R. T. (1985). Maximal monotone relations and the second derivatives on nonsmooth functions. *Ann. Inst. H. Poincaré Anal. Non Linéaire.* 2(3): 167–184.

Sun, D., Han, J. (1997). On a conjecture in Moreau–Yosida regularization of a nonsmooth convex function. *Chinese Sci. Bull.* 42(17):1140–1143.

Yosida, K. (1964). *Functional Analysis*. Berlin: Springer Verlag.

Zhao, G., Meng, F. (2000). On Piecewise smoothness of a lagrangian-dual function and its Moreau-Yosida regularization. Research Report, DAMTP of University of Cambridge.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016