On Smoothness of the Moreau-Yosida Regularization of a Lagrangian-dual Function [1]

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

Department of Mathematics

National University of Singapore

10 Kent Ridge Crescent, Singapore 119260, Singapore

Abstract

In this paper we consider the Lagrangian-dual of a class of optimization problems. First we discuss the piecewise smoothness of the Lagrangian-dual function φ. This property is then used to investigate the second-order properties of the Moreau-Yosida regularization η of the function φ. We show that the gradient of regularized function η is semismooth for certain instances of the optimization problems. We also find a simple instance of the optimization problem, whose Lagrangian-dual function is not piecewise smooth. However, its regularized function still possesses nice second-order properties, e.g. the gradient of the regularized function is semismooth.

Keywords. Moreau-Yosida Regularization, Piecewise C^k Functions, Lagrangian-dual Functions.

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

---

[1] Research is partially supported by NUS Academic Research Grant R-146-000-006-112.

[2] The paper is written during this author's visit to DAMTP of the University of Cambridge. The visit is granted by the Commonwealth Fellowship. This author would like to thank Professor Mike Powell for his support and hospitality.

1 Introduction

Consider the following large scale convex programming:

$$
\begin{aligned}
\min \quad & c^T x \\
s.t. \quad & Ax = a, \\
& f_i(x) \le 0, \quad i = 1, 2, \ldots, \theta, \\
& x \in R^n,
\end{aligned} \tag{1}
$$

where $f_i(x)$ are smooth, convex on R^n, $i = 1, 2, \ldots, \theta$, $A \in R^{m \times n}$ and $\text{rank}(A) = m \le n$.

It is known that a lot of practical problems can be inducted into problem (1). Especially, the program (1) has played a great role in multi-stage stochastic nonlinear programming (MSSNP), which has been intensively studied in the past years [1, 5, 15, 18, 19].

Let $\mathcal{F} := \{x \in R^n : f_i(x) \le 0, i \in \hat{I}\}$ where $\hat{I} = \{1, 2, \ldots, \theta\}$. In many circumstances, in particular in (MSSNP) $\mathcal{F}$ are separable, while the constraint $Ax = a$ is not separable. Thus we first seek to relax the constraint $Ax = a$ by using Lagrangian dual of problem (1) as follows:

$$\min\{\varphi(v) \mid v \in R^m\}, \tag{2}$$

where

$$\varphi(v) = \max\{-c^T x + v^T (Ax - a) \mid x \in \mathcal{F}\}. \tag{3}$$

A substantial obstacle in solving problem (2) is that the function $\varphi(v)$ is nondifferentiable. To overcome this, we convert problem (2) into another convex problem by using the so-called Moreau [10] -Yosida [17] regularization of φ, associated with M, defined by

$$\min\{\eta(v) \mid v \in R^m\}, \tag{4}$$

and

$$\eta(v) = \min_{w \in R^m} \{\varphi(w) + \frac{1}{2}||w - v||_M^2\}, \quad v \in R^m. \tag{5}$$

where M is a symmetric positive definite $m \times m$ matrix and for any $v \in R^m$, let $\|v\|_M^2 = v^T M v$.

It is known that the set of minimizers of the problem (4) is exactly the set of minimizers of the problem (2). It can be shown that η is continuously differentiable and its gradient $g = \nabla\eta$ is globally Lipschitz continuous with modulus $||M||$. For more properties of Moreau-Yosida regularization, readers are referred to [6, 8]. For the problem discussed in the present paper, there are some remarkable advantages by using the Moreau-Yosida regularization:

(a) The minimization in (5) is separable if φ is separable. This is very important for solving large-scale problems. For instance, in multi-stage stochastic programming the functions f_i are separable in the following sense: $x = (x_{I_1}; \ldots; x_{I_k})$ where x_{I_j} is a subvector of x and f_i, $i = 1, 2, \ldots, \theta$, can be divided into k groups, so that for $i \in I_j$, the function f_i depends only on the subvector x_{I_j}, i.e., $f_i(x) = \tilde{f}_i(x_{I_j})$. Thus φ is separable and the problem (5) preserves the separability.

(b) Fukushima and Qi [4] have shown that the superlinear convergence can be guaranteed by using approximate solutions of the problem (5) to construct search directions for minimizing η. While finding an exact solution for a nonsmooth function φ is difficult, the computation of an approximate solution is not difficult in general. We can, e.g., consider a parameterized function $\varphi(w,\mu)$ where $\varphi(w,\mu) \to \varphi(w)$ as $\mu \to 0$ and $\varphi(w,\mu)$ is smooth for any $\mu > 0$ as in the case of the barrier function method used in [19]. For any prescribed accuracy, we can choose an appropriate $\mu > 0$ such that the minimizer of the $\varphi(w,\mu) + (1/2)\|w - v\|_M^2$ is a desirable approximate solution to problem (5).

An interesting point about (b) is that both parameterized function $\varphi(\cdot,\mu)$ and the regularized function η (without parameter) are used to smooth the nonsmooth function φ. However, they function in different ways and have different properties: the former is successful in global convergence and the latter can speed up local convergence. Incorporating the parameterizations into the Moreau-Yosida regularization can be a way to combine advantages in both approaches.

Besides the parameterizations mentioned above, there are many other methods for computing approximate minimizers of φ. Each of these methods can be incorporated into the Moreau-Yosida regularization, and gives rise to an enhanced method for minimizing the nonsmooth function φ. Hence establishing theoretical framework of the Moreau-Yosida regularization can benefit a variety of algorithms. One of the most important properties about Moreau-Yosida regularization is the semismoothness of the gradient of the regularized function, $g := \nabla\eta$, which plays a key role in the superlinear convergence of the Moreau-Yosida regularization method, cf. [4].

Recently, several authors have considered the semismoothness of g for some classes of φ, for example, see [7, 8, 9, 11, 14]. In [11], Qi conjectured that g is semismooth under a regularity condition if φ is the maximum of several twice continuously differentiable convex functions. In [16], Sun and Han 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 [9], Mifflin, Qi and Sun investigated the case where φ is piecewise C^2 which is a generalization of the maximum of convex C^2 functions. They proposed a qualification, called (AIPCQ), which is weaker than (CRCQ) used in [16]. Under (AIPCQ) they derived the same result for g as in [16].

