On a conjecture in Moreau-Yosida approximation of a nonsmooth convex function

SUN Defeng and HAN Jiye

Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing 100080, China

Keywords: convex function, Moreau-Yosida approximation, piecewise smooth.

CONSIDER

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

where $f: \mathbb{R}^n \to [-\infty, +\infty]$ is an extended valued closed proper convex function. The Moreau-Yosida approximation F_λ of f is defined by

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

where λ is a positive parameter and $\| \cdot \|$ denotes the Euclidean norm.

From ref. [1] we know that F_λ is a differentiable convex function defined in the whole space of $\mathbb{R}^n$. The derivative of F_λ is

$$G_\lambda(x) \equiv \frac{1}{\lambda}(x - p_\lambda(x)) \in g(p_\lambda(x)), \tag{3}$$

where $g = \partial f$ is the subdifferential mapping of f and $p_\lambda(x)$ is the unique minimizer of eq. (2). An often discussed case is

$$f(x) = \max\{f_i(x): i \in J\}, \tag{4}$$

where J is a finite index set and f_i, $i \in J$, are proper convex functions. When f_i, $i \in J$, are linear functions, Qi[1)] proved that G_λ is piecewise affine, hence semismooth, under a regular

1) Qi, L., Second-order analysis of the Moreau-Yosida approximation of a convex function, *Mathematical Programming*, 1997 (to appear).

assumption. When f_i, $i \in J$, are twice continuously differentiable nonlinear convex functions, Qi[1)] posed a question: will G_λ be piecewise smooth[2], hence semismooth, under a similar regularity condition? We will give a positive answer to this question under a more relaxed condition here.

Define $J(p_\lambda(x)) = \{i \in J: f_i(p_\lambda(x)) = f(p_\lambda(x))\}$.

Constant Rank Constraint Qualification (CRCQ). CRCQ is said to hold at $p_\lambda(x)$ if there exists a neighborhood V of $p_\lambda(x)$ such that for every subset $K \subset J(p_\lambda(x))$, the family of the vectors

$$\left\{\begin{pmatrix} \nabla f_i(z) \\ 1 \end{pmatrix}: i \in K\right\} \tag{5}$$

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

Remark 1. CRCQ will hold if $|J(p_\lambda(x))| = 1$ or the linear independence constraint qualification (LICQ) holds. CRCQ holds automatically if all f_i, $i \in J$, are linear functions.

Let $\mathscr{M}(x)$ denote the set of all multipliers $\alpha(x)$ such that

$$\begin{cases} G_\lambda(x) = \dfrac{1}{\lambda}(x - p_\lambda(x)) = \displaystyle\sum_{i \in J(p_\lambda(x))} \alpha_i(x) \nabla f_i(p_\lambda(x)), \\ \alpha_i(x) \geqslant 0,\ j \in J;\ \alpha_i(x) = 0,\ i \in J \setminus J(p_\lambda(x)),\ \displaystyle\sum_{i \in J(p_\lambda(x))} \alpha_i(x) = 1. \end{cases} \tag{6}$$

For a nonnegative vector $d \in \mathbb{R}^{|J|}$, let supp(d) be the subset of $\{1, \cdots, |J|\}$ consisting of the indexes i for $d_i > 0$. Define the family $\mathscr{B}(x)$ of subsets of J as follows: $K \in \mathscr{B}(x)$ if and only if supp$(\alpha(x)) \subseteq K \subseteq J(p_\lambda(x))$ for some $\alpha(x) \in \mathscr{M}(x)$ and the vectors

$$\left\{\begin{pmatrix} \nabla f_i(p_\lambda(x)) \\ 1 \end{pmatrix}: i \in K\right\}$$

are linearly independent. This family $\mathscr{B}(x)$ is nonempty because $\mathscr{M}(x)$ has an extreme point which easily yields a desired index set K with the stated properties.

