J Glob Optim (2009) 44:375–394
DOI 10.1007/s10898-008-9333-7

Moreau–Yosida regularization of Lagrangian-dual functions for a class of convex optimization problems

Fanwen Meng

Received: 28 December 2007 / Accepted: 29 July 2008 / Published online: 19 August 2008

Abstract In this paper, we consider the Lagrangian dual problem of a class of convex optimization problems, which originates from multi-stage stochastic convex nonlinear programs. We study the Moreau–Yosida regularization of the Lagrangian-dual function and prove that the regularized function η is piecewise C^2, in addition to the known smoothness property. This property is then used to investigate the semismoothness of the gradient mapping of the regularized function. Finally, we show that the Clarke generalized Jacobian of the gradient mapping is BD-regular under some conditions.

Keywords Lagrangian dual · Moreau–Yosida regularization · Piecewise C^k functions · Semismoothness

Mathematics Subject Classification (2000) 90C25 · 65K10 · 52A41

1 Introduction

In this paper, we consider the following convex program:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & f_i(x) \le 0, \quad i = 1, 2, \ldots, \theta, \\ & Ax = a, \qquad x \in \mathbb{R}^n, \end{array} \tag{1}$$

where f, $f_i, i = 1, 2, \ldots, \theta$, are smooth and convex on $\mathbb{R}^n$, and $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = m$ and $0 < m < n$. It is known that many practical problems can be converted to problem (1). For example, some recent studied multi-stage stochastic convex nonlinear programming models can be formulated as (1). See [20, Chap. 1] and [25] for the detailed modelling in this regard.

F. Meng (✉)
The Logistics Institute – Asia Pacific, National University of Singapore, 7 Engineering Drive 1, Singapore 117574, Singapore
e-mail: tlimf@nus.edu.sg

Let $\mathcal{F} := \{x \in \mathbb{R}^n : f_i(x) \leq 0,\ i \in \hat{I}\}$ where $\hat{I} = \{1, \ldots, \theta\}$. In many circumstances, particularly in multistage stochastic programming, f and $\mathcal{F}$ are separable into scenarios in the following sense: $x = (x^1, \ldots, x^K)$, $x^k \in \mathbb{R}^{n_k}$, $\sum_{k=1}^K n_k = n$, and f, f_i, $i = 1, \ldots, \theta$, can be divided into K groups which can be represented for all x as sums:

$$f(x) = \sum_{k=1}^{K} f^k(x^k), \quad f_i(x) = \sum_{k=1}^{K} f_i^k(x^k), \ \ i = 1, \ldots, \theta.$$

However, the constraint $Ax = a$, which is called *nonantipicipativity* constraint, is non separable. Thus, we seek to relax the constraint $Ax = a$ using the Lagrangian dual of problem (1) as follows:

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

where

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

The subproblem in (3) is separable and then is solvable using the well-developed parallel algorithms. However, note that (2) is a nonsmooth problem due to the nonsmoothness of φ. To overcome this, we adopt the well known Moreau [13]–Yosida [24] regularization to convert (2) into a smooth problem as follows:

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

where η is the Moreau–Yosida regularization of φ given below

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

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

It is well known [5,6] that (4) is equivalent to (2) in the sense that the set of minimizers of problem (4) coincides with the set of minimizers of (2). Note that η is continuously differentiable and its gradient $g = \nabla\eta$ is globally Lipschitz continuous with modulus $||M||$.

Fukushima and Qi [5] established the superlinear convergence of the generalized Newton's method for solving (4) by using approximate solutions of the problem (5) to construct search directions for minimizing η. Since it is difficult to find an exact solution for (5), the computation of an approximate solution appears easier. For instance, we can use a parameterized function $\varphi(w, \mu)$, which is smooth for $\mu > 0$, such that $\varphi(w, \mu)$ tends to $\varphi(w)$ when the smoothing parameter μ is driven to zero, as in the case of the barrier function method. Actually, this method was employed for solving multi-stage stochastic nonlinear problems recently [25], in which the underlying stochastic problem was formulated exactly as problem (1). For any prescribed accuracy, people may choose an appropriate $\mu > 0$ as long as the solution of minimizing $\varphi(w, \mu) + (1/2)\|w - v\|_M^2$ is a desirable approximate solution to problem (5). In fact, in addition to the above parameterization method, any method for computing approximate minimizers of (2) in literature can be incorporated into the Moreau–Yosida regularization, and gives rise to an enhanced method for minimizing the nonsmooth problem (2). Hence, it is important and meaningful to establish the theoretical framework of the Moreau–Yosida regularization which can benefit a variety of algorithms.

Note that, for the problem considered in this paper, one of the most important properties about the Moreau–Yosida regularization is the semismoothness of the gradient of the regularized function, which has played a key role in establishing the superlinear convergence of the generalized Newton's method for nonsmooth convex problems by using the Moreau–Yosida regularization scheme [5]. The notion of semismooth functions was first introduced by Mifflin [11] which is an important subclass of Lipschitz functions. In order to establish the superlinear convergence of generalized Newton's method for solving nonsmooth equations, Qi and Sun [16] extended the definition of semismoothness to vector valued functions. After the work of Qi and Sun, semismoothness was extensively used to establish superlinear/quadratic convergence of Newton's method for solving the convex best interpolation problem [3,4], the nondifferentiable equations in which the underlying functions are slant differentiable functions [1], complementarity problems and variational inequalities [17], and the inverse eigenvalue problem [23], for instance.

In this paper, our focus is on a special case of semismooth functions, *piecewise C^k functions*, which is a large class of locally Lipschitz continuous functions, found in most practical problems. In the past few years, many people have studied piecewise smoothness of nonsmooth functions and designed algorithms based on Newton's method for solving the associated nonsmooth equations. For instance, the analysis was mainly focused on piecewise C^k functions in [8,12,22], where the authors investigated properties of g for some specific classes of φ. Specifically, Sun and Han [22] showed the semismoothness of g if φ is the maximum of several twice continuously differentiable convex functions under a constant rank constraint qualification (CRCQ). Meng and Hao [8] derived the same result for the case of unconstrained problem (1) where the objective function f is piecewise C^2 under a weaker sequential constant rank constraint qualification. Later, Mifflin et al. [12] investigated the case where φ is piecewise C^2 under an affine independence preserving constraint qualification (AIPCQ). Recently, Meng et al. [10] proved that the Lagrangian-dual function φ in the case of the objective f in (1) being an affine function is piecewise C^2 and, using the result established in [12], they derived the semismoothness of g. Note that the notions of CRCQ and AIPCQ mentioned above are defined with respect to piecewise smooth functions under consideration whereas the usual constraint qualifications in nonlinear programming, e.g., Linear Independence Constraint Qualification (LICQ), Mangasarian-Fromovitz Constraint Qualification (MFCQ), and CRCQ, are defined in terms of constraint functions.

Having motivated the importance of the notions of semismoothness and the Moreau–Yosida regularization in nonsmooth analysis, in this paper, we will investigate properties of the Moreau–Yosida regularization η and the gradient mapping g of η. The analysis of the present paper differs from that of Meng et al. [10] in two main ways. First, we consider a more general problem where the objective function f in (1) is a nonlinear convex function while the analysis in [10] is mainly focused on the case where f is an affine function. Second, the methods used here are quite different from those of [10] where the semismoothness of g is derived based on the study of piecewise C^2-ness of the Lagrangian dual function φ. In this paper, we study the semismoothness of g by investigating the piecewise smoothness of the regularized function η directly. Additionally, in the analysis, we introduce a so called generalized affine independence preserving constraint qualification (GAIPCQ). Unlike AIPCQ defined for piecewise function in [12], GAIPCQ is defined in terms of constraint functions.

Our contributions are as follows. In this paper, we show that the regularized function η is piecewise twice continuously differentiable. It is well known that the regularized function η is smooth, however, the second-order properties of η is unclear yet, which is very important in constructing higher convergence generalized Newton methods for solving (4). Second, we derive the piecewise smoothness and thereby the semismoothness of g. Further, we examine the conditions under which the Clarke generalized Jacobian of the gradient g is BD-regular (See Definition 4 in Sect. 4) and then derive that every element in the Clarke generalized Jacobian is symmetric and positive definite.

The rest of the paper is organized as follows. In Sect. 2, basic definitions and properties are collected. Section 3 investigates the piecewise C^2-ness of the function η and the semismoothness of g. Section 4 studies the BD-regularity of the Clarke generalized Jacobian of the gradient g. Section 5 concludes.

2 Preliminaries

In this section, we briefly recall some concepts, such as semismoothness, piecewise smoothness, MFCQ, and GAIPCQ, which will be used in this paper.

It is known that the regularized function η is a continuously differentiable and convex function on $\mathbb{R}^m$ even though φ may be nondifferentiable. The gradient of η at v (see, [6]) is

$$g(v) \equiv \nabla\eta(v) = M(v - p(v)), \quad v \in \mathbb{R}^m, \tag{6}$$

where $p(v)$ represents the unique solution of the minimization problem in (5). In order to use Newton's method or modified Newton's methods for solving (4), it is important to study the Hessian of η, i.e., the Jacobian of g. A remarkable feature of semismoothness is that superlinear or quadratic convergence of generalized Newton method for solving nonsmooth equations can be obtained under the assumption of semismoothness. See [5,15,16] for the relevant discussions. The notion of semismoothness for vector-valued functions is defined as follows [16].

Definition 1 Suppose $F : \mathbb{R}^n \to \mathbb{R}^m$ 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 $h \in \mathbb{R}^n$, where $\partial F(u)$ denotes the Clarke generalized Jacobian [2].