The main task of this paper is to study the semismoothness of g for the Lagrangian dual function φ defined in (3). The main tool used in this study is Proposition 1 in the next section which is established by Mifflin, Qi and Sun in [9] based on the notion of piecewise smooth functions. We will consider different types of constraint functions f_i which define $\mathcal{F}$ and, in turn, φ. We will show that if all f_i's are affine functions or all $\nabla^2 f_i$'s are positive definite then the function φ is piecewise C^2 and satisfies (AIPCQ), and by Proposition 1 g is semismooth. We will also present an example in which some of f_i's are affine and some possess positive definite Hessian matrices. In this case, the function φ is, surprisingly, *not* piecewise C^2, and, equally surprisingly, the gradient g of the regularization of φ *is* semismooth.

The rest of the paper is organized as follows. In Section 2, basic definitions and propositions are collected. Section 3 investigates the piecewise smoothness of the function φ. Section 4 studies the semismoothness of the gradient g. The example is presented in Sections 3 and 4. Section 5 concludes.

2 Preliminaries

In this section we recall some concepts, such as semismoothness, piecewise C^k, and the (AIPCQ) condition, which are basic elements of this paper.

It is well known that η is a continuously differentiable convex function defined on R^m even though φ may be nondifferentiable. The gradient of η at v is

$$g(v) \equiv \nabla\eta(v) = M(v - p(v)), \tag{6}$$

where $p(v)$ is the unique solution of (5). 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 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 [12] introduced the definition of semismoothness for vector-valued functions below:

Definition 1. Suppose $F : R^n \to R^m$ is a locally Lipschitzian vector-valued function, we say that F is *semismooth* at u if

$$\lim_{\substack{V\in\partial F(u+th'),\\ h'\to h,\, t\downarrow 0}} \{Vh'\}, \tag{7}$$

exists for any $h \in R^n$, where $\partial F(u)$ is the generalized Jacobian defined by Clarke in [3].

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 [4]. A direct verification of semismoothness is difficult in general. The Lagrangian dual function φ defined in (3) has a special feature: it is piecewise smooth. We shall make use of this special feature to investigate the semismoothness of g. The following is the definition for piecewise smooth functions, cf [2].

Definition 2. A continuous function ψ is said to be a *piecewise* C^k function on a set $D \subseteq R^n$ if there exist a finite index set $I = \{1, \ldots, q\}$, closed sets $D_1, \ldots, D_q$, open sets $U_1, \ldots, U_q$, and functions $\psi_1, \ldots, \psi_q$, such that

(i) $D \subseteq \cup_{j=1}^{q} D_j$ and $D \cap D_j \subseteq U_j$ for each $j \in I$,

(ii) $\psi_j \in C^k(U_j)$ for each $j \in I$,

(iii) $\psi(u) = \psi_j(u)$ for any $u \in D_j$ and each $j \in I$.

We refer to $\{D_j, U_j, \psi_j\}_{j \in I}$ as a *representation* of ψ.

Remark: If every $U_j \supseteq D$, then Definition 2 can be simply stated as follows. A continuous function ψ is a *piecewise* C^k function on a set $D \subseteq R^n$ if there exist a finite set of functions $\psi_j \in C^k(U_j)$ for $j = 1, \ldots, q$, such that for any $u \in D$, $\psi(u) \in \{\psi_1(u), \ldots, \psi_q(u)\}$.

For the Moreau-Yosida regularization of a piecewise smooth functions to be smooth, the pieces ψ_j must be joined together properly. Rigorously, Mifflin, Qi and Sun [9] introduced the following constraint qualification—(AIPCQ). For any $u \in D$, we write

$$I(u) = \{i \in I : u \in D_i\}.$$

Definition 3. *Affine Independence Preserving Constraint Qualification*(AIPCQ) is said to hold for a piecewise smooth function ψ at u if for every subset $K \subseteq I(u)$ for which there exists a sequence $\{u^k\}$ with $\{u^k\} \to u$, $K \subseteq I(u^k)$ and the vectors

$$\left\{ \begin{pmatrix} \nabla \psi_i(u^k) \\ 1 \end{pmatrix} : i \in K \right\}, \tag{8}$$

being linearly independent, it follows that the vectors

$$\left\{ \begin{pmatrix} \nabla \psi_i(u) \\ 1 \end{pmatrix} : i \in K \right\}, \tag{9}$$

are linearly independent.

Remark: The set $I(u)$ defined in this paper and the corresponding set in [9], denoted as $I_{[9]}(u)$, are slightly different. In [9], they define

$$I_{[9]}(u) = \{j \in I : \psi_j(u) = \psi(u)\}.$$

Since $u \in D_j$ implies $\psi_j(u) = \psi(u)$, we have $I(u) \subseteq I_{[9]}(u)$. For $u \in U_j \setminus D_j$, $\psi_j(u)$ can be set to any value (as long as $\psi_j \in C^k(U_j)$), hence we can assume, without loss of generality, that $\psi_j(u) \neq \psi(u)$ for all $u \in U_j \setminus D_j$. Under this assumption,

$$I(u) = I_{[9]}(u), \qquad \forall\, u \in D.$$

By virtue of (AIPCQ), Mifflin, Qi, and Sun [9] derived the following theorem.

Proposition 1. *Suppose that φ is a piecewise C^2 convex function and (AIPCQ) holds at the proximal point $p(v)$ for a given $v \in \mathrm{dom}\varphi$. Then there exists an open neighborhood $N(v)$ about v, such that the gradient g of the function η, the Moreau-Yosida regularization of φ, is piecewise C^1(smooth) on $N(v)$. Hence g is semismooth at v.*

3 Piecewise smoothness of φ

In this section, we will study the piecewise smoothness of the Lagrangian-dual function φ which is defined by

$$\varphi(v) = \max\{-c^T x + v^T(Ax - a) \mid x \in \mathcal{F}\}.$$

The piecewise smoothness is an important characteristic of the Lagrangian-dual function φ. The investigation of this characteristic is helpful for methods which use the Lagrangian dual. Hence the results in this section have their own great interest. In the next section, the piecewise smoothness of φ will be used to prove the semismoothness of the gradient of the Moreau-Yosida regularization. In order to study the Lagrangian dual of the problem (1), we first make the following assumptions.

Assumption 1: $c \notin \{A^T s : s \in R^m\}$.

Assumption 2: $f_i \in C^2(R^n)$ for all $i \in \hat{I}$.

Remark: If $c = A^T s$ for some $s \in R^m$, then $c^T x = s^T A x = s^T a$. This means that any feasible solution of (1) is an optimal solution. The above Assumption 1 is to rule out this degenerate case.