Theorem 1. *Suppose that f_i, $i \in J$, are twice continuously differentiable convex functions and CRCQ holds at $p_\lambda(x)$. Then there exists an open neighborhood N of x such that $G_\lambda(x)$, the derivative of the Moreau-Yosida approximation of f, is piecewise smooth in N.*

Proof. First we can prove that there exists a neighborhood U of x such that

$$\mathscr{B}(y) \subseteq \mathscr{B}(x) \text{ for all } y \in U. \tag{7}$$

For y close to x, $J(p_\lambda(y)) \subseteq J(p_\lambda(x))$. Hence if $|J(p_\lambda(x))| = 1$, then $G_\lambda(x)$ is continuously differentiable in a neighborhood of x. In the following, we assume that $|J(p_\lambda(x))| > 1$. By eqs. (6) and (7), for any $y \in U$ and $K \in \mathscr{B}(y)$, there exist $\alpha^K(y) \in \mathscr{M}(y)$, supp$(\alpha^K(y)) \subseteq K$, such that

$$\begin{cases} G_\lambda(y) = \dfrac{1}{\lambda}(y - p_\lambda(y)) = \displaystyle\sum_{i \in K} \alpha_i^K(y) \nabla f_i(p_\lambda(y)), \\ \displaystyle\sum_{i \in K} \alpha_i^K(y) = 1. \end{cases} \tag{8}$$

For all $i \in K$, we have

$$f(p_\lambda(y)) = f_i(p_\lambda(y)). \tag{9}$$

Without loss of generality, for $K \in \mathscr{B}(x)$, we will assume that $K = \{1, \cdots, m\}$, $m = |K|$. Consider the following systems:

---

1) See footnote 1) on page 1423.

$$H^K(z, q, y) := \begin{pmatrix} \lambda(f_1(z) - f_m(z)) \\ \vdots \\ \lambda(f_{m-1}(z) - f_m(z)) \\ \lambda \sum_{i=1}^{m-1} q_i \nabla f_i(z) + \left(1 - \sum_{i=1}^{m-1} q_i\right) \nabla f_m(z) + z - y \end{pmatrix},$$

where $(z, q, y) \in \mathbb{R}^n \times \mathbb{R}^{m-1} \times \mathbb{R}^n$. If we set $z^0 = p_\lambda(x)$, $q_i^0 = \alpha_i^K(x)$, $i = 1, \cdots, m-1$, and $y^0 = x$, then from eqs. (8) and (9) we have $H^K(z^0, q^0, y^0) = 0$. Denote

$$A^K(z, q, y) = \begin{pmatrix} \lambda \nabla \bar{f}_{\bar{K}}(z)^{\mathrm{T}} & 0 \\ I + \lambda\left(\sum_{i \in \bar{K}} q_i \nabla^2 f_i(z) + \left(1 - \sum_{i \in \bar{K}} q_i\right) \nabla^2 f_m(z)\right) & \lambda \nabla \bar{f}_{\bar{K}}(z) \end{pmatrix},$$

where $\nabla \bar{f}_i(z) = \nabla f_i(z) - \nabla f_m(z)$, $i \in \bar{K}$, and $\bar{K} = \{1, \cdots, m-1\}$. Denote

$$B^K(z, q, y) = I + \lambda\left(\sum_{i \in \bar{K}} q_i \nabla^2 f_i(z) + \left(1 - \sum_{i \in \bar{K}} q_i\right) \nabla^2 f_m(z)\right).$$

Since f_i, $i \in J$, are twice continuously differentiable convex functions and $q_i^0 = \alpha_i^K(x) > 0$, $i = 1, \cdots, |\bar{K}|$, $1 - \sum_{i \in \bar{K}} q_i^0 = 1 - \sum_{i \in \bar{K}} \alpha_i(x) = \alpha_m(x) > 0$, there exists a neighborhood V^K of (z^0, q^0, y^0) such that $B^K(z, q, y)$ is a symmetric positive definite matrix when $(z, q, y) \in V^K$. By CRCQ and since the vectors

$$\left\{ \begin{pmatrix} \nabla f_i(z^0) \\ 1 \end{pmatrix} : i \in K \right\}$$

are linearly independent, there exists a neighborhood (we still denote it by V^K) of (z^0, q^0, y^0) such that the vectors