Note that in general a direct verification of semismoothness is difficult. Some equivalent definitions of semismooth functions and further studies on semismoothness can be found in [9,15] and the references therein. In particular, as discussed earlier, piecewise smooth functions is a special case of semismooth function.

To investigate the semismoothness or piecewise smoothness of g, we will study the piecewise C^2-ness of the regularized function η in the subsequent analysis. It is known that, to establish the higher convergence of Newton's method based algorithms for solving nonsmooth equations, people often assume that the underlying gradient is semismooth in a vicinity of the optimal solution. Due to this, we would like to investigate the piecewise C^2-ness of η around a given point. We now give a definition of piecewise smooth functions below, which is slightly different from the one given in [21].

Definition 2 A continuous function $\psi : \mathbb{R}^n \to \mathbb{R}^l$ is said to be a *piecewise* C^k function on a set $D \subseteq \mathbb{R}^n$ if there exist a finite set of functions ψ_j, $j = 1, \ldots, q$, such that each $\psi_j \in C^k(U)$ for some open set U containing D, and $\psi(u) \in \{\psi_1(u), \ldots, \psi_q(u)\}$ for any $u \in D$.
We refer to $\{\psi_j\}_{i\in\tilde{I}}$ as a representation of ψ on D, where $\tilde{I} = \{1, \ldots, q\}$.

Note that in the case when D is an open set, functions ψ_j, $i = 1, \ldots, q$, in Definition 2 are required to be defined on D only. Since our focus is on the study of the piecewise C^k-ness of η on some open neighborhood around a given point, we seek to find the corresponding pieces defined on this open neighborhood as well.

The following problem will be used to examine properties of η:

$$\tilde{\zeta}(u) = \max_{x \in \mathcal{F}} \left\{ u^T x - \frac{1}{2} \|Ax\|^2_{M^{-1}} - f(x) \right\}, \quad u \in \mathbb{R}^n. \tag{8}$$

Note that u serves as a perturbation parameter of the optimization problem in (8). For $u \in \mathbb{R}^n$, we denote by (P_u) the underlying perturbed problem above. Note also that (P_u) is a concave problem for any given u, and the corresponding Karush–Kuhn–Tucker (KKT) conditions can be written as:

$$\begin{cases} -\nabla f(x) - A^T M^{-1} Ax - \sum_{i \in \hat{I}} \lambda_i \nabla f_i(x) + u = 0, \\ f_i(x) \leq 0, \quad i \in \hat{I}, \\ \lambda_i f_i(x) = 0, \quad \lambda_i \geq 0, \quad i \in \hat{I}. \end{cases} \tag{9}$$

For constraints f_i in $\mathcal{F}$, the active index set is defined by $I(x) := \{i \in \hat{I} \mid f_i(x) = 0\}$. Recall that LICQ is said to hold at $x \in \mathcal{F}$, if $\{\nabla f_i(x) : i \in I(x)\}$ is linearly independent; MFCQ holds at $x \in \mathcal{F}$ if there exists $y \in \mathbb{R}^n$ such that $\nabla f_i(x)^T y < 0$ for all $i \in I(x)$; and CRCQ is said to hold at $x \in \mathcal{F}$, if for every subset $K \subseteq I(x)$, there exists a neighborhood $\mathcal{N}(x)$ of x, such that $\{\nabla f_i(y) : i \in K\}$ has the same rank (dependent on K) for all $y \in \mathcal{N}(x)$. It is known that CRCQ, studied by Janin [7], Pang and Ralph [14] amongst others, is a generalization of the LICQ. Evidently, LICQ implies CRCQ but not vice versa; and LICQ implies MFCQ but not vice versa. It is also known from [7] that CRCQ neither implies nor is implied by MFCQ.

In this paper, we introduce a constraint qualification, called *Generalized Affine Independent Preserving Constraint Qualification (GAIPCQ)* as follows.

Definition 3 The *GAIPCQ* is said to hold at $x \in \mathcal{F}$, if for every subset $K \subseteq I(x)$ for which there exists a sequence $\{x^k\}$ with $\{x^k\} \to x$, $K \subseteq I(x^k)$ and the vectors $\{\nabla f_i(x^k) : i \in K\}$ being linearly independent, it follows that the vectors $\{\nabla f_i(x) : i \in K\}$ are linearly independent.

It is evident that GAIPCQ is weaker than CRCQ as the latter implies the former but not vice versa.

Set

$$V(x, \lambda) := A^T M^{-1} A + \nabla^2 f(x) + \sum_{i \in I(x)} \lambda_i \nabla^2 f_i(x), \tag{10}$$

and

$$G(x, \lambda) := \left\{ d \in \mathbb{R}^n : \nabla f_i(x)^T d = 0, \ \forall\, i \in I_0(x, \lambda) \right\}, \tag{11}$$

where $I_0(x, \lambda) := \{i \in \hat{I} : f_i(x) = 0, \ \ \lambda_i > 0\}$. A point $(x, \lambda) \in \mathbb{R}^{n+\theta}$ is said to satisfy the *strong second-order sufficiency* condition for (8) if it satisfies the first-order KKT conditions (9) and

$$d^T V(x, \lambda) d > 0, \quad \text{for all} \quad d \in G(x, \lambda) \setminus \{0\}. \tag{12}$$

Note that if $I(x) = \emptyset$, the last term in (10) will disappear automatically and the set $G(x, \lambda)$ will become to the whole n-dimensional space.

3 Piecewise C^2-ness of η and semismoothness of g

As discussed earlier, the regularized function η of the Lagrangian-dual function possesses the smoothness property. In this section, we investigate its higher order properties. Specifically, we will study the piecewise C^2-ness of η and the semismoothness of its gradient. First, we make the following assumptions throughout the forthcoming paper.

Assumption 1 $\mathcal{F}$ is nonempty, bounded, and $\mathrm{int}(\mathcal{F}) \neq \emptyset$.

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

The motivation of Assumption 1 is to guarantee the properness of the function $\tilde{\zeta}$ defined in (8) and to make the von Neumann duality applicable as shown by Proposition 1. Assumption 1 also ensures that the optimal solution of perturbed problem (P_u) will lie in the interior of the feasible set $\mathcal{F}$, if not falls on its boundary. In other words, we rule out the situation where the optimal solution lies in the relative interior of $\mathcal{F}$. Assumption 2 is a natural assumption of smoothness. Under these two assumptions, we derive an alternative expression of the Moreau–Yosida regularization function η as follows.

Proposition 1 *For η defined in (5), under Assumptions 1 and 2 we have*

$$\eta(v) = \tilde{\eta}(v) - v^T a - \frac{1}{2}\|a\|^2_{M^{-1}}, \quad v \in \mathbb{R}^m, \tag{13}$$

where $\tilde{\eta}(v) := \max_{x\in\mathcal{F}} \left\{ (v + M^{-1}a)^T Ax - \frac{1}{2}\|Ax\|^2_{M^{-1}} - f(x) \right\}$.

Proof By the definitions of η, φ, and by Assumptions 1 and 2,

$$\begin{aligned}
\eta(v) &= \min_{w\in\mathbb{R}^m} \left\{ \varphi(w) + \frac{1}{2}\|w - v\|^2_M \right\} \\
&= \min_{w\in\mathbb{R}^m} \left\{ \max_{x\in\mathcal{F}} \left\{ -f(x) + w^T(Ax - a) \right\} + \frac{1}{2}\|w - v\|^2_M \right\} \\
&= \min_{w\in\mathbb{R}^m} \max_{x\in\mathcal{F}} \left\{ -f(x) + w^T(Ax - a) + \frac{1}{2}\|w - v\|^2_M \right\}.
\end{aligned}$$

Let $K_1(x, w) := -f(x) + w^T(Ax - a) + \frac{1}{2}\|w - v\|^2_M$. Since f is a convex function, it follows that $K_1(\cdot, \cdot)$ is a continuous, finite, and concave–convex function on $\mathcal{F} \times \mathbb{R}^m$. Further, $\mathcal{F}$ is a closed set due to the continuity of f_i, $i \in \hat{I}$. By Assumption 1 and the well known von Neumann duality theory (see [19, Corollary 37.3.2]), we have

$$\begin{aligned}
\eta(v) &= \min_{w\in\mathbb{R}^m} \max_{x\in\mathcal{F}} \left\{ -f(x) + w^T(Ax - a) + \frac{1}{2}\|w - v\|^2_M \right\} \\
&= \max_{x\in\mathcal{F}} \min_{w\in\mathbb{R}^m} \left\{ -f(x) + w^T(Ax - a) + \frac{1}{2}\|w - v\|^2_M \right\} \\
&= \max_{x\in\mathcal{F}} \left\{ \min_{w\in\mathbb{R}^m} \left\{ w^T(Ax - a) + \frac{1}{2}\|w - v\|^2_M \right\} - f(x) \right\}.
\end{aligned}$$

Let $K_2(w) = w^T(Ax - a) + \frac{1}{2}\|w - v\|^2_M$. Obviously, $\nabla K_2(w) = Ax - a + M(w - v)$. So, the optimal solution for the unconstrained inner minimization problem above is

$w^* = v - M^{-1}(Ax - a)$. Therefore,

$$\begin{aligned}\eta(v) &= \max_{x\in\mathcal{F}}\left\{\min_{w\in\mathbb{R}^m}\left\{w^T(Ax-a)+\frac{1}{2}\|w-v\|_M^2\right\}-f(x)\right\}\\ &= \max_{x\in\mathcal{F}}\left\{\left(v-M^{-1}(Ax-a)\right)^T(Ax-a)+\frac{1}{2}\|M^{-1}(Ax-a)\|_M^2-f(x)\right\}\\ &= \max_{x\in\mathcal{F}}\left\{(v+M^{-1}a)^TAx-\frac{1}{2}\|Ax\|_{M^{-1}}^2-f(x)\right\}-v^Ta-\frac{1}{2}\|a\|_{M^{-1}}^2.\end{aligned}$$