Define

$$\zeta(u) = \max\{u^T x \mid x \in \mathcal{F}\}. \tag{10}$$

The Lagrangian-dual function can be written as

$$\varphi(v) = \zeta(A^T v - c) - a^T v. \tag{11}$$

Proposition 2. *If ζ is piecewise C^2 on $R^n \setminus \{0\}$, then φ is piecewise C^2 on R^m under Assumption 1.*

Proof. Since ζ is piecewise C^2, there exist closed sets D_i, open sets U_i and functions $\zeta_i \in C^2(U_i)$ which satisfy Definition 2. Let $\varphi_i(v) = \zeta_i(A^T v - c) - a^T v$, $E_i = \{v : A^T v - c \in D_i\}$ and $V_i = \{v : A^T v - c \in U_i\}$. Then it is evident that $\varphi_i \in C^2(V_i)$, E_i is closed and V_i is open. Furthermore, they satisfy (i-iii) in Definition 2. Hence φ is a piecewise C^2 function. $\square$

By virtue of the above proposition, we need only consider the function ζ. Due to Assumption 1, we will consider ζ only on the open set $R^n \setminus \{0\}$. Actually, being the supporting function of $\mathcal{F}$, ζ has a singularity at $u = 0$. We define some notations in this paper:

(i) Q is said to be a *facet* of $\mathcal{F}$ if there exists an index subset $I_Q \subset \hat{I}$, such that $Q = \{x \in \mathcal{F} : f_i(x) = 0, \forall i \in I_Q\}$. I_Q is referred to as the *index set of the facet Q*.

(ii) The *domain* of ζ is defined as $\mathrm{dom}\zeta := \{u \in R^n \setminus \{0\} : \zeta(u) < \infty\}$, and the domain of φ, if Assumption 1 is satisfied, is defined as $\mathrm{dom}\varphi := \{v \in R^m : \varphi(v) < \infty\}$.

Proposition 3. *Suppose that for every $i \in \hat{I}$, f_i is an affine function on R^n. Then the function φ defined in (3) is a piecewise C^2 convex function on its domain. Especially, φ is piecewise affine on its domain.*

Proof. By Proposition 2, it suffices to show that ζ is piecewise C^2 on $R^n \setminus \{0\}$. According to the remarks after Definition 2, it suffices to show that there exist twice continuously differentiable functions ζ_j on R^n, $j \in \hat{J}$ a finite index set, such that for any $u \in \mathrm{dom}\zeta$,

$$\zeta(u) \in \{\zeta_j(u) : j \in \hat{J}\}. \tag{12}$$

It is known that the polytope $\mathcal{F}$ can be represented by its vertices $\{x_1, \dots, x_p\}$ and rays $\{r_1, \dots, r_q\}$ in the form

$$\mathcal{F} = \{x = \sum_{i=1}^{p} \alpha_i x_i + \sum_{i=1}^{q} \lambda_i r_i \ : \ \alpha_i \geq 0, \sum_{i=1}^{p} \alpha_i = 1, \lambda_i \geq 0\}.$$

Define

$$\bar{\mathcal{F}} = \{x = \sum_{i=1}^{p} \alpha_i x_i + \sum_{i=1}^{q} \lambda_i r_i \ : \ \alpha_i \geq 0, \sum_{i=1}^{p} \alpha_i = 1, \ 0 \leq \lambda_i \leq 1\}.$$

We claim that for any $u \in \mathrm{dom}\zeta$, $\max\{u^T x : x \in \mathcal{F}\} = \max\{u^T x : x \in \bar{\mathcal{F}}\}$. Assume by contradiction that there exists a $\bar{x} \in \mathcal{F} \setminus \bar{\mathcal{F}}$, such that $u^T \bar{x} > \max\{u^T x : x \in \bar{\mathcal{F}}\}$. Denote $J := \{i : \lambda_i > 1\}$, where λ_i's are the coefficients in the representation of $\bar{x}$. Let $\hat{x} \in \bar{\mathcal{F}}$ be defined by the same representation of $\bar{x}$ except of changing the λ_i, $i \in J$, to 1. Then $\bar{x} - \hat{x} = \sum_{i \in J} (\lambda_i - 1) r_i$. Since $\hat{x} \in \bar{\mathcal{F}}$, we have $u^T \hat{x} < u^T \bar{x}$, i.e.,

$$\sum_{i \in J} (\lambda_i - 1) u^T r_i > 0.$$

So there exists at least an $\bar{i} \in J$ with $u^T r_{\bar{i}} > 0$. For any fixed $x_0 \in \mathcal{F}$ and any $\lambda \geq 0$, $x_0 + \lambda r_{\bar{i}} \in \mathcal{F}$. Thus $\zeta(u) \geq u^T x_0 + \lambda u^T r_{\bar{i}} \to \infty$, as $\lambda \to \infty$, which contradicts to the fact

$u \in \mathrm{dom}\zeta$. This shows that for any $u \in \mathrm{dom}\zeta$,

$$\zeta(u) = \max\{u^T x \mid x \in \bar{\mathcal{F}}\}.$$

Note that $\bar{\mathcal{F}}$ is a bounded polytope. Without loss of generality, let $\{\bar{x}_1, \ldots, \bar{x}_k\}$ be all vertices of $\bar{\mathcal{F}}$, and define $\zeta_j(u) = \bar{x}_j^T u$. Then $\zeta_j \in C^2(R^n)$, (here $U_j = R^n$). For any $u \in \mathrm{dom}\zeta$, the set of maximizers of $\zeta(u)$ must contain at least a vertex, say $\bar{x}_j$, of $\bar{\mathcal{F}}$. It follows that

$$\zeta(u) = \bar{x}_j^T u = \zeta_j(u),$$

which shows (12). Thus, $\zeta(u)$ is piecewise C^2 convex on its domain. Evidently, $\zeta(u)$ is also piecewise affine on its domain, and so is φ. $\square$

Now we consider the case where all $\nabla^2 f_i$ $(i \in \hat{I})$ are positive definite. For any facet Q of $\mathcal{F}$ with the index set I_Q, we define

$$W := \{(x, \lambda) \in R^n \times R^{|I_Q|} : f_i(x) = 0,\ i \in I_Q,\ \sum_{i \in I_Q} \lambda_i \nabla^2 f_i(x) \succ 0\}, \tag{13}$$

where $B \succ 0$ means that the matrix B is symmetric positive definite,

$$U := \{u = \sum_{i \in I_Q} \lambda_i \nabla f_i(x) \in R^n : (x, \lambda) \in W\}. \tag{14}$$

Without loss of generality, let $I_Q = \{1, \ldots, k\}$. Denote $\tilde{f} = (f_1, \ldots, f_k)^T$, and define a mapping $\Gamma : R^n \times R^k \to R^{n+k}$ by

$$\Gamma(x, \lambda) := \begin{pmatrix} \sum_{i=1}^k \lambda_i \nabla f_i(x) \\ \\ \tilde{f}(x) \end{pmatrix}, \tag{15}$$

The KKT-conditions for problem (10) can be written as $\Gamma(x, \lambda) = (u; 0)$.

