SIAM J. OPTIM.
Vol. 19, No. 1, pp. 39–61

LAGRANGIAN-DUAL FUNCTIONS AND MOREAU–YOSIDA REGULARIZATION*

FANWEN MENG†, GONGYUN ZHAO‡, MARK GOH§, AND ROBERT DE SOUZA†

Abstract. In this paper, we consider the Lagrangian-dual problem of a class of convex optimization problems. We first discuss the semismoothness 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, e.g., the semismoothness of the gradient g of the regularized function η. We show that φ and g are piecewise C^2 and semismooth, respectively, for certain instances of the optimization problem. We establish a relationship between the original problem and the Fenchel conjugate of the regularization of the corresponding Lagrangian dual problem. We also find some instances of the optimization problem whose Lagrangian-dual function φ is not piecewise smooth. However, its regularized function still possesses nice second-order properties. Finally, we provide an alternative way to study the semismoothness of the gradient under the structure of the epigraph of the dual function.

Key words. Lagrangian dual, Moreau–Yosida regularization, piecewise C^k functions, semismoothness, Fenchel conjugate

AMS subject classifications. 90C25, 65K10, 52A41

DOI. 10.1137/060673746

1. Introduction.

Consider the following convex program:

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

where f, f_i, $i = 1, 2, \ldots, \theta$, are smooth and convex on $\mathbb{R}^n$, and where $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = m$ and $0 < m < n$.

It is known that many practical problems can be converted to problem (1) above. For instance, some recently studied multistage stochastic programming models can be formulated as (1). See [19, Chapter 1] for the detailed modeling in this regard.

Let $\mathcal{F} := \{x \in \mathbb{R}^n : f_i(x) \le 0, i \in \hat{I}\}$. In many circumstances, particularly in multistage stochastic programming, f and $\mathcal{F}$ are separable, while the constraint $Ax = a$ is nonseparable. Thus, we seek to relax the constraint $Ax = a$ using the Lagrangian dual of problem (1) as follows:

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

*Received by the editors October 30, 2006; accepted for publication (in revised form) August 7, 2007; published electronically February 6, 2008. The work of the first, third, and fourth authors was supported in part by the Economic Development Board of Singapore Research grant R-385-000-019-414.

http://www.siam.org/journals/siopt/19-1/67374.html

†The Logistics Institute—Asia Pacific, National University of Singapore, 7 Engineering Drive 1, Singapore 117574 (tlimf@nus.edu.sg, tlirbrtm@nus.edu.sg).

‡Department of Mathematics, National University of Singapore, 2 Science Drive 2, Singapore 117543 (matzgy@nus.edu.sg). The work of this author was supported in part by the National University of Singapore Academic Research grant R-146-000-057-112.

§Department of Decision Sciences, National University of Singapore, 1 Business Link, Singapore 117592, and The Logistics Institute—Asia Pacific, National University of Singapore, 7 Engineering Drive 1, Singapore 117574 (bizgohkh@nus.edu.sg, tligohkh@nus.edu.sg).

where

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

In these circumstances, the subproblem in (3) is separable, and then is solvable through the well-developed parallel algorithms. This makes the evaluation of φ much easier in general. However, an obstacle in solving problem (2) is that the function φ is nondifferentiable. To overcome this and noticing that the underlying function φ is convex on $\mathbb{R}^m$, we then use the well-known regularization of Moreau [14] and Yosida [23] 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 as defined 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 that the set of minimizers of problem (4) is exactly the set of minimizers of (2). It can be shown that η is continuously differentiable and that its gradient $g = \nabla\eta$ is globally Lipschitz continuous with modulus $||M||$. For the properties of the Moreau–Yosida regularization, the reader is referred to [8, 7]. For the problem discussed in the present paper, there are some advantages of using the Moreau–Yosida regularization, as given next.

Fukushima and Qi [7] have shown that 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 relatively easier. 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. This method was utilized for solving multistage stochastic nonlinear problems recently in [24], in which the underlying stochastic problem was formulated as problem (1). For any prescribed accuracy, we can now choose an appropriate $\mu > 0$ such that the minimizer of $\varphi(w, \mu) + (1/2)\|w - v\|_M^2$ is a desirable approximate solution to (5).