This completes the proof. □

Note that η shares some properties of interest with $\tilde{\eta}$, such as, smoothness or piecewise C^k-ness. Due to this, in what follows, we mainly investigate the properties of $\tilde{\eta}$. In addition, evidently, under Assumption 1, $\mathrm{dom}\tilde{\zeta} = \mathbb{R}^n$ and $\mathrm{dom}\tilde{\eta} = \mathbb{R}^m$, where $\mathrm{dom}h$ represents the *domain* of a function $h : \mathbb{R}^l \to \mathbb{R} \cup \{+\infty\}$ defined by $\mathrm{dom}h := \{z \in \mathbb{R}^l : h(z) < +\infty\}$. Note also that $\tilde{\zeta}$ is a convex function on $\mathbb{R}^n$. Further, by virtue of the definition of $\tilde{\zeta}$ defined in (8), $\tilde{\eta}$ can be rewritten as follows:

$$\tilde{\eta}(v) = \tilde{\zeta}(A^T(v + M^{-1}a)), \quad v \in \mathbb{R}^m.$$

Hence, to investigate function $\tilde{\eta}$, we need to study the properties of $\tilde{\zeta}$ instead in the analysis.

3.1 Second-order properties of $\tilde{\zeta}$

In this subsection, we will investigate some second order properties of the optimal value function $\tilde{\zeta}$ of (P_u), based on the different distribution of its corresponding optimal solution, i.e., the solution falls on the boundary or the interior of the feasible set. For the former case, our attention turns to study second order properties of some auxiliary functions $\tilde{\zeta}_Q$ which are closely related to a facet Q of $\mathcal{F}$.

We first state a result which characterizes the relationship between $\tilde{\eta}$ and $\tilde{\zeta}$ in terms of the piecewise smoothness.

Proposition 2 *Let $\bar{v} \in \mathbb{R}^m$. If $\tilde{\zeta}$ is piecewise C^2 on an open neighborhood $\mathcal{N}(\bar{u})$ of $\bar{u} := A^T\bar{v} + A^TM^{-1}a$, then $\tilde{\eta}$ is piecewise C^2 on an open neighborhood of $\bar{v}$.*

Proof Since $\tilde{\zeta}$ is piecewise C^2 on $\mathcal{N}(\bar{u})$, by definition, there exists functions $\tilde{\zeta}_i \in C^2(\mathcal{N}(\bar{u}))$, $i = 1, \ldots, k$, such that $\tilde{\zeta}(u) \in \{\tilde{\zeta}_1(u), \ldots, \tilde{\zeta}_k(u)\}$ for any $u \in \mathcal{N}(\bar{u})$. Let

$$\tilde{\eta}_i(v) := \tilde{\zeta}_i(A^T(v + M^{-1}a)), \quad \mathcal{N}(\bar{v}) := \{v : A^Tv + A^TM^{-1}a \in \mathcal{N}(\bar{u})\}.$$

Then, it is evident that $\tilde{\eta}_i \in C^2(\mathcal{N}(\bar{v}))$, and $\mathcal{N}(\bar{v})$ is an open neighborhood of $\bar{v}$. Furthermore, $\tilde{\eta}(v) \in \{\tilde{\eta}_1(v), \ldots, \tilde{\eta}_k(v)\}$ for every $v \in \mathcal{N}(\bar{v})$. This completes the proof. □

According to Proposition 2, to study the piecewise smoothness of $\tilde{\eta}$, it is enough to investigate that of $\tilde{\zeta}$. For a given $u \in \mathbb{R}^n$, we consider the following two cases: (i) the optimal solution of (P_u) lies in the interior of $\mathcal{F}$; (ii) the optimal solution of (P_u) lies on the boundary $\mathrm{bd}\mathcal{F}$ of $\mathcal{F}$, respectively.

For a symmetric matrix B, we denote by $B \succ 0$ a symmetric positive definite matrix. We derive the following result concerning case (i).

Proposition 3 *Let $\bar{u} \in \mathbb{R}^n$. Assume that the optimal solution, $\bar{x}$, of $(P_{\bar{u}})$ lies in the interior of $\mathcal{F}$. Suppose that $\nabla^2 f(\bar{x}) + A^TM^{-1}A \succ 0$. Then, there exists an open neighborhood $\mathcal{N}(\bar{u})$ of $\bar{u}$ such that $\tilde{\zeta} \in C^2(\mathcal{N}(\bar{u}))$.*

Proof Since, by assumption, $\bar{x}$ is an interior point of $\mathcal{F}$, then $\bar{x}$ together with a Lagrangian multiplier $\bar{\lambda}$ satisfies the KKT conditions:

$$\begin{cases} \nabla f(\bar{x}) + A^T M^{-1} A\bar{x} = \bar{u}, \\ f_i(\bar{x}) < 0, \quad i \in \hat{I}, \\ \bar{\lambda}_i = 0, \quad i \in \hat{I}. \end{cases} \tag{14}$$

Define a mapping $\Upsilon : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$\Upsilon(x, u) = \nabla f(x) + A^T M^{-1} Ax - u.$$

Then, we have $\Upsilon(\bar{x}, \bar{u}) = 0$. Obviously, Υ is smooth on $\mathbb{R}^n \times \mathbb{R}^n$. Note also that, by assumption, $\nabla_x \Upsilon(x, u) = \nabla^2 f(x) + A^T M^{-1} A$ is nonsingular at $(\bar{x}, \bar{u})$. Hence, by the implicit function theorem, there exists an open neighborhood $\mathcal{N}(\bar{u})$ of $\bar{u}$ and a unique continuously differentiable function $x(u)$ defined on $\mathcal{N}(\bar{u})$ such that $x(\bar{u}) = \bar{x}$, and $\Upsilon(x(u), u) = 0$ for every $u \in \mathcal{N}(\bar{u})$.

On the other hand, shrinking $\mathcal{N}(u)$ when necessary, we have $I(x(u)) = \emptyset$ for $u \in \mathcal{N}(\bar{u})$. For any $u \in \mathcal{N}(\bar{u})$, it is easy to verify that $x(u)$ together with $\lambda = 0 \in \mathbb{R}^\theta$ satisfies the KKT conditions of (P_u), which leads to that $x(u)$ is the unique optimal solution of (P_u). Then,

$$\tilde{\zeta}(u) = -f(x(u)) - \frac{1}{2}\|Ax(u)\|^2_{M^{-1}} + u^T x(u), \quad u \in \mathcal{N}(\bar{u}). \tag{15}$$

Hence, $\tilde{\zeta}$ is smooth on $\mathcal{N}(\bar{u})$ due to the smoothness of $x(u)$.

Next, we further prove that $\tilde{\zeta}$ is twice continuously differentiable on $\mathcal{N}(\bar{u})$. Let $u \in \mathcal{N}(\bar{u})$. For any $u' \in \mathbb{R}^n$, it holds

$$\begin{aligned} \tilde{\zeta}(u') &= \max_{x \in \mathcal{F}} \left\{ u'^T x - \frac{1}{2}\|Ax\|^2_{M^{-1}} - f(x) \right\} \\ &\geq u'^T x(u) - \frac{1}{2}\|Ax(u)\|^2_{M^{-1}} - f(x(u)) \\ &= u^T x(u) - \frac{1}{2}\|Ax(u)\|^2_{M^{-1}} - f(x(u)) + (u' - u)^T x(u) \\ &= \tilde{\zeta}(u) + (u' - u)^T x(u). \end{aligned}$$

Recall that $\tilde{\zeta}$ is a convex function on $\mathbb{R}^n$, thus $x(u)$ is a subgradient of $\tilde{\zeta}$ at u. So,

$$\nabla \tilde{\zeta}(u) = x(u), \quad \forall\, u \in \mathcal{N}(\bar{u}),$$

which, together with the smoothness of $x(u)$ on $\mathcal{N}(\bar{u})$, implies that $\tilde{\zeta} \in C^2(\mathcal{N}(\bar{u}))$ immediately. This completes the proof. □

Remark 1 For $u \in \mathbb{R}^n$, let x_{opt} denote the optimal solution of (P_u). Define

$$\widetilde{U} := \{u \in \mathbb{R}^n : I(x_{\text{opt}}) = \emptyset\}, \qquad \widehat{U} := \mathbb{R}^n \setminus \widetilde{U}.$$

Then, according to the proof of Proposition 3, it is not hard to show that $\widetilde{U}$ is an open set in $\mathbb{R}^n$. Note further that, in this case, we can find a twice continuously differentiable function, denoted by $\tilde{\zeta}_0$, which is defined on the whole space $\mathbb{R}^n$, such that $\tilde{\zeta}_0(u) = \tilde{\zeta}(u)$ for every $u \in \widetilde{U}$.

Next, we investigate the case where the optimal solution of (P_u) lies on the boundary of $\mathcal{F}$. In other words, $u \in \widehat{U}$. We need the following non-degenerate assumption.

Assumption 3 Let $u \in \widehat{U}$. For any KKT point (x, λ) associated with problem (P_u), the Lagrangian multiplier $\lambda \neq 0$.

We now define some notations, which will be used in the sequel. 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. We denote by $|S|$ the *cardinality* of a set S.