The following lemma plays a fundamental role in the analysis.

Lemma 1. *Let W, U be defined as in (13), (14), respectively. Suppose that all $\nabla^2 f_i(i \in$*

I_Q) are positive definite and $\{\nabla f_i(x)\}_{i\in I_Q}$ are linearly independent. Then (i) U is an open set in R^n; (ii) there exists a continuously differentiable bijective mapping $\xi = (\xi_x, \xi_\lambda) : U \to W$ such that for all $u \in U$, $\Gamma(\xi_x(u), \xi_\lambda(u)) = (u; 0)$, i.e., ξ is the inverse mapping of Γ restricted on U.

Proof. (i) For any $(\bar{x}, \bar{\lambda}) \in W$ let $(\bar{u}; \bar{v}) = \Gamma(\bar{x}, \bar{\lambda})$. Then $\bar{u} = \sum_{j=1}^k \bar{\lambda}_j \nabla f_j(\bar{x})$, and $\bar{v} = 0$. In the following we need to show that $\bar{u}$ is an interior point of U. Let us denote $\nabla\tilde{f} := (\nabla f_1, \ldots, \nabla f_k) \in R^{n\times k}$. Then, $\nabla\tilde{f}(x)$ have full column rank (the rank $= k$) by assumption. Since $f_i \in C^2(R^n)(i \in \hat{I})$, there exists a neighborhood of $(\bar{x}, \bar{\lambda})$, denoted by $\mathcal{N}_x$, such that $\sum_{i=1}^k \lambda_i \nabla^2 f_i(x) \succ 0$ for all $(x, \lambda) \in \mathcal{N}_x$. The Jacobian of Γ is

$$\nabla\Gamma(x, \lambda) = \begin{pmatrix} \sum_{i=1}^k \lambda_i \nabla^2 f_i(x) & \nabla\tilde{f}(x) \\ & \\ \nabla\tilde{f}(x)^T & 0 \end{pmatrix}.$$

It is easy to see that $\nabla\Gamma(x, \lambda)$ is nonsingular on $\mathcal{N}_x$ under the above assumptions. By the inverse function theorem, there exists a neighborhood of $(\bar{u}, \bar{v})$, denoted by $\mathcal{N}_u$, such that there exists an inverse mapping Ψ of Γ defined on $\mathcal{N}_u$, such that for any $(u, v) \in \mathcal{N}_u$, $(x; \lambda) := \Psi(u, v) \in \mathcal{N}_x$ and $\Gamma(x, \lambda) = (u; v)$. In particular, for any $(u, v) \in \mathcal{N}_u$ with $v = \bar{v} = 0$, $(x, \lambda) = \Psi(u, 0)$ satisfies

$$u = \sum_{i=1}^k \lambda_i \nabla f_i(x), \qquad \tilde{f}(x) = 0, \qquad \sum_{i=1}^k \lambda_i \nabla^2 f_i(x) \succ 0. \tag{16}$$

This implies $(x, \lambda) \in W$ and thus $u \in U$ for all $(u, 0) \in \mathcal{N}_u$. Since $U_0 := \{u : (u, 0) \in \mathcal{N}_u\}$ is an open set in R^n, and $\bar{u} \in U_0 \subset U$, $\bar{u}$ is an interior point of U. Thus U is open.

(ii) Since the Jacobian $\nabla\Gamma(x, \lambda)$ is nonsingular and continuous on the entire set W and Γ maps W onto $U \times \{0\}$, the inverse mapping Ψ of Γ defined in (i) is a continuously differentiable bijective mapping from $U \times \{0\}$ onto W. Define a mapping $\xi : U \to W$ by $\xi(u) = \Psi(u, 0)$. Then ξ is continuously differentiable and bijective, and $\Gamma(\xi(u)) = (u; 0)$. $\square$

As a consequence of Lemma 1, we obtain the following result:

Lemma 2. *Let* $\zeta_Q(u) = u^T \xi_x(u)$ *where* $\xi_x(u)$ *is defined as Lemma 1. Then* $\zeta_Q \in C^2(U)$ *and for any* $u \in U$

$$\nabla \zeta_Q(u) = \xi_x(u). \tag{17}$$

Proof. From Lemma 1 and the first equation of $\Gamma(\xi(u)) = (u; 0)$, it follows that

$$u = \sum_{i \in I_Q} \xi_{\lambda_i}(u) \nabla f_i(\xi_x(u)).$$

So,

$$\begin{aligned} \nabla \zeta_Q(u) &= \xi_x(u) + \nabla \xi_x(u) u \\ &= \xi_x(u) + \nabla \xi_x(u) \textstyle\sum_{i \in I_Q} \xi_{\lambda_i}(u) \nabla f_i(\xi_x(u)) \\ &= \xi_x(u) + \textstyle\sum_{i \in I_Q} \xi_{\lambda_i}(u) \nabla \xi_x(u) \nabla f_i(\xi_x(u)). \end{aligned}$$

According to the second equation in $\Gamma(\xi(u)) = (u; 0)$, we have $f_i(\xi_x(u)) = 0$ for all $u \in U$ and $i \in I_Q$. Differentiating these functions, we obtain $\nabla \xi_x(u) \nabla f_i(\xi_x(u)) = 0$, $\forall\, i \in I_Q$. Hence,

$$\sum_{i \in I_Q} \xi_{\lambda_i}(u) \nabla \xi_x(u) \nabla f_i(\xi_x(u)) = 0.$$

Thus, it follows that

$$\nabla \zeta_Q(u) = \xi_x(u).$$

By Lemma 1, $\xi_x(u)$ is continuously differentiable on U. Therefore, ζ_Q is twice continuously differentiable on U. □

Lemma 3. *If all* f_i*,* $i \in I_Q$*, are strictly convex, then for any* $u \in \mathrm{int}(\mathrm{dom}\zeta)$ *where* ζ *is defined by (10), there exists a unique optimal solution to the problem* $\sup\{u^T x : x \in \mathcal{F}\}$.