It is interesting that both the 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, while 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, giving rise to an enhanced method for minimizing the nonsmooth function φ. Hence establishing the theoretical framework of the Moreau–Yosida regularization can benefit a variety of algorithms.

For the problem under consideration, 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 method for nonsmooth convex problems by combining the Moreau–Yosida regularization scheme in (5) [7].

The concept of semismooth functions, an important subclass of Lipschitz functions, was first introduced by Mifflin [12]. In order to study the superlinear convergence of Newton method for solving nondifferentiable 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 [4, 5], nondifferentiable equations in which the underlying functions are slant differentiable functions [1], and complementarity problems and variational inequalities [6], for instance.

In this paper, we will focus 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 [20, 17]. In the past few years, many people have studied the piecewise smoothness of nonsmooth functions and designed algorithms based on Newton's method for solving the associated nonsmooth equations or nonlinear optimization problems. For example, the analysis was mainly focused on the concept of piecewise C^k functions in [10, 13, 21], where the authors have considered properties of g for some specific classes of φ. Specifically, Sun and Han [21] showed the semismoothness of g if φ is the maximum of several twice continuously differentiable convex functions under a constant rank constraint qualification (CRCQ). Later, Meng and Hao [10] derived the same result for the case of unconstrained problem (1) with the objective function f being a piecewise C^2 function under a weaker sequential constant rank constraint qualification. In [13], 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 under a so-called affine independence preserving constraint qualification (AIPCQ).

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 Lagrangian-dual function φ and the gradient of its Moreau–Yosida regularization η. Further, studying the properties of Lagrangian-dual function φ has its own interest as well; see [22] and the references therein, for instance. Since piecewise smooth functions as a special class of semismooth functions possess more enjoyable properties than semismooth functions [12, 16, 17, 20], we will concentrate on the study of piecewise smoothness of φ and the gradient g of the regularized function η in the context. We have adopted two different methods in analyzing properties of g. In terms of the first method, the main tool used in this study is based on Proposition 1 (see section 2), which was established by Mifflin, Qi, and Sun [13] using the notion of piecewise smoothness. We will first study the piecewise smoothness of φ. This property will then be used to show the semismoothness of g. For the problem with the linear objective function $f(x) = c^T x$, we can show that the function φ is piecewise C^2 and satisfies AIPCQ, and thus g is semismooth by Proposition 1 if all f_i's are affine functions or all $\nabla^2 f_i$'s are positive definite. We also present an example whose region $\mathcal{F}$ is defined by a linear constraint and a strictly convex constraint. In this example, the function φ is, surprisingly, *not* piecewise C^2, and, equally surprisingly, the gradient g of the regularization of φ *is* still semismooth. For general convex objective functions f and constraint functions f_j, it is completely unknown how smooth φ and g should be. This issue is considered by analyzing some special cases where the objective function possesses a positive definite Hessian. The second method is mainly based on the metric projection operator under the structure of the epigraph of the Lagrangian-dual function. Using the projection mapping, the study of the properties of g is equivalently converted to the study of the properties of solutions to a system of nonsmooth equations. The analysis is basically based on the framework established by Meng, Sun, and Zhao [11] recently. The results obtained complement and enrich the framework of piecewise smooth functions [20, 17] and also enhance the recent results on the Moreau–Yosida regularization [11].

Another topic of interest is the study of the duality of the original problem (1). It is well known that the duality theory is a fundamental issue in optimization both theoretically and numerically. For problem (1) with a linear objective, we derive an interesting result regarding the original problem and the Fenchel conjugate of Moreau–Yosida regularization of its Lagrangian-dual function, characterizing a relationship between the conjugate and the Lagrangian-dual. This provides a new way to look at the Lagrangian-dual and the Moreau–Yosida regularization. We believe that the established results complement the dual theory in optimization, particularly the theory of Magnanti [9] to some extent.

The rest of the paper is organized as follows. In section 2, basic definitions and properties are collected. The analysis of problems with the linear objective functions covers the next two sections. Section 3 investigates the piecewise smoothness of the function φ. Section 4 studies the semismoothness of the gradient g and the conjugate of the Moreau–Yosida regularization. Illustrative examples are presented in sections 3 and 4. Section 5 discusses the case of general convex objective functions. Section 6 concludes.