For each facet Q of $\mathcal{F}$ with the index set I_Q, we define the following sets with respect to $x \in Q$ and $\lambda \in \mathbb{R}^{|I_Q|}$:

$$\begin{aligned} I_Q(\lambda) &:= \{i \in I_Q : \lambda_i > 0\}, \\ G_Q(x, \lambda) &:= \left\{ d \in \mathbb{R}^n : \nabla f_i(x)^T d = 0, \ \forall\, i \in I_Q(\lambda) \right\}, \\ V_Q(x, \lambda) &:= A^T M^{-1} A + \nabla^2 f(x) + \sum_{i \in I_Q} \lambda_i \nabla^2 f_i(x), \end{aligned}$$

and

$$\begin{aligned} W_Q := \Big\{ &(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^{|I_Q|} : f_i(x) = 0, \ \forall\, i \in I_Q, \ \text{and} \\ &d^T V_Q(x, \lambda) d > 0 \text{ for all } d \in G_Q(x, \lambda) \setminus \{0\} \Big\}, \end{aligned} \tag{16}$$

where $G_Q(x, \lambda) = \mathbb{R}^n$ when $I_Q(\lambda) = \emptyset$.

Let $\tilde{f}(x)$ denote the column vector consisting of $f_i(x)$ for each $i \in I_Q$. Define a mapping $\Gamma_Q : \mathbb{R}^n \times \mathbb{R}^{|I_Q|} \to \mathbb{R}^{n+|I_Q|}$ by

$$\Gamma_Q(x, \lambda) := \begin{pmatrix} \nabla f(x) + \sum_{i \in I_Q} \lambda_i \nabla f_i(x) + A^T M^{-1} A x \\ \tilde{f}(x) \end{pmatrix}. \tag{17}$$

The following lemma plays an important role in our further analysis.

Lemma 1 *Let W_Q, Γ_Q be defined by (16), (17), respectively. Suppose that $\{\nabla f_i(x)\}_{i \in I_Q}$ are linearly independent. Then the Jacobian of Γ_Q is nonsingular at each $(x, \lambda) \in W_Q$.*

Proof Note that for every $(x, \lambda) \in W_Q$, the Jacobian of Γ_Q is

$$\nabla \Gamma_Q(x, \lambda) = \begin{pmatrix} V_Q(x, \lambda) & \nabla \tilde{f}_{I_0}(x) & \nabla \tilde{f}_{I_1}(x) \\ \nabla \tilde{f}_{I_0}(x)^T & 0 & 0 \\ \nabla \tilde{f}_{I_1}(x)^T & 0 & 0 \end{pmatrix},$$

where $\nabla \tilde{f}_{I_0}(x) := \{\nabla f_i(x)\}_{i \in I_0}$, $\nabla \tilde{f}_{I_1}(x) := \{\nabla f_i(x)\}_{i \in I_1}$, $I_0 := I_Q(\lambda)$, and $I_1 := I_Q \setminus I_0$. Let

$$\nabla \Gamma_Q(x, \lambda) \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = 0,$$

where $d_1 \in \mathbb{R}^n$, $d_2 \in \mathbb{R}^{|I_0|}$, and $d_3 \in \mathbb{R}^{|I_1|}$. Then, it follows that

$$0 = V_Q(x, \lambda) d_1 + \nabla \tilde{f}_{I_0}(x) d_2 + \nabla \tilde{f}_{I_1}(x) d_3, \tag{18}$$
$$0 = \nabla \tilde{f}_{I_0}(x)^T d_1, \tag{19}$$
$$0 = \nabla \tilde{f}_{I_1}(x) d_1. \tag{20}$$

By (19), $d_1 \in G_Q(x,\lambda)$. Multiplying (18) by d_1^T, we have

$$d_1^T V_Q(x,\lambda)d_1 + d_1^T \nabla \tilde{f}_{I_0}(x)d_2 + d_1^T \nabla \tilde{f}_{I_1}(x)d_3 = 0.$$

Then, by virtue of (19) and (20),

$$d_1^T V_Q(x,\lambda)d_1 = 0.$$

By the definition of W_Q, $d_1 = 0$. So, (18) yields that

$$\nabla \tilde{f}_{I_0}(x)d_2 + \nabla \tilde{f}_{I_1}(x)d_3 = 0.$$

Since $\{\nabla f_i(x)\}_{i\in I_Q}$ are linearly independent, $d_2 = 0$ and $d_3 = 0$. Thereby, $d_1 = 0$, $d_2 = 0$, and $d_3 = 0$, which leads to the nonsingularity of $\nabla\Gamma_Q(x,\lambda)$. This completes the proof. □

By virtue of Lemma 1, we obtain the following result.

Lemma 2 *Let W_Q be given by (16). Suppose that $\{\nabla f_i(x)\}_{i\in I_Q}$ are linearly independent. For $(\bar{x},\bar{\lambda}) \in W_Q$, let $\Gamma_Q(\bar{x},\bar{\lambda}) = \begin{pmatrix}\bar{u}\\0\end{pmatrix}$. Then, there exist open neighborhoods $\mathcal{N}(\bar{u})$ of $\bar{u}$, $\mathcal{N}(\bar{x},\bar{\lambda})$ of $(\bar{x},\bar{\lambda})$ such that, when $u \in \mathcal{N}(\bar{u})$, the equation*

$$\Gamma_Q(x,\lambda) = \begin{pmatrix}u\\0\end{pmatrix}$$

has a unique solution $\xi(u) = (\xi_x,\xi_\lambda)(u) \in \mathcal{N}(\bar{x},\bar{\lambda})$. Further, $(\xi_x,\xi_\lambda)(u)$ is continuously differentiable on $\mathcal{N}(\bar{u})$.

Proof Define a mapping $\digamma_Q(x,\lambda,u) := \Gamma_Q(x,\lambda) - \begin{pmatrix}u\\0\end{pmatrix}$. Clearly, $\digamma_Q$ is smooth on $\mathbb{R}^n \times \mathbb{R}^{|I_Q|} \times \mathbb{R}^m$, and

$$\digamma_Q(\bar{x},\bar{\lambda},\bar{u}) = \Gamma_Q(\bar{x},\bar{\lambda}) - \begin{pmatrix}\bar{u}\\0\end{pmatrix} = 0.$$

Further, according to Lemma 1, $\nabla_{(x,\lambda)}\digamma_Q(x,\lambda,u) = \nabla\Gamma_Q(x,\lambda)$ is nonsingular at $(\bar{x},\bar{\lambda},\bar{u})$. Hence, the desired results follow immediately by virtue of the implicit function theorem. This completes the proof. □

As a consequence of Lemma 2, we obtain the following result.

Proposition 4 *For $(\bar{x},\bar{\lambda}) \in W_Q$, let $\bar{u} = \nabla f(\bar{x}) + \sum_{i\in I_Q} \bar{\lambda}_i \nabla f_i(\bar{x}) + A^T M^{-1} A\bar{x}$. Define*

$$\tilde{\zeta}_Q(u) := -f(\xi_x(u)) - \frac{1}{2}\|A\xi_x(u)\|^2_{M^{-1}} + u^T\xi_x(u), \quad u \in \mathcal{N}(\bar{u}),$$

where $\xi_x(u)$ and $\mathcal{N}(\bar{u})$ are defined as Lemma 2. Then, (i) for $u \in \mathcal{N}(\bar{u})$, $\nabla\tilde{\zeta}_Q(u) = \xi_x(u)$; (ii) $\tilde{\zeta}_Q \in C^2(\mathcal{N}(\bar{u}))$.

Proof From Lemma 2 and the first equation of $\Gamma_Q(\xi(u)) - \begin{pmatrix}\bar{u}\\0\end{pmatrix} = 0$, it follows that

$$u = \nabla f(\xi_x(u)) + \sum_{i\in I_Q} \xi_{\lambda i}(u)\nabla f_i(\xi_x(u)) + A^T M^{-1} A\xi_x(u).$$

So,

$$
\begin{aligned}
\nabla\tilde{\zeta}_Q(u) &= -\nabla\xi_x(u)\nabla f(\xi_x(u)) + \xi_x(u) + \nabla\xi_x(u)u - \nabla\xi_x(u)A^T M^{-1}A\xi_x(u)\\
&= \xi_x(u) + \nabla\xi_x(u)\left[-\nabla f(\xi_x(u)) - A^T M^{-1}A\xi_x(u) + u\right]\\
&= \xi_x(u) + \nabla\xi_x(u)\Bigg[- \nabla f(\xi_x(u)) - A^T M^{-1}A\xi_x(u) + \nabla f(\xi_x(u))\\
&\qquad + \sum_{i\in I_Q}\xi_{\lambda_i}(u)\nabla f_i(\xi_x(u)) + A^T M^{-1}A\xi_x(u)\Bigg]\\
&= \xi_x(u) + \nabla\xi_x(u)\sum_{i\in I_Q}\xi_{\lambda_i}(u)\nabla f_i(\xi_x(u))\\
&= \xi_x(u) + \sum_{i\in I_Q}\xi_{\lambda_i}(u)\nabla\xi_x(u)\nabla f_i(\xi_x(u)).
\end{aligned}
$$

Since $f_i(\xi_x(u)) = 0$ for every $u \in \mathcal{N}(\bar{u})$ and each $i \in I_Q$, differentiating these functions, we have $\nabla\xi_x(u)\nabla f_i(\xi_x(u)) = 0, \forall\ i \in I_Q$. Hence,

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

which leads to that $\nabla\tilde{\zeta}_Q(u) = \xi_x(u)$. Again, by Lemma 2, $\xi_x(u)$ is continuously differentiable on $\mathcal{N}(\bar{u})$. Therefore, $\tilde{\zeta}_Q(u)$ is twice continuously differentiable on $\mathcal{N}(\bar{u})$. □

In this subsection, we show that $\tilde{\zeta}$ is twice continuously differentiable on an open neighborhood of u when $u \in \widetilde{U}$. In the case of $u \in \widehat{U}$, we define function $\tilde{\zeta}_Q$ associated with the facet Q of $\mathcal{F}$ and prove that $\tilde{\zeta}_Q$ is twice continuously differentiable in a vicinity of u as well. These functions $\tilde{\zeta}_Q$, as we shall prove in next subsection, actually serve as the pieces of a representation for $\tilde{\zeta}$, which then implies the piecewise C^2-ness of $\tilde{\zeta}$ for the latter case, i.e., $u \in \widehat{U}$.