Proof. Assume that the optimal solution does not exist. Then there exist a sequence of $\{x_i\}$ in $\mathcal{F}$ such that $u^T x_i \to \beta^* := \sup\{u^T x : x \in \mathcal{F}\}$. If the sequence is bounded, then there exists a subsequence with a limit point $\bar{x}$ which is in $\mathcal{F}$ since $\mathcal{F}$ is closed. So $\bar{x}$ is an optimal solution, that contradicts to our assumption. Therefore, the sequence $\|x_i\| \to \infty$. Let K be an index set of a subsequence such that

$$x_i/\|x_i\| \to \bar{y}, \qquad \text{for } i \in K,\, i \to \infty.$$

Let $\bar{u} = u + \epsilon\bar{y}$, where the constant $\epsilon > 0$ is chosen so small such that $\bar{u} \in \mathrm{int}(\mathrm{dom}\zeta)$. Then for $i \in K$, we have

$$\lim_{i\to\infty} \bar{u}^T x_i = \lim_{i\to\infty}(u^T x_i + \epsilon\bar{y}^T x_i) = \lim_{i\to\infty}(u^T x_i + \epsilon\|x_i\|) = \beta^* + \infty = \infty.$$

This contradicts the fact $\zeta(\bar{u}) < \infty$. The contradiction shows that an optimal solution must exist. The uniqueness follows obviously from the strict convexity of $\mathcal{F}$. $\square$

The following is one of the main results in this paper, showing the piecewise smoothness of the function φ.

Proposition 4. *For φ defined by (3), suppose that for all $i \in \hat{I}$, $\nabla^2 f_i(x)$ are positive definite and for any facet Q of $\mathcal{F}$ with the index set I_Q, $\{\nabla f_i(x)\}_{i\in I_Q}$ are linearly independent. Then φ is piecewise C^2 convex on its domain.*

Proof. Let us first consider the function ζ defined by (10). Let $\{Q_1, \ldots, Q_q\}$ be the set of all facets of $\mathcal{F}$. Let W_i, U_i and ξ_i be defined in (13), (14) and Lemma 1 for the facet $Q_i = \{x \in \mathcal{F} : f_l(x) = 0, l \in I_i\}$. Define

$$D_i := \{u = \sum_{l\in I_i} \lambda_l \nabla f_l(x) : x \in Q_i,\, \lambda_l \geq 0\}, \tag{18}$$

which is evidently a closed set. By Lemma 1, U_i is open. Define $\zeta_i(u) := u^T \xi_{ix}(u)$. In what follows, we show that (i), (ii) and (iii) in Definition 2 hold.

(i) For any $u \in \mathrm{int}(\mathrm{dom}\zeta)$, by Lemma 3, there exists an optimal solution x^*, which together with a Langrangian multiplier $\bar{\lambda}^*$ satisfy the KKT-conditions:

$$\begin{aligned} & u = \textstyle\sum_{i=1}^{\theta} \bar{\lambda}_i^* \nabla f_i(x^*), \\ & \bar{\lambda}_i^* \geq 0, \\ & f_i(x^*) \leq 0, \\ & \bar{\lambda}_i^* f_i(x^*) = 0,\ i = 1, 2, \ldots, \theta. \end{aligned} \tag{19}$$

Let $Q_j = \{x \in \mathcal{F} : f_i(x) = 0,\ i \in I_j\}$ be the *smallest* facet at x^*. By "smallest" we mean that for any $i \notin I_j,\ \ f_i(x^*) \neq 0$. Then $\bar{\lambda}_i^* = 0$ for all $i \notin I_j$. Let λ^* denote the subvector of $\bar{\lambda}^*$ consisting of components in I_j. Then $\lambda^* \geq 0,\ \ x^* \in Q_j$ and $u = \sum_{i \in I_j} \lambda_i^* \nabla f_i(x^*)$. Hence $u \in D_j$. This shows that $\mathrm{int}(\mathrm{dom}\zeta) \subseteq \cup_{i=1}^q D_i$. Thereby, $\mathrm{dom}\zeta \subseteq \cup_{i=1}^q D_i$ since each D_i is closed. For any $u \in D_j \cap \mathrm{dom}\zeta$, let $\bar{x} \in Q_j$ and $\bar{\lambda} \geq 0$ represent u as in (18). $\bar{\lambda} \neq 0$ since $u \neq 0$. This implies that $\sum_{i \in I_j} \bar{\lambda}_i \nabla^2 f_i(\bar{x}) \succ 0$ since all f_i possess positive definite Hessian matrices. Thus $(\bar{x}, \bar{\lambda}) \in W_j$ and $u \in U_j$. This shows $D_j \cap \mathrm{dom}\zeta \subseteq U_j$.

(ii) By Lemma 2, $\ \zeta_i \in C^2(U_i)$ for $i = 1, \ldots, q$.

(iii) For any $u \in D_j \cap \mathrm{dom}\zeta$, let $\bar{x} \in Q_j$ and $\bar{\lambda} \geq 0$ represent u as in (18). Let $\bar{\lambda}^*$ be defined by $\bar{\lambda}_i^* = \bar{\lambda}_i$ for $i \in I_j$ and $\bar{\lambda}_i^* = 0$ for $i \notin I_j$. Then $(\bar{x}, \bar{\lambda}^*)$ satisfies the KKT-conditions (19). Thus $\zeta(u) = u^T \bar{x}$. On the other hand, the second part of (i) shows that $(\bar{x}, \bar{\lambda}) \in W_j$. Using the relation in (18), we have $(u; 0) = \Gamma_j(\bar{x}, \bar{\lambda})$, where Γ_j is the mapping defined in (15). Since ξ_j is the inverse of Γ_j restricted on U_j, $\xi_j(u) = (\bar{x}, \bar{\lambda})$. By definition, $\zeta_j(u) = u^T \xi_{jx}(u) = u^T \bar{x}$. Thus $\zeta(u) = \zeta_j(u)$.

The above shows that ζ is piecewise C^2 on $\mathrm{dom}\zeta$. By virtue of Proposition 2, φ is piecewise C^2 on its domain. $\square$

In the proof of Proposition 4, we have defined a representation $\{D_j, U_j, \zeta_j\}_{j \in I}$ of ζ. This representation induces a representation $\{E_j, V_j, \varphi_j\}_{j \in I}$ of φ as defined in the proof of Proposition 2, we will use these notations below. For simplicity, in what follows we assume that

$$\varphi_j(v) \neq \varphi(v), \qquad \forall j \in I,\, v \in V_j \setminus E_j, \tag{20}$$

(recall the remark after Definition 3).