2. Preliminaries.

In this section, we briefly recall some concepts, such as semismoothness, piecewise smoothness, and AIPCQ, which will be used in the rest of this paper.

It is known that the regularized function η is a continuously differentiable convex function defined on $\mathbb{R}^m$, even though φ may be nondifferentiable. The gradient of η at v (see [8]) is

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

where $p(v)$ represents the unique solution of the minimization problem in (5). In order to use Newton method or modified Newton's methods for solving (4), it is important to study the Hessian of η, i.e., the Jacobian of g. Note that, in general, g may not be differentiable. To extend the definition of Jacobian to certain classes of nonsmooth functions, Qi and Sun [16] introduced the definition of semismoothness [12] for vector-valued functions. See [16] for details.

A remarkable feature of semismoothness is that superlinear or quadratic convergence of a generalized Newton method for solving nonsmooth equations can be obtained under the assumption of semismoothness. See [7, 15, 16] for the relevant discussions. 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 [11, 15] and the references therein. As for the underlying Lagrangian dual function φ, it has a special feature; i.e., φ is piecewise smooth. We shall make use of this special feature to investigate the semismoothness of g in the subsequent analysis. We now give a definition of piecewise smooth functions below, which is slightly different from the one given in [20].

DEFINITION 1. *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 index set $I = \{1, \dots, q\}$, closed sets $D_1, \dots, D_q$, open sets $U_1, \dots, U_q$ (or relatively open with respect to the affine hull of D), and functions $\psi_1, \dots, \psi_q$ such that*

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

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

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

Note that for the Moreau–Yosida regularization of a piecewise smooth function to be smooth, the pieces ψ_j must be joined together properly. Mifflin, Qi, and Sun [13] introduced the following constraint qualification—AIPCQ. For any $u \in D$, we write

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

Definition 2. *The* 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*

$$\left\{ \begin{pmatrix} \nabla\psi_i(u^k) \\ 1 \end{pmatrix} : i \in K \right\} \tag{7}$$

being linearly independent, it follows that the vectors

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

are linearly independent.

Remark 2. The set $I(u)$ defined in this paper and the corresponding set in [13], denoted by $I'(u)$, are slightly different. In [13], they define

$$I'(u) = \{j \in I : \psi_j(u) = \psi(u)\}.$$

Since $u \in D_j$ implies $\psi_j(u) = \psi(u)$, we have $I(u) \subseteq I'(u)$. 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,

$$I(u) = I'(u) \qquad \forall\, u \in D.$$

By virtue of the AIPCQ, Mifflin, Qi, and Sun [13] derived the following result, which will be used in the analysis of this paper.

Proposition 1. *Suppose that the convex function* φ *is piecewise* C^2 *on* $\mathbb{R}^m$ *and that the AIPCQ holds at the proximal point* $p(v)$ *for a given* $v \in \mathbb{R}^m$. *Then there exists an open neighborhood* $\mathcal{N}(v)$ *about* v *such that the gradient* g *of the function* η, *the Moreau–Yosida regularization of* φ, *is piecewise* C^1 *(smooth) on* $\mathcal{N}(v)$. *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 for the case $f(x) = c^T x$ in (3), which is defined by

$$\varphi(v) = \sup\{-c^T x + v^T(Ax - a) \mid x \in \mathcal{F}\}. \tag{9}$$

The piecewise smoothness is an important characteristic of the Lagrangian-dual function φ. The investigation of this characteristic is helpful to optimization methods which use the Lagrangian dual. Hence the results in this section are significant in their own right. In the next section, the piecewise smoothness of φ will then be used to prove the semismoothness of the gradient of the Moreau–Yosida regularization. Denote

$$\Omega := \{u = A^T v - c : v \in \mathbb{R}^m\}.$$

Clearly, Ω is an m-dimensional affine set in $\mathbb{R}^n$ since $\text{rank}(A) = m$. We make the following assumptions throughout the paper.

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

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

Assumption 3. $\mathcal{F} \neq \emptyset$ and $\Omega \cap \mathcal{F}^b \neq \emptyset$.