3.2 Piecewise C^2-ness of $\tilde{\zeta}$

In this subsection, we study the piecewise C^2-ness of $\tilde{\zeta}$. We will use the following result concerning the continuity of the local optimal solution of perturbed problem (8), which is taken from [18, Theorem 3.2].

Proposition 5 *Let $u \in \mathbb{R}^n$ be given. Suppose that for some $x \in \mathcal{F}$, MFCQ holds at x. Suppose further that for each (x, λ) satisfying the KKT conditions (9), the strong second-order sufficient condition holds at (x, λ). Then, x is a strict local optimal solution of (P_u). And, the optimal solution to $(P_{u'})$ tends to x as $u' \to u$.*

For any $u \in \widehat{U}$, let x be the optimal solution of (P_u). Let $\mathcal{M}(u)$ denote the set of multipliers $\lambda \in \mathbb{R}^\theta$ such that x together with λ satisfies the KKT conditions (9). For a nonnegative vector $\alpha \in \mathbb{R}^l$, we denote by $\mathrm{supp}(\alpha)$ the *support* of α which is defined by $\mathrm{supp}(\alpha) := \{i : \alpha_i > 0,\ i = 1, \ldots, l\}$. We define an index set

$$
\begin{aligned}
\mathcal{B}(u) := \{K \mid\ &\text{there exists } \lambda \in \mathcal{M}(u) \text{ such that } \mathrm{supp}(\lambda) \subseteq K \subseteq I(x),\\
&\text{and the vectors } \{\nabla f_i(x) : i \in K\} \text{ are linearly independent}\}.
\end{aligned}
$$

Note that, under the assumption of MFCQ at x, $\mathcal{B}(u)$ is nonempty since $\mathcal{M}(u)$ has an extreme point, which can easily yield an index set K as defined in $\mathcal{B}(u)$.

Lemma 3 *Let $u \in \widehat{U}$ and $x(u)$ be the optimal solution of (P_u). Suppose the conditions of Proposition 5 hold. Suppose further that GAIPCQ is satisfied at $x(u)$. Then, there exists an open neighborhood $\mathcal{N}_1(u)$ of u such that $\mathcal{B}(w) \subseteq \mathcal{B}(u)$ for any $w \in \mathcal{N}_1(u)$.*

Proof To ease the notation, let $x(w)$ denote the optimal solution of (P_w). Evidently, by definition, $|I(x(u))| \geq 1$ since $u \in \widehat{U}$. According to Proposition 5, for any w close to u,

$$I(x(w)) \subseteq I(x(u)). \tag{21}$$

Suppose on the contrary that the proposed result is false. Noticing that $I(x(u))$ is a finite index set, then there exists a sequence $\{u^k\}$ tending to u such that there exists an index set K satisfying $K \in \mathcal{B}(u^k)\setminus\mathcal{B}(u)$ for all k. So, for each k, the vectors $\{\nabla f_i(x(u^k)) : i \in K\}$ are linearly independent, and there exists $\lambda(u^k) \in \mathcal{M}(u^k)$ such that $\text{supp}(\lambda(u^k)) \subseteq K \subseteq I(x(u^k))$, but $K \notin \mathcal{B}(u)$. Then, by GAIPCQ, the vectors $\{\nabla f_i(x(u)) : i \in K\}$ must be linearly independent. Moreover, it follows from (21) that $K \subseteq I(x(u))$. Since $K \notin \mathcal{B}(u)$, we then derive the following statement:

$$\textit{There does not exist } \lambda(x(u)) \in \mathcal{M}(u) \textit{ such that } \text{supp}(\lambda(u)) \subseteq K \subseteq I(x(u)). \tag{22}$$

On the other hand, by the definition of $\lambda(x(u^k))$,

$$-\nabla f(x(u^k)) - A^T M^{-1} A x(u^k) - \sum_{i\in K} \lambda_i(x(u^k))\nabla f_i(x(u^k)) + u^k = 0. \tag{23}$$

Since $\{\nabla f_i(x(u^k)) : i \in K\}$ are linearly independent and $\nabla f_i(x(u^k)) \to \nabla f_i(x(u))$ as $k \to \infty$, the sequence of vectors $\{(\lambda_i(x(u^k)))_{i\in K}\}$ is bounded. Thus, $\{(\lambda(x(u^k)))_{i\in K}\}$ must have an accumulation point, say, $\bar{\lambda}^K(x(u))$. It then follows from (23) that

$$-\nabla f(x(u)) - A^T M^{-1} A x(u) - \sum_{i\in K} \bar{\lambda}_i^K(x(u))\nabla f_i(x(u)) + u = 0.$$

Define

$$\lambda_i(x(u)) = \begin{cases} \bar{\lambda}_i^K(x(u)), & \text{if } i \in K, \\ 0, & \text{if } i \in I(x(u)) \setminus K, \end{cases}$$

then, $\lambda(x(u)) \in \mathcal{M}(u)$ and $\text{supp}(\lambda(x(u))) \subseteq K \subseteq I(x(u))$, which leads to a contradiction with (22). This completes the proof. □

For $u \in \widehat{U}$, we study the piecewise C^2-ness of $\tilde{\zeta}$ on an open neighborhood of u for the following two cases: (i) $u \in \text{int}\widehat{U}$; (ii) $u \in \text{bd}\widehat{U}$, respectively.

Proposition 6 *For $u \in \text{int}\widehat{U}$, let x be the optimal solution of (P_u). Suppose that (a) both GAIPCQ and MFCQ hold at x; (b) the strong second-order sufficient condition holds at each KKT point of problem (P_u). Then, there exist an open neighborhood $\mathcal{N}(u)$ of u and functions $\tilde{\zeta}_i(\cdot)$ defined on $\mathcal{N}(u)$ such that $\tilde{\zeta}_i \in C^2(\mathcal{N}(u))$, $i = 1, \ldots, k$, and $\tilde{\zeta}(w) \in \{\tilde{\zeta}_1(w), \ldots, \tilde{\zeta}_k(w)\}$ for every $w \in \mathcal{N}(u)$. That is, $\tilde{\zeta}$ is piecewise C^2 on $\mathcal{N}(u)$.*

Proof For any $K \in \mathcal{B}(u)$, there exists $(x, \lambda_K(u))$ with $\lambda_K(u) \in \mathcal{M}(u)$ such that x^* lies on the facet, denoted by $Q_K := \{x \in \mathcal{F} \mid f_i(x) = 0, i \in K\}$. Then, by the strong second-order sufficiency condition, it follows that $(x^*, \lambda_K(u)) \in W_{Q_K}$, where W_{Q_K} is defined as in (16). Also, by the definition of $\mathcal{B}(u)$, $\{\nabla f_i(x^*) : i \in K\}$ are linearly independent. Hence, by

Lemma 2, there exist open neighborhoods $\mathcal{N}_K(u)$ of u, $\mathcal{N}_K(x^*, \lambda_K(u))$ of $(x^*, \lambda_K(u))$, and a continuously differentiable function $\xi^K(w) = (\xi_x^K, \xi_\lambda^K(w)) \in \mathcal{N}_K(x^*, \lambda_K(u))$ satisfying

$$\Gamma_{Q_K}(\xi_x^K(w), \xi_\lambda^K(w)) = \begin{pmatrix} w \\ 0 \end{pmatrix}$$

for all $w \in \mathcal{N}_K(u)$. Set

$$\tilde{\zeta}_{Q_K}(u) := -f(\xi_x^K(u)) - \frac{1}{2}\|A\xi_x^K(u)\|_{M^{-1}}^2 + u^T\xi_x^K(u).$$

Then, by Proposition 4, $\tilde{\zeta}_{Q_K} \in C^2(\mathcal{N}_K(u))$. Since $\mathcal{B}(u)$ is a finite index set, we hence derive a finitely many set of such functions $\tilde{\zeta}_{Q_K} : \mathcal{N}_K(u) \to \mathbb{R}^n$ with $\tilde{\zeta}_{Q_K} \in C^2(\mathcal{N}_K(u))$, $K \in \mathcal{B}(u)$.