Lemma 4. *Let φ be defined by (3) with the set $\mathcal{F}$ defined by all convex constraints with positive definite Hessian matrices. For any $v \in \mathrm{dom}\varphi$ and any $i,\, j \in I$, if $\varphi_i(v) = \varphi_j(v) = \varphi(v)$, then $\nabla\varphi_i(v) = \nabla\varphi_j(v)$.*

Proof. First let us consider the function ζ. For any $v \in V_i$ let $u = A^Tv - c$ and $\xi_i : U_i \to W_i$ be defined as in Lemma 1, then $\zeta_i(u) = u^T\xi_{ix}(u)$. By (17), $\nabla\zeta_i(u) = \xi_{ix}(u)$. Denote $\bar{x}_i := \xi_{ix}(u)$. Since $\varphi_i(v) = \varphi(v)$ and $v \in V_i$, it follows from (20) that $v \in E_i$. This implies $u \in D_i$ where D_i is defined by (18), thus $\bar{x}_i \in Q_i \subset \mathcal{F}$. Since $\varphi_i(v) = \varphi(v)$, it follows that $\zeta_i(u) = \zeta(u)$, i.e. $u^T\bar{x}_i = \zeta(u)$. Hence $\bar{x}_i \in \mathcal{F}$ maximizes u^Tx on $\mathcal{F}$. Since $\mathcal{F}$ is strictly convex, thus the solution to $\max\{u^Tx \mid x \in \mathcal{F}\}$ is unique, which is denoted by x^*. So, $\bar{x}_i = x^*$, namely, $\nabla\zeta_i(u) = x^*$. Hence $\nabla\zeta_i(u) = \nabla\zeta_j(u)$ if $\varphi_i(v) = \varphi_j(v) = \varphi(v)$. By (11), $\nabla\varphi_i(v) = A^T\nabla\zeta_i(A^Tv - c) - a$, so we have $\nabla\varphi_i(v) = \nabla\varphi_j(v)$ if $\varphi_i(v) = \varphi_j(v) = \varphi(v)$. $\square$

The above lemma will be used in the next section to prove Proposition 6. Besides it, this lemma leads to a nice property of the dual function φ.

Corollary 5. *Let φ be defined by (3) where the set $\mathcal{F}$ is defined by all convex constraints with positive definite Hessian matrices. Then φ is continuously differentiable on its domain.*

Proof. For any $v \in \mathrm{dom}\varphi$, denote $I(v) = \{i \in I : \varphi_i(v) = \varphi(v)\}$. The subdifferential of φ at v is the convex hull of $\{\nabla\varphi_i(v) : i \in I(v)\}$. By Lemma 4, the subdifferential is a singleton, thus φ is differentiable at v. Furthermore, as is shown in the proof of Lemma 4, $\nabla\varphi(v) = x^*(v)$, where $x^*(v)$ is the unique maximizer of $\max\{-c^Tx + v^T(Ax - a) \mid x \in \mathcal{F}\}$. Since $\mathcal{F}$ is strictly convex, $x^*(v)$ is continuous. Hence, φ is continuously differentiable. $\square$

Remark: In Propositions 3 and 4 we conclude that φ is piecewise C^2 convex under the assumption that the constraints for $\mathcal{F}$ are either all linear or all have positive definite Hessian matrices. A natural question arises: can φ be piecewise C^2 for more general $\mathcal{F}$? The following example considers a $\mathcal{F}$ which is defined by a linear constraint and a strictly convex constraint with the positive definite Hessian, and gives a negative answer to the above question.

Example 1. Let

$$\varphi(v) = \max\{-c^T x + v^T(Ax - a) \mid x \in \mathcal{F}\},$$

where

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad c = (0, 0, -1)^T, \qquad a = (0, 0)^T,$$

and

$$\mathcal{F} = \{x \in R^3 : x_1^2 + x_2^2 + x_3^2 - 1 \leq 0,\ x_3 \leq 0\}.$$

It is easy to see that Assumptions 1 and 2 are satisfied. We can write

$$\varphi(v) = \max\{(v_1, v_2, 1)x \mid x \in \mathcal{F}\}.$$

It is easy to verify that the maximizer is $x = (v_1/\|v\|, v_2/\|v\|, 0)^T$. It follows that

$$\varphi(v) = \sqrt{v_1^2 + v_2^2}.$$

Obviously, φ is smooth at any point $v \neq 0$. Let E_1 be a closed set containing those points close to $(0, 0)$ and φ_1 be a function representing φ on E_1. Since E_1 is closed, it must also contain points $(0, 0)$. For any non-zero $v \in E_1$, the gradient and the Hessian of $\varphi_1(= \varphi)$ can be written as

$$\nabla\varphi_1(v) = \begin{pmatrix} v_1/\sqrt{v_1^2 + v_2^2} \\ v_2/\sqrt{v_1^2 + v_2^2} \end{pmatrix},$$

$$\nabla^2\varphi_1(v) = \begin{pmatrix} v_2^2/(v_1^2+v_2^2)^{3/2} & -v_1v_2/(v_1^2+v_2^2)^{3/2} \\ \\ -v_1v_2/(v_1^2+v_2^2)^{3/2} & v_1^2/(v_1^2+v_2^2)^{3/2} \end{pmatrix}.$$

Now consider $v \in E_1$ satisfying $v_1 = v_2 \neq 0$ and let $v \to v^* := (0,0)^T$, we have

$$\lim_{v\to v^*} \frac{\partial^2\varphi_1(v)}{\partial v_1 \partial v_1} = \lim_{v\to v^*} \frac{v_2^2}{(v_1^2+v_2^2)^{3/2}} = \lim_{v_1\to 0+} \frac{v_1^2}{(2v_1^2)^{3/2}} = \infty.$$

This shows that φ_1 is not a C^2 function on E_1. Therefore, φ is not a piecewise C^2 function on its domain.

4 Semismoothness of the gradient of η

In this section we study the semismoothness of g $(= \nabla\eta)$. First, we seek to show that φ defined in (3) satisfies (AIPCQ) in the cases: (i) all f_i are affine; (ii) all f_i possess positive definite Hessian matrices. Suppose $\varphi(v)$ is a piecewise smooth function with the representation $\{E_i, V_i, \varphi_i\}_{i\in I}$. For any $v \in R^m$, define

$$I(v) := \{i \in I : v \in E_i\}.$$

When $\mathcal{F}$ is defined by all linear constraints, we obtain the following result for φ:

Proposition 5. *Suppose for every $i \in \hat{I}$, f_i is an affine function on R^n. Then for the piecewise affine function $\varphi(v)$ defined by (3), (AIPCQ) holds at every v on its domain.*

Proof. By Proposition 3, $\varphi(v)$ is piecewise affine. φ can be represented by $\{\varphi_i\}_{i\in I}$ where $\varphi_i(v) = \beta_i^T v - \alpha_i$. For any $w \in R^m$ and any index set $K \subseteq I$,

$$\left\{ \begin{pmatrix} \nabla\varphi_i(w) \\ 1 \end{pmatrix} : i \in K \right\} = \left\{ \begin{pmatrix} \beta_i \\ 1 \end{pmatrix} : i \in K \right\},$$

is a set of constant vectors. Therefore (AIPCQ) holds at any v in the domain of φ. $\square$

Now we consider the case that the set $\mathcal{F}$ is defined by all convex functions f_j with positive definite Hessian matrices.