Here, $\mathcal{F}^b$ denotes the *barrier cone* of the convex set $\mathcal{F}$ defined by

$$\mathcal{F}^b = \{y \in \mathbb{R}^n \mid \exists\, \beta \in \mathbb{R} \text{ such that } y^T x \leq \beta \ \ \forall\, x \in \mathcal{F}\}.$$

Remark 3. If $c = A^T s$ for some $s \in \mathbb{R}^m$, then $Ax = a$ implies $c^T x = s^T Ax = s^T a$. This means that any feasible solution of (1) is an optimal solution. Assumption 1 should rule out this degenerate case. Assumption 1 can also be written as $0 \notin \Omega$. Assumption 2 is a natural assumption of smoothness. The motivation of Assumption 3 is to guarantee the properness of the function φ, as shown by Lemma 1 below.

Define ζ, the *support function* of $\mathcal{F}$ in $\mathbb{R}^n$, as follows

$$\zeta(u) = \delta^*(u \mid \mathcal{F}) := \sup\{\langle u, x\rangle \mid x \in \mathcal{F}\}, \quad u \in \mathbb{R}^n. \tag{10}$$

Then the Lagrangian-dual function φ defined in (9) can be rewritten as

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

We now define some notation which will be used in the 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) For a convex function $h : \mathbb{R}^s \to \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$, the *domain* of h, denoted by $\text{dom}h$, is defined by $\text{dom}h := \{z \in \mathbb{R}^s : h(z) < +\infty\}$.

LEMMA 1. *The Lagrangian-dual function φ is a proper convex function on $\mathbb{R}^m$ if and only if Assumption* 3 *holds. One also has*

$$\text{dom}\varphi = \{v \in \mathbb{R}^m \mid A^T v - c \in \mathcal{F}^b\}.$$

Proof. It is evident that $\varphi(v)$ can never be $-\infty$ if $\mathcal{F} \neq \emptyset$, and if $\mathcal{F} = \emptyset$, then $\varphi \equiv -\infty$.

By (11), we have

$$\text{dom}\varphi = \{v \in \mathbb{R}^m \mid A^T v - c \in \text{dom}\zeta\}.$$

Hence, $\text{dom}\varphi \neq \emptyset$ if and only if $\Omega \cap \text{dom}\zeta \neq \emptyset$. Since ζ is the support function of $\mathcal{F}$, it is easy to see that

$$\text{dom}\zeta = \mathcal{F}^b.$$

Therefore, the second condition in Assumption 3 is a necessary and sufficient condition for $\text{dom}\varphi \neq \emptyset$. ☐

PROPOSITION 2. *If ζ is piecewise C^2 on a set $D \subseteq \mathbb{R}^n \setminus \{0\}$, then φ is piecewise C^2 on the set $E := \{v : A^T v - c \in D\} \subseteq \mathbb{R}^m$ under Assumption* 1.

Proof. Since ζ is piecewise C^2, there exist closed sets D_i, open sets U_i, and functions $\zeta_i \in C^2(U_i)$, $i \in l$, where l is a finite index set, which satisfy Definition 1. Let

$$\varphi_i(v) := \zeta_i(A^T v - c) - a^T v, \quad E_i := \{v : A^T v - c \in D_i\}, \quad V_i := \{v : A^T v - c \in U_i\}.$$

Then it is evident that $\varphi_i \in C^2(V_i)$, E_i is closed, and V_i is open. Furthermore, $\varphi_i (i \in l)$ satisfy (i)–(iii) in Definition 1. Hence φ is a piecewise C^2 function. ☐

PROPOSITION 3. *Suppose that f_i is an affine function on $\mathbb{R}^n$ for every $i \in \hat{I}$. Then the function φ defined in* (9) *is a piecewise C^2 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 $\Omega \cap \mathrm{dom}\zeta$. According to the remark after Definition 1, it suffices to show that there exist twice continuously differentiable functions ζ_j on $\mathbb{R}^n (= U_j)$, $j \in \hat{J}$ a finite index set, such that for any $u \in \Omega \cap \mathrm{dom}\zeta$