On the other hand, by Lemma 3, for any $w \in \mathcal{N}_1(u)$, there exist $K \in \mathcal{B}(w) \subseteq \mathcal{B}(u)$ and $(x(w), \lambda_K(w))$ such that

$$\begin{cases} w = \nabla f(x(w)) + A^T M^{-1} A x(w) + \sum_{i\in K} \lambda_{Ki}(w)\nabla f_i(x(w)) \\ \lambda_{Ki}(w) \geq 0, \quad i \in K, \\ f_i(x(w)) \leq 0, \quad i \in K, \\ \lambda_{Ki}(w) f_i(x(w)) = 0, \quad i \in K. \end{cases} \tag{24}$$

It is not hard to show that $x(w)$ is the optimal solution of (P_w) and $x(w) \to x(u)$ as $w \to u$, where $x(u) = x$. By virtue of the linear independence of $\{\nabla f_i(x(w)) : i \in K\}$ and the continuity of $x(\cdot)$ and $\nabla f_i(\cdot)$, together with the first equation in (24), we have $\lambda_K(w) \to \lambda_K(u)$ as $w \to u$. Thus, we can choose an open neighborhood $\mathcal{N}(u)$ of u with $\mathcal{N}(u) \subseteq \big((\cap_{K\in\mathcal{B}(u)}) \cap \mathcal{N}_1(u) \cap \mathrm{int}\widehat{U}\big)$ such that $(x(w), \lambda_K(w)) \in \mathcal{N}_K(x^*, \lambda_K(u))$ for any $w \in \mathcal{N}(u)$.

For $K \in \mathcal{B}(u)$, let $V_K(u) = \{w \in \mathcal{N}(u) : K \in \mathcal{B}(w)\}$. Then, $\mathcal{N}(u) = \cup_{K\in\mathcal{B}(u)} V_K(u)$. So, for each $w \in \mathcal{N}(u)$, there exists $K \in \mathcal{B}(u)$ such that $\Gamma_{Q_K}(x(w), \lambda_K(w)) = \begin{pmatrix} w \\ 0 \end{pmatrix}$ and $(x(w), \lambda_K(w)) \in \mathcal{N}_K(x^*, \lambda_K(u))$. By the implicit function theorem, it follows that

$$(x(w), \lambda_K(w)) = (\xi_x^K(w), \xi_\lambda^K(w)),$$

which implies that $\xi_x^K(w)$ is an optimal solution of (P_w), $w \in \mathcal{N}(w)$. Then, the corresponding optimal value function $\tilde{\zeta}(w)$ is as follows:

$$\tilde{\zeta}(w) = \tilde{\zeta}_{Q_K}(w) = -f(\xi_x^K(w)) - \frac{1}{2}\|A\xi_x^K(w)\|_{M^{-1}}^2 + w^T\xi_x^K(w), \quad w \in \mathcal{N}(u).$$

Thus, $\tilde{\zeta}(w) \in \{\tilde{\zeta}_{Q_K}(w) : K \in \mathcal{B}(u)\}$ for every $w \in \mathcal{N}(u)$. Note that $\mathcal{B}(u)$ is a finite set, so there are finitely many such functions $\tilde{\zeta}_{Q_K}$. Thereby, $\tilde{\zeta}$ is piecewise C^2 on $\mathcal{N}(u)$. This completes the proof. □

Next, we consider the case where $u \in \mathrm{bd}\widehat{U}$ and derive the piecewise C^2-ness of $\tilde{\zeta}$ as follows.

Proposition 7 *For $u \in \mathrm{bd}\widehat{U}$, let x be the optimal solution of (P_u). Suppose all the conditions in Proposition 6 are satisfied. Then there exist an open neighborhood $\mathcal{N}(u)$ of u, a finitely many set of functions $\tilde{\zeta}_i \in C^2(\mathcal{N}(u))$, $i = 0, 1, \ldots, l$, such that*

$$\tilde{\zeta}(w) \in \{\tilde{\zeta}_0(w), \tilde{\zeta}_1(w), \ldots, \tilde{\zeta}_l(w)\}, \quad \forall\, w \in \mathcal{N}(u).$$

Proof In this case, the whole proof is similar to that of Proposition 6. The only problem comes from the points which lie in the intersection of some neighborhood of u and $\widetilde{U}$, i.e., $\{w \in \mathbb{R}^n \mid w \in \mathcal{N}(u) \cap \widetilde{U}\}$. For sake of brevity, we will roughly describe the proof. With similar arguments as in Proposition 6 and noticing the only required concern is that the optimal solution $x \in \mathrm{bd}\mathcal{F}$, but no any restriction on u, we then obtain $\tilde{\zeta}_i \in C^2(\mathcal{N}_i(u))$ and $\mathcal{N}(u) \subseteq \cap_{i\in I_1}\mathcal{N}_i(u)$, where $\mathcal{N}_i(u)$ is an open neighborhood of u and I_1 is a finite index set, which is denoted by $\{1, \ldots, l\}$. Now, for any $w \in \mathcal{N}(u)$, there are two cases needed to consider:

Case i. $w \in \widehat{U}$. Then, the corresponding optimal solution $x(w) \in \mathrm{bd}\mathcal{F}$. Thereby, there exists $i \in I_1$ such that $\tilde{\zeta}(w) = \tilde{\zeta}_i(w)$ as discussed in Proposition 6.

Case ii. $w \in \widetilde{U}$. According to Remark 1, there exists a function $\tilde{\zeta}_0 \in C^2(\mathbb{R}^n)$, such that $\tilde{\zeta}(w) = \tilde{\zeta}_0(w)$.

We define an index set $\tilde{I} := I_1 \cup \{0\}$. Then, based on the above arguments, we have $\tilde{\zeta}(w) \in \{\tilde{\zeta}_i(w)\}_{i\in\tilde{I}}$ for any $w \in \mathcal{N}(u)$. Thereby, $\tilde{\zeta}$ is piecewise C^2 on $\mathcal{N}(u)$. This completes the proof. □

3.3 Piecewise C^2-ness of η

With the arguments in previous subsections, we are ready to derive the piecewise C^2-ness of η and the semismoothness of its gradient g.

Proposition 8 *For any $u \in \mathbb{R}^n$, let x be the optimal solution of (P_u). Suppose MFCQ and GAIPCQ both hold at x. Suppose further that the strong second-order sufficiency condition holds at every KKT point of (P_u). Then, for any $v \in \mathbb{R}^m$, the Moreau–Yosida regularization η is piecewise C^2 on an open neighborhood of v.*

Proof For any $v \in \mathbb{R}^m$, let $u := A^T v + A^T M^{-1} a \in \mathbb{R}^n$. Then, by the assumption and Propositions 3, 6, and 7, $\tilde{\zeta}$ is piecewise C^2 on some neighborhood $\mathcal{N}(u)$ with a representation $\{\tilde{\zeta}_j\}_{j\in\tilde{I}}$, where $\tilde{I}$ is a finite index set. By virtue of Proposition 2, this representation induces a representation $\{\tilde{\eta}_j\}_{j\in I}$ of $\tilde{\eta}$ and there exists an open neighborhood $\mathcal{N}(v)$ of v such that $\tilde{\eta}$ is piecewise C^2 on $\mathcal{N}(v)$. Therefore, by Proposition 1, η is piecewise C^2 on $\mathcal{N}(v)$ as well. □

By Proposition 8, we obtain the following result immediately.

Proposition 9 *Suppose that the assumptions in Proposition 8 are all satisfied. Then, for any $v \in \mathbb{R}^m$, the gradient g of the Moreau–Yosida regularization η is piecewise smooth on an open neighborhood $\mathcal{N}(v)$ of v. Furthermore, g is semismooth on $\mathcal{N}(v)$.*

Proof For any $v \in \mathbb{R}^m$, it follows from Proposition 8 that η is piecewise C^2 on an open neighborhood $\mathcal{N}(v)$ of v. In addition, since η is smooth on $\mathbb{R}^m$, thus, $g(v) = \nabla\eta(v)$ is piecewise smooth on $\mathcal{N}(v)$ as well, which implies that g is semismooth on $\mathcal{N}(v)$. □

4 BD-regularity of ∂g

In the previous section, we investigate the semismoothness of the gradient g of the regularized function η. In order to obtain the superlinear convergence of the generalized Newton method for solving nonsmooth equation, people also assume that all matrices in ∂g to be positive definite. In this section, we will study under which conditions on the objective function f and constraints f_j, $j \in \hat{I}$, all elements in the generalized Jacobian of g are positive definite. First, we recall the definition of the Clarke generalized Jacobian.

Definition 4 Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz continuous function. The B(ouligand)-subdifferential of F at $x \in \mathbb{R}^n$, denoted by $\partial_B F(x)$, is set of V such that

$$V = \lim_{k\to\infty} JF(x^k),$$

where $\{x^k\} \in D_F$ is a sequence converging to x and D_F denotes the set where F is differentiable. The Clarke generalized Jacobian of F at x [2], denoted by $\partial F(x)$, is the convex hull of $\partial_B F(x)$, i.e.,

$$\partial F(x) = \mathrm{conv}\,\{\partial_B F(x)\}.$$

We say F is BD-regular at x if every $V \in \partial_B F(x)$ is nonsingular, and CD-regular at x if every $V \in \partial F(x)$ is nonsingular.

Remark 2 Note for any $v \in D_g$, $Jg(v)$ is symmetric and positive semidefinite since g is the gradient mapping of the convex function η. So, every $V \in \partial_B g(v)$ is symmetric and positive semidefinite. Then, all $V \in \partial g(v)$ are positive definite if g is BD-regular at v. This implies that, to study the positive definiteness of ∂g, it suffices to investigate the BD-regularity of g.

By Proposition 1, $g(v) = \nabla\eta(v) = \nabla\tilde{\eta}(v) - a$ for any $v \in \mathbb{R}^m$. Then, $\partial g(v) = \partial\tilde{g}(v)$ where $\tilde{g}(v) := \nabla\tilde{\zeta}(v)$. By Proposition 8, $\tilde{g}$ is piecewise smooth on an open neighborhood $\mathcal{N}(v)$ of v with a representation $\{\nabla\tilde{\eta}_j\}_{j\in I}$, I is a finite index set. In addition, by [14],

$$\partial g(v) = \partial\tilde{g}(v) \subseteq \mathrm{conv}\{\nabla^2\tilde{\eta}_j(v) : j \in I(v)\}, \tag{25}$$

where $I(v) = \{j \in I : \tilde{g}(v) = \nabla\tilde{\eta}_j(v)\}$.