Proposition 6. *Let φ be defined by (3) with the set $\mathcal{F}$ defined by all convex functions $f_j (j \in I)$ with positive definite Hessian matrices. Suppose that for any facet Q of $\mathcal{F}$ with the index set I_Q, vectors $\{\nabla f_j(x)\}_{j \in I_Q}$ are linearly independent for any $x \in Q$. Then (AIPCQ) holds at each $v \in \mathrm{dom}\varphi$.*

Proof. Let $v \in \mathrm{dom}\varphi$ and $K \subseteq I(v)$. To verify (AIPCQ), we shall show that the conditions for (AIPCQ) are satisfied for cases $|K| = 1$ and $|K| \geq 2$. If $|K| = 1$, the vectors in the set in (9) are evidently linearly independent, (actually, the set is a singleton). So the conditions for (AIPCQ) are satisfied. If $|K| \geq 2$, then for any $w \neq v$ with $K \subseteq I(w)$ and for any $i \neq j \in K$, $\varphi_i(w) = \varphi_j(w) = \varphi(w)$ implies that $\nabla\varphi_i(w) = \nabla\varphi_j(w)$ by Lemma 4. Thus the set of vectors in (8) can never be linearly independent. This means that the conditions for (AIPCQ) hold automatically. $\square$

From the piecewise C^2 smoothness of φ shown in Section 3 and the qualification (AIPCQ) verified in this section, we have the semismoothness of g, as is stated below.

Proposition 7. *Let φ be defined by (3) where the constraint functions f_i, $i \in \hat{I}$ are either all affine or all possess positive definite Hessian matrices. And, for the latter case suppose that for any facet Q of $\mathcal{F}$ with the index set I_Q, $\{\nabla f_i(x)\}_{i \in I_Q}$ are linearly independent. Then the gradient $g(v)$ $(= \nabla\eta(v))$ of the Moreau-Yosida regularization η is piecewise smooth, and thereby semismooth, on* $\mathrm{dom}\varphi$.

Proof. By Proposition 1, g is piecewise smooth on $\mathrm{dom}\varphi$ if φ is piecewise C^2 on $\mathrm{dom}\varphi$ and (AIPCQ) is satisfied at $p(v)$ for any $v \in \mathrm{dom}\varphi$. Propositions 3 and 4 show the property of piecewise C^2 of φ. For each $v \in \mathrm{dom}\varphi$, if $p(v) \notin \mathrm{dom}\varphi$, then $\varphi(p(v)) = \infty$ that implies $p(v)$ cannot be the minimizer of (5). Thus $p(v)$ must be on $\mathrm{dom}\varphi$. Thereby, Propositions 5 and 6 imply that (AIPCQ) is satisfied at $p(v)$ for this two cases. $\square$

Remark: The above proposition shows that g is semismooth if constraints defining $\mathcal{F}$ are either all linear or all possess positive definite Hessian matrices. In Section 3 Example 1, we found that for some simple such two mixed constraints, the Lagrangian dual function φ is not piecewise C^2. Actually, the second-order derivatives of φ tend to infinity at some point. Since the semismoothness of g is closely related to the piecewise C^2 smoothness of φ, we might expect that for this example g is not semismooth either. However, to our surprise, the gradient g of the Moreau-Yosida regularization of this function φ turns out to be semismooth, as is shown in the following example.

Example 1 (continued). It is known that $\varphi(v) = \sqrt{v_1^2 + v_2^2}$. For convenience in description we set $M = I$, so we have

$$\eta(v) = \min\{\sqrt{w_1^2 + w_2^2} + \frac{1}{2}\|w - v\|^2 \mid w \in R^2\}.$$

It is easy to verify that for $\|v\| \le 1$,

$$\eta(v) = (v_1^2 + v_2^2)/2, \qquad p(v) = (0,\ 0)^T,$$

and for $\|v\| \ge 1$,

$$\eta(v) = \sqrt{v_1^2 + v_2^2} - 1/2, \quad p(v) = (1 - 1/\|v\|)v.$$

Let $\hat{V}_1 = \{v \in R^2 : \|v\| \le 1\}$ and $\hat{V}_2 = \{v \in R^2 : \|v\| \ge 1\}$. By (6) it suffices to study the semismoothness of p. For $v \in \hat{V}_1$, the Jacobian of p is

$$J(p(v)) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \tag{21}$$

and for $v \in \hat{V}_2$,

$$J(p(v)) = \begin{pmatrix} 1 - v_2^2/\|v\|^3 & v_1 v_2/\|v\|^3 \\ v_1 v_2/\|v\|^3 & 1 - v_1^2/\|v\|^3 \end{pmatrix}. \tag{22}$$

From the Jacobian of p above we can see that p is smooth on the interior of $\hat{V}_i (i = 1, 2)$. Thus we only need to investigate the semismoothness of p on the region where the two sets meet, namely, $\{v \in R^2 : \|v\| = 1\}$. Let $\bar{v} = (\bar{v}_1, \bar{v}_2)^T$ be any point on this region, we will show that p is semismooth at $\bar{v}$. By Definition 1, it suffices to show

$$\lim_{h' \to h, t \to 0+} \{Vh' : V \in \partial p(\bar{v} + th')\}, \tag{23}$$

exists for any $h \in R^2$. Let $S_1(\bar{v}) = \{h \in R^2 : h^T \bar{v} < 0\}$, $S_2(\bar{v}) = \{h \in R^2 : h^T \bar{v} > 0\}$, $S_3(\bar{v}) = \{h \in R^2 : h^T \bar{v} = 0\}$. Write $v' = \bar{v} + th'$. Then $\|v'\|^2 = \|\bar{v}\|^2 + 2t\bar{v}^T h' + t^2\|h'\|^2$.

If $h \in S_1(\bar{v})$ or $h \in S_2(\bar{v})$, then for any sufficiently small $t > 0$ and h' close to h, $v' \in \mathrm{int}\hat{V}_1$ or $v' \in \mathrm{int}\hat{V}_2$. It is evident that the limit in (23) exists.