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

are linearly independent when $(z, q, y) \in V^K$. Then it follows that the vectors $\{\nabla \bar{f}_i(z) : i \in \bar{K}\}$ are linearly independent. So the nonsingularity of $A^K(z, q, y)$ follows easily when $(z, q, y) \in V^K$. By the implicit function theorem, there exist an open neighborhood U^K of $y^0(=x)$ and an open neighborhood W^K of (z^0, q^0) such that when $y \in \mathrm{cl}\, U^K$, the equation $H^K(z, q, y) = 0$ has a unique solution $(z^K(y), q^K(y)) \in \mathrm{cl}\, W^K$, where $\mathrm{cl}\, S$ denotes the closure of a set S. Moreover, $(z^K(y), q^K(y))$ is continuously differentiable in U^K. Define $G^K: U^K \to \mathbb{R}^n$ as $G^K(y) = \frac{1}{\lambda}(y - z^K(y))$, $y \in U^K$. Then $G^K(y)$ is continuously differentiable in U^K.

Let $N \subseteq \{\bigcap_{K \in \mathscr{B}(x)} U^K\} \cap U$ be an open neighborhood of $y^0(=x)$ such that for any $y \in N$ and $K \in \mathscr{B}(x)$,

$$\left(p_\lambda(y), \alpha_{\bar{K}}^K(y)\right) \in \mathrm{cl}\, W^K. \tag{10}$$

The above relation (10) can be satisfied due to the facts that $\alpha^K(y) \to \alpha^K(x)$ and $p_\lambda(y) \to p_\lambda(x)$ as $y \to x$. For $K \in \mathscr{B}(x)$, denote $N^K = \{y \in N : K \in \mathscr{B}(y)\}$. Then from eq. (7) $N = \bigcup_{K \in \mathscr{B}(x)} N^K$. So for any $y \in N$, there exists $K \in \mathscr{B}(x)$ such that $y \in N^K$. But from eqs. (8) and (9) we know that

$$H^K\left(p_\lambda(y), \alpha_{\bar{K}}^K(y), y\right) = 0.$$

Then it follows that

$$\left(p_\lambda(y),\ \alpha_{\bar{K}}^{K}(y)\right) = (z^K(y),\ q^K(y))$$

from eq. (10) and the uniqueness of the solution of the equation $H^K(z, q, y) = 0$ in $\mathrm{cl}\,W^K$ for $y \in N^K \subseteq \mathrm{cl}\,U^K$. So for $y \in N^K$,

$$G_\lambda(y) = \frac{1}{\lambda}(y - p_\lambda(y)) = \frac{1}{\lambda}(y - z^K(y)) = G^K(y),$$

which means that for any $y \in N$, there exists at least a continuously differentiable function G^K: $U^K \supseteq N \rightarrow \mathbb{R}^n$ such that $G_\lambda(y) = G^K(y)$. This shows that in the neighborhood N of x, G_λ is piecewise smooth.

When f_i, $i \in J$, are linear functions, CRCQ holds automatically, so we have

Theorem 2. *Suppose that* f_i, $i \in J$, *are linear functions. Then* G_λ, *the derivative of the Moreau-Yosida approximation of* f, *is a piecewise affine function, hence a semismooth function in a neighborhood of any* $x \in \mathbb{R}^n$.

(*Received January* 27, 1997)

References

1 Rockafellar, R. T., Monotone operators and the proximal point algorithm, *SIAM J. Control and Optimization*, 1976, 14: 877.

2 Chaney, R. W., Piecewise C^K functions in nonsmooth analysis, *Nonlinear Analysis, Theory, Methods and Applications*, 1990, 15: 649.

Acknowledgement This work was supported by the National Natural Science Foundation of China (Grant No. 19136012) and the Management, Decision and Information System Lab, Chinese Academy of Sciences.