Proposition 10 *Let $\xi_x(u)$, $\xi_\lambda(u)$ be defined as in Lemma 2. Then,*

$$\begin{aligned}\nabla\xi_x(u) &= S(S^T HS)^{-1}S^T,\\ \nabla\xi_\lambda(u) &= (R^T R)^{-1}\left(R^T - R^T HS(S^H S)^{-1}S^T\right),\end{aligned}$$

where

$$H = A^T M^{-1}A + \nabla^2 f(\xi_x(u)) + \sum_{i\in I_Q} \xi_{\lambda i}(u)\nabla^2 f_i(\xi_x(u)), \tag{26}$$

$$R = (\nabla f_i(\xi_x(u)))_{i\in I_Q} \in \mathbb{R}^{n\times|I_Q|}, \tag{27}$$

and S is a matrix whose column vectors are orthogonal and the spanning space of all column vectors is the null space of matrix R, Null(R).

Proof By Lemma 2, the solution to equation

$$\Gamma_Q(x,\lambda) = \begin{pmatrix} A^T M^{-1}Ax + \nabla f(x) + \sum_{i\in I_Q}\lambda_i\nabla f_i(x) \\ \tilde{f}(x)\end{pmatrix} = \begin{pmatrix} u\\ 0\end{pmatrix} \tag{28}$$

is $(x,\lambda) = (\xi_x(u), \xi_\lambda(u))$. That is,

$$\Gamma_Q(\xi_x(u), \xi_\lambda(u)) = \begin{pmatrix} u\\ 0\end{pmatrix}.$$

Differentiating the above equation with respect to u, we obtain

$$\nabla\Gamma_Q(\xi_x(u), \xi_\lambda(u)) \begin{pmatrix} \nabla\xi_x(u) \\ \nabla\xi_\lambda(u) \end{pmatrix} = \begin{pmatrix} I_{n\times n} \\ 0 \end{pmatrix},$$

where

$$\nabla\Gamma_Q(\xi_x(u), \xi_\lambda(u)) := \begin{pmatrix} H & R \\ R^T & 0 \end{pmatrix}.$$

Then, we have

$$H\nabla\xi_x(u) + R\nabla\xi_\lambda(u) = I_{n\times n}, \tag{29}$$
$$R^T\nabla\xi_x(u) = 0. \tag{30}$$

Let $L_Q(u) = \left\{d \in \mathbb{R}^n : \nabla f_i(\xi_x(u))^T d = 0,\ i \in I_Q\right\}$ and S be a matrix consisting of orthogonal column vectors which span the subspace $L_Q(u)$. In fact, $L_Q(u)$ is the null space of R. By definition, $R^T S = 0$. So,

$$\nabla\xi_x(u) = S\nabla\xi_{xS}(u) + R\nabla\xi_{xR}(u), \tag{31}$$

where $\nabla\xi_{xS}(u) \in \mathbb{R}^{(n-|I_Q|)\times n}$ and $\nabla\xi_{xR}(u) \in \mathbb{R}^{|I_Q|\times n}$. Multiplying on the left by R^T of both sides of (31), with the help of (30), we get

$$R^T\nabla\xi_x(u) = R^T S\nabla\xi_{xS}(u) + R^T R\nabla\xi_{xR}(u) = 0.$$

In addition, as $R^T R$ is nonsingular, $\nabla\xi_{xR}(u) = 0$. It yields that

$$\nabla\xi_x(u) = S\nabla\xi_{xS}(u). \tag{32}$$

Then, (29) can be rewritten as $HS\nabla\xi_{xS}(u) + R\nabla\xi_\lambda(u) = I_{n\times n}$. Multiplying the above equation on the left by S^T, we have

$$S^T HS\nabla\xi_{xS}(u) + S^T R\nabla\xi_\lambda(u) = S^T.$$

Again, by virtue of $R^T S = 0$, it yields that

$$S^T HS\nabla\xi_{xS}(u) = S^T. \tag{33}$$

Since

$$\begin{aligned} S^T HS &= S^T(A^T M^{-1}A)S + S^T\nabla^2 f(\xi_x(u))S + \sum_{i\in I_Q} \xi_{\lambda i}(u)S^T\nabla^2 f(\xi_x(u))S \\ &= S^T\left(A^T M^{-1}A + \nabla^2 f(\xi_x(u)) + \sum_{i\in I_Q(\lambda)} \xi_{\lambda i}(u)\nabla^2 f_i(\xi_x(u))\right)S \\ &\quad + \sum_{i\in I_Q\setminus I_Q(\lambda)} \xi_{\lambda i}(u)S^T\nabla^2 f_i(\xi_x(u))S, \end{aligned}$$

by the definition of W_Q, the first term of in the above equation is positive definite and the second term is positive semidefinite, hence, $S^T HS$ is positive definite. It follows from (33) that $\nabla\xi_{xS}(u) = (S^T HS)^{-1}S^T$. By virtue of (32),

$$\nabla\xi_x(u) = S(S^T HS)^{-1}S^T. \tag{34}$$

Multiplying (29) on the left by R^T gives

$$R^T H\nabla\xi_x(u) + R^T R\nabla\xi_\lambda(u) = R^T.$$

So,

$$\nabla\xi_\lambda(u) = (R^T R)^{-1}(R^T - R^T H\nabla\xi_x(u)).$$

By (34), we obtain

$$\nabla\xi_\lambda(u) = (R^T R)^{-1}(R^T - R^T HS(S^T HS)^{-1}S^T).$$

This completes the proof. □

Proposition 11 *Let $x(u)$ and $\tilde{\zeta}(u)$ be defined as in Proposition 3. Then,*

$$\nabla^2\tilde{\zeta}(u) = \nabla x(u) = \left(\nabla^2 f(x(u)) + A^T M^{-1}A\right)^{-1}, \quad u \in \mathcal{N}(\bar{u}),$$

where $\bar{u} \in \widetilde{U}$ as defined in (1) and $\mathcal{N}(\bar{u})$ is defined as in Proposition 3.

Proof By Proposition 3, the solution $x(u)$ to equation

$$\Upsilon(x, u) = \nabla f(x) + A^T M^{-1}Ax - u = 0,$$

is smooth on $\mathcal{N}(\bar{u})$. Differentiating the above equation with respect to u, we get

$$(\nabla^2 f(x(u)) + A^T M^{-1}A)\nabla x(u) - I_{n\times n} = 0.$$

Thereby, $\nabla^2\tilde{\zeta}(u) = \nabla x(u) = \left(\nabla^2 f(x(u)) + A^T M^{-1}A\right)^{-1}$. This completes the proof. □

Remark 3 Note that for $\bar{u} \in \widehat{U}$, by Proposition 4, the Hessian of $\tilde{\zeta}_Q(u)$ equals to $\nabla\xi_x(u)$ on a neighborhood $\mathcal{N}(\bar{u})$ since $\nabla\tilde{\zeta}_Q(u) = \xi_x(u)$. Thereby, by Proposition 10, we derive an expression of $\nabla^2\tilde{\zeta}_Q(u)$ for any facet Q of $\mathcal{F}$. In the case of $\bar{u} \in \widetilde{U}$, Proposition 11 provides an expression of $\nabla^2\tilde{\zeta}(u)$ in a vicinity of $\bar{u}$. Thus, by Propositions 2 and 8, we can derive expressions of the Hessians of all pieces for η (or $\tilde{\eta}$) on some neighborhood of $v \in \mathbb{R}^m$.

Lemma 4 *Let $C = ADA^T$, where $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = m \le n$, $D \in \mathbb{R}^{n\times n}$, and $D^T = D$. Then, C is positive definite on $\mathbb{R}^m$ if and only if D is positive definite on the subspace $\mathcal{R}(A^T) := \{A^T d \mid \forall\, d \in \mathbb{R}^m\}$.*

Proof Because C is positive definite if and only if $\forall\ d \neq 0 \in \mathbb{R}^m$, $d^T Cd > 0$, which is equivalent to $(A^T d)^T D(A^T d) > 0$ for every $d \neq 0 \in \mathbb{R}^m$. Since $\operatorname{rank}(A^T) = m$, there exist an orthogonal matrix P and a nonsingular upper triangular matrix R such that $A^T = P\begin{pmatrix} R \\ 0 \end{pmatrix}$. Let $\alpha = A^T d$, it the follows that $\alpha = P\begin{pmatrix} Rd \\ 0 \end{pmatrix}$. Note that $d \neq 0 \Leftrightarrow Rd \neq 0 \Leftrightarrow \alpha \neq 0$. Hence, the above statement is further equivalent to $\alpha^T D\alpha > 0$ for any $0 \neq \alpha \in \mathcal{R}(A^T)$. Namely, D is positive definite on $\mathcal{R}(A^T)$. This completes the proof. □

By Lemma 4, we derive the following result on the positive definiteness of elements in ∂g.

Proposition 12 *For $u \in \mathbb{R}^n$, let x be the optimal solution of (P_u). Suppose that MFCQ and GAIPCQ hold at x. Suppose further that the strong second order sufficiency condition holds at each KKT point of (P_u). Suppose that for any facet Q_j of $\mathcal{F}$ with the index set I_{Q_j}*

$$C^j = S_j(S_j^T H_j S_j)^{-1} S_j^T \text{ is positive definite on } \mathcal{R}(A^T),$$

where H_j, S_j, and R_j are defined as Proposition 10 with respect to facet Q_j. Then, g is BD-regular at every $v \in \mathbb{R}^m$. Thereby, all $V \in \partial g(v)$ are positive definite.