If $h \in S_3(\bar{v})$. Then for any sufficiently small $t > 0$ and h' close to h, there are the following three cases: If $\|v'\| < 1$, we have

$$\lim_{h' \to h, t \to 0+} Vh' = \lim_{h' \to h, t \to 0+} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} h'_1 \\ h'_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{24}$$

If $\|v'\| > 1$, we have

$$\begin{aligned} \lim_{h' \to h, t \to 0+} Vh' &= \lim_{h' \to h, t \to 0+} \begin{pmatrix} 1 - {v'_2}^2/\|v'\|^3 & v'_1 v'_2/\|v'\|^3 \\ v'_1 v'_2/\|v'\|^3 & 1 - {v'_1}^2/\|v'\|^3 \end{pmatrix} \begin{pmatrix} h'_1 \\ h'_2 \end{pmatrix} \\ &= \lim_{h' \to h, t \to 0+} \begin{pmatrix} (1 - {v'_2}^2/\|v'\|^3)h'_1 + (v'_1 v'_2/\|v'\|^3)h'_2 \\ (v'_1 v'_2/\|v'\|^3)h'_1 + (1 - {v'_1}^2/\|v'\|^3)h'_2 \end{pmatrix}. \end{aligned} \tag{25}$$

Since

$$\begin{aligned} & \lim_{h' \to h, t \to 0+} [(1 - {v'_2}^2/\|v'\|^3)h'_1 + (v'_1 v'_2/\|v'\|^3)h'_2] \\ = \; & (1 - \bar{v}_2^2/\|\bar{v}\|^3)h_1 + (\bar{v}_1 \bar{v}_2/\|\bar{v}\|^3)h_2 = (1 - \bar{v}_2^2)h_1 + \bar{v}_1 \bar{v}_2 h_2 \\ = \; & \bar{v}_1^2 h_1 + \bar{v}_1 \bar{v}_2 h_2 = \bar{v}_1(h_1 \bar{v}_1 + h_2 \bar{v}_2) = 0, \end{aligned}$$

and similarly, $\lim_{h' \to h, t \to 0+}[(v'_1 v'_2/\|v'\|^3)h'_1 + (1 - {v'_1}^2/\|v'\|^3)h'_2] = 0$, by (25), we have

$$\lim_{h' \to h, t \to 0+} Vh' = (0, 0)^T. \tag{26}$$

Hence, Vh' tends to the same limit in these two cases by (24) and (26). If $\|v'\| = 1$, by the definition of generalized Jacobian and (21), (22), it follows that V is a convex combination of the corresponding expressions in the above two cases. Thus, Vh' also tends to the same limit as the above two cases. Thereby, the limit in (23) exists if $h \in S_3(\bar{v})$. The above shows that p is semismooth on R^2. Therefore, g is semismooth on R^2 as well.

5 Conclusion

The Lagrangian dual is widely used for large-scale problems. A significant feature of Lagrangian dual functions is the piecewise smoothness, which is studied in this paper and employed in the analysis of the Moreau-Yosida regularization of the Lagrangian dual function φ. We investigate the semismoothness of the gradient g of the Moreau-Yosida regularization of φ, which plays a key role in the superlinear convergence analysis of approximate Newton methods. We show that the gradient g is piecewise smooth and thereby semismooth if the inequality constraints in (1) are either all linear or all possess positive definite Hessian matrices. An example with a mixed set of these two different convex constraints is constructed. We find surprisingly that the Lagrangian dual function of this problem is *not* piecewise C^2, and equally surprisingly, that the gradient g of the Moreau-Yosida regularization of this Lagrangian dual function *is* semismooth. We have not been able to prove the semismoothness of the gradient g for general problems with such mixed constraints. Actually, there are still a lot of questions left open. For instance, for problems with general strictly or non-strictly convex inequality constraints, we have neither a confirmative proof nor a counterexample for the semismoothness of the gradient g. It is a great challenge to answer these questions.

References

[1] J.R. Birge, F. Louveaux, *Introduction to stochastic programming*, Springer-Verlag, 1997.

[2] R.W. Chaney, *Piecewise C^k functions in nonsmooth analysis*, Nonlinear Analysis, Theory, Methods & Applications, 15 (1990), 649–660.

[3] F.H. Clarke, *Optimization and nonsmooth analysis*, John Wiley and Sons, New York,1983.

[4] M. Fukushima, L. Qi, *A global and superlinear convergent algorithm for nonsmooth convex minimization*, SIAM Journal on Optimization, 6 (1996), 1106–1120.

[5] J.L. Higle, S. Sen, *Stochastic decomposition*, Kluwer Academic Publishers, 1996.

[6] J.B. Hiriart-Urruty, C. Lemaréchal, *Convex analysis and minimization algorithms II.* Springer Verlag, Berlin, 1993.

[7] C. Lemaréchal, C. Sagastizábal, *An approach to variable metric methods*, in: J. Henry and J. -P. Yvon, eds., Systems modelling and optimization, Lecture notes in control and information sciences 197 (Springer-Verlag, Berlin, 1994) pp.144-162.

[8] C. Lemaréchal, C. Sagastizábal, *Practical aspects of the Moreau-Yosida regularization I: theoretical preliminaries,* SIAM Journal on Optimization, 7(1997), 367–385.

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

[10] J. J. Moreau, *Proximite et dualite dans un espace hilbertien*, Bulletin de la Societe Mathematique de France, 93 (1965), 273–299.

[11] L. Qi, *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, 1995).

[12] L. Qi, J. Sun, *A nonsmooth version of Newton's method,* Mathematical Programming, 58 (1993), 353-367.

[13] R.T. Rockafellar, *Convex analysis.* Princeton, New Jersey, 1970.

[14] R.T. Rockafellar, *Maximal monotone relations and the second derivatives on nonsmooth functions*, Annales de l'sInstitute Henri Poincaré, Analyse Non Linéaire, 2 (1985), 167–186.

[15] A. Ruszczyński, *Decomposition methods in stochastic programming*, Mathematical Programming, 79 (1997), 333-353.

[16] D. Sun, J. Han, *On a conjecture in Moreau-Yosida regularization of a nonsmooth convex function*, Chinese Science Bulletin, 42 (1997), 1140–1143.

[17] K. Yosida, *Functional analysis*, Springer Verlag, Berlin, 1964.

[18] G. Zhao, *A log-barrier method with Benders decomposition for solving two-stage stochastic programs,* to appear in Mathematical Programming.

[19] G. Zhao, *Lagrange dual method with self-concordant barriers for multi-stage stochastic nonlinear programming,* Research Report, Department of Mathematics, National University of Singapore, Singapore, 1999.