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

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

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

Define

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

We claim that, for any $u \in \mathrm{dom}\zeta$, $\sup\{u^T x : x \in \mathcal{F}\} = \sup\{u^T x : x \in \bar{\mathcal{F}}\}$. Assume by contradiction that there exist a $u \in \mathrm{dom}\zeta$ and a $\bar{x} \in \mathcal{F} \setminus \bar{\mathcal{F}}$ such that $u^T \bar{x} > \sup\{u^T x : x \in \bar{\mathcal{F}}\}$. Denote $J := \{i : \lambda_i > 1\}$, where the 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 for 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.,

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

Thus 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 the fact $u \in \mathrm{dom}\zeta$. This shows that for any $u \in \mathrm{dom}\zeta$

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

Note that $\bar{\mathcal{F}}$ is a bounded polytope. Without loss of generality, let $\{\bar{x}_1, \dots, \bar{x}_k\}$ be all vertices of $\bar{\mathcal{F}}$, and define $\zeta_j(u) = \bar{x}_j^T u$. Then $\zeta_j \in C^2(\mathbb{R}^n)$ (here $U_j = \mathbb{R}^n$). For any $u \in \Omega \cap \mathrm{dom}\zeta$, because $u \neq 0$ by Assumption 1, 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 on its domain. Evidently, $\zeta(u)$ is also piecewise affine on its domain, and so is φ. ∎

Next, we consider the case where all $\nabla^2 f_i$ ($i \in \hat{I}$) are positive definite. Our analysis will proceed as follows. For each facet Q (of any dimension) of $\mathcal{F}$, we will define an open set U and a C^2 function on U. Roughly speaking, we first define a mapping from x-space to an open set in u-space (actually the mapping is defined on enlarged spaces), prove that this mapping is bijective, and then use the inverse of this mapping to define a function on the open set in u-space. For any facet Q of $\mathcal{F}$ with the index set I_Q, we define

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

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

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

Note that for $(x,\lambda) \in W$, x is not required to be in Q. Actually, x need not be in $\mathcal{F}$. Without loss of generality, let $I_Q = \{1,\ldots,k\}$. Denote $\tilde{f} = (f_1,\ldots,f_k)^T$, and define a mapping $\Gamma : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^{n+k}$ by

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

Note that the Karush–Kuhn–Tucker (KKT) conditions for problem (10) can be written as

$$\Gamma(x,\lambda) = (u;0).$$

The following lemma plays a fundamental role in our analysis.

LEMMA 2. *Let W, U be defined by* (13), (14), *respectively. Suppose that for any $x \in \mathbb{R}^n$ all $\nabla^2 f_i(x)$ ($i \in I_Q$) are positive definite and $\{\nabla f_i(x)\}_{i\in I_Q}$ are linearly independent. Then* (i) *U is an open set in $\mathbb{R}^n$; and* (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 seek to show that $\bar{u}$ is an interior point of U. Let us denote $\nabla\tilde{f} := (\nabla f_1,\ldots,\nabla f_k) \in \mathbb{R}^{n\times k}$. Then, $\nabla\tilde{f}(x)$ has full column rank, i.e., $\mathrm{rank}(\nabla\tilde{f}(x)) = k$, by assumption. By the continuity of $\nabla^2 f_i$, $i \in \hat{I}$, 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 $(x,\lambda) \in \mathcal{N}_x$. Thus, the Jacobian of Γ,

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

is nonsingular on $\mathcal{N}_x$. 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

$\mathcal{N}_u$, and for any $(u,v) \in \mathcal{N}_u$, $\Psi(u,v) \in \mathcal{N}_x$ and $\Gamma(\Psi(u,v)) = (u;v)$. 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 $\mathbb{R}^n$, 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 since 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 $\Gamma(\xi(u)) = (u;0)$. ∎

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

LEMMA 3. *Let $\zeta_Q(u) = u^T \xi_x(u)$, where ξ_x is defined in Lemma* 2. *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 2 and the first equation of $\Gamma(\xi(u)) = (u;0)$, it follows that

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

Thus,

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

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 $\nabla \xi_x(u) \nabla f_i(\xi_x(u)) = 0$ for all $i \in I_Q$. 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 2, $\xi_x(u)$ is continuously differentiable on U. Therefore, ζ_Q is twice continuously differentiable on U. ∎

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

PROPOSITION 4. *For φ defined by* (9), *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 on its domain.*

Proof. Let us first consider the function ζ defined by (10) on the set D, where $D = \Omega \