Proof By the definition of BD-regularity and (25), it is enough to show that $\nabla^2\tilde{\eta}_j(v)$ is positive definite for any $j \in I(v)(\subseteq I)$, where $\{\tilde{\eta}_j\}_{j\in I}$ is a representation of $\tilde{\eta}$ on a neighborhood around v and I is a finite index set. For any $v \in \mathbb{R}^m$, let $u = A^T(v + M^{-1}a)$. We consider the following three cases.

Case i. $u \in \widetilde{U}$. By Proposition 3, $|I| = 1$, which means that $\tilde{\eta}$ has only one piece generated by that of $\tilde{\zeta}$. In this case, by (15) and Proposition 11,

$$\nabla^2\eta(v) = A\nabla^2\tilde{\zeta}(u)A^T = A\nabla x(u)A^T = A(A^T M^{-1}A + \nabla^2 f(x))^{-1}A^T,$$

which is positive definite due to $\text{rank}(A) = m$.

Case ii. $u \in \text{int}\widehat{U}$. For any facet Q_j and the corresponding function $\tilde{\zeta}_j$ as discussed in Lemma 2 and Proposition 4, it follows that

$$\nabla\tilde{\eta}_j(v) = A\nabla\tilde{\zeta}_j(A^T(v + M^{-1}a)).$$

Thereby,

$$\nabla^2\tilde{\eta}_j(v) = A\nabla^2\tilde{\zeta}_j(A^T(v + M^{-1}a))A^T = A\nabla\xi_{x\,j}(A^T(v + M^{-1}a))A^T.$$

Thus, according to Proposition 10 and Lemma 4, the desired result follows immediately.

Case iii. $u \in \text{bd}\widehat{U}$. By Propositions 7 and 8, it is easy to see that any piece $\tilde{\eta}_j$ should be a piece either in Case i or Case ii. Hence, the desired result is valid. This completes the proof. □

Remark 4 Note that, for sake of simplicity in analysis, for any facet Q of $\mathcal{F}$, we may redefine set $G_Q(x, \lambda)$ defined in Sect. 3 as follows: $G_Q(x, \lambda) := \{d : \nabla f_i(x)^T d = 0,\ \forall\, i \in I_Q\}$. In addition, the MFCQ condition can be strengthened by LICQ. In this case, the GAIPCQ condition becomes to be redundant because LICQ implies CRCQ and CRCQ implies GAIPCQ. The results obtained above are still valid after making some necessary changes.

By Proposition 12, we derive the following result which simplifies the conditions of Proposition 12.

Proposition 13 *For $u \in \mathbb{R}^n$, let x be the optimal solution of (P_u). Suppose that MFCQ and GAIPCQ hold at x, and that the strong second order sufficiency condition holds at every KKT point of (P_u). Assume that*

$$\mathcal{R}(A^T) \cap \text{span}\{\nabla f_i(x)\}_{i\in I_Q} = \{0\} \tag{35}$$

for any facet Q of $\mathcal{F}$. Then, g is BD-regular at any $v \in \mathbb{R}^m$. Thereby, all $V \in \partial g(v)$ are positive definite.

Proof According to the linear independence of $\{\nabla f_i(x)\}_{i\in I_Q}$ and the definition of S defined in Proposition 10 with respect to Q, it follows that

$$\text{span}\{\nabla f_i(x)\}_{i\in I_Q} \dotplus \text{span}(S) = \mathbb{R}^n,$$

where "$\dot{+}$" denotes the orthogonal decomposition operation in the vector space $\mathbb{R}^n$. Let $R = (\nabla f_i(x))_{i \in I_Q} \in \mathbb{R}^{n \times |I_Q|}$. Since, by assumption, $\text{rank}(A) = m$, we then have $A^T d \neq 0$ for any $d \neq 0 \in \mathbb{R}^m$. Hence, it follows from (35) that $A^T d \notin \text{span}\{\nabla f_i(x)\}_{i \in I_Q}$. Thereby, there exist $\alpha \in \mathbb{R}^{|I_Q|}$ and $\beta \neq 0 \in \mathbb{R}^{n-|I_Q|}$ such that

$$A^T d = R\alpha + S\beta.$$

Note that $R^T S = 0$, it yields that

$$\begin{aligned}
&(A^T d)^T S(S^T HS)^{-1} S^T (A^T d) \\
&\quad = (R\alpha + S\beta)^T S(S^T HS)^{-1} S^T (R\alpha + S\beta) \\
&\quad = \alpha^T R^T S(S^T HS)^{-1} S^T (R\alpha + S\beta) + \beta^T S^T S(S^T HS)^{-1} S^T (R\alpha + S\beta) \\
&\quad = \beta^T S^T S(S^T HS)^{-1} S^T S\beta > 0,
\end{aligned}$$

where H is defined as Proposition 12 with respect to facet Q. Hence, all the conditions in Proposition 12 hold, which implies that the desired results are valid. This completes the proof. □

5 Conclusion

The Lagrangian dual is widely used for large-scale problems. We investigate the semismoothness of the gradient g of the Moreau–Yosida regularization of the Lagrangian-dual function, which plays a key role in the superlinear or quadratic convergence analysis of generalized Newton methods for solving nonsmooth equations. Besides the well known smoothness property, we have showed that the regularized function η possesses a nice feature, i.e., η is piecewise C^2, which is a large class of locally Lipschitz continuous functions. We have obtained the piecewise smoothness and thereby semismoothness of the gradient g of the regularized function. We have also investigated the conditions, under which the Clarke generalized Jacobian of g is BD-regular and thereby is CD-regular. For future research, we will investigate the the relationship between the projection mapping over the epigraph of the regularized function η and that of the Lagrangian-dual function φ which was studied in [9,10].

Acknowledgements The author would like to thank the two anonymous referees for their detailed comments and suggestions which helped improve the presentation of this paper.

References

1. Chen, X., Nashed, Z., Qi, L.: Smoothing methods and semismooth methods for nondifferentiable operator equations. SIAM J. Numer. Anal. **38**, 1200–1216 (2000)
2. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
3. Dontchev, A.L., Qi, H.-D., Qi, L.: Convergence of Newton method for convex best interpolation. Numer. Math. **87**, 435–456 (2001)
4. Dontchev, A.L., Qi, H.-D., Qi, L.: Quadratic convergence of Newton method for convex interpolation and smoothing. Constr. Approx. **19**, 1230–143 (2003)
5. Fukushima, M., Qi, L.: A global and superlinear convergent algorithm for nonsmooth convex minimization. SIAM J. Optim. **6**, 1106–1120 (1996)
6. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms. Springer Verlag, Berlin (1993)
7. Janin, R.: Directional derivative of the marginal function in nonlinear programming. Math. Program. Stud. **21**, 110–126 (1984)

8. Meng, F., Hao, Y.: The property of piecewise smoothness of Moreau–Yosida approximation for a piecewise C^2 convex function. Adv. Math. (China) **30**, 354–358 (2001)
9. Meng, F., Sun, D., Zhao, G.: Semismoothness of solutions to generalized equations and the Moreau–Yosida regularization. Math. Program. **104**, 561–581 (2005)
10. Meng, F., Zhao, G., Goh, M., Souza, R.D.: Lagrangian-dual functions and Moreau–Yosida regularization. SIAM J. Optim. **19**, 39–61 (2008)
11. Mifflin, R.: Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Optim. **15**, 957–972 (1977)
12. Mifflin, R., Qi, L., Sun, D.: Properties of the Moreau–Yosida regularization of a piecewise C^2 convex function. Math. Program. **84**, 269–281 (1999)
13. Moreau, J.J.: Proximite et dualite dans un espace hilbertien. Bull. Soc. Math. France **93**, 273–299 (1965)
14. Pang, J.-S., Ralph, D.: Piecewise smoothness, local invertibility, and parametric analysis of normal maps. Math. Oper. Res. **21**, 401–426 (1996)
15. Pang, J.-S., Sun, D., Sun, J.: Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems. Math. Oper. Res. **28**, 39–63 (2003)
16. Qi, L., Sun, J.: A nonsmooth version of Newton's method. Math. Program. **58**, 353–367 (1993)
17. Qi, L., Sun, D., Zhou, G.: A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities. Math. Program. **87**, 1–35 (2000)
18. Robinson, S.M.: Generalized equations and their solutions, Part II: application to nonlinear programming. Math. Program. Stud. **19**, 200–211 (1982)
19. Rockafellar, R.T.: Convex Analysis. Princeton, New Jersey (1970)
20. Ruszczyński, A., Shapiro, A. (eds.): Stochastic Programming. Handbook in Operations Research and Management Science. Elsevier Science, Amsterdam (2003)
21. Scholtes, S.: Introduction to Piecewise Smooth Equations. Habilitation thesis, University of Karlsruhe, Karlsruhe, Germany (1994)
22. Sun, D., Han, J.: On a conjecture in Moreau–Yosida regularization of a nonsmooth convex function. Chin. Sci. Bull. **42**, 1140–1143 (1997)
23. Sun, D., Sun, J.: Semismooth matrix valued functions. Math. Oper. Res. **27**, 150–169 (2002)
24. Yosida, K.: Functional Analysis. Springer Verlag, Berlin (1964)
25. Zhao, G.: A Lagrangian dual method with self-concordant barrier for multi-stage stochastic convex nonlinear programming. Math. Program. **102**, 1–24 (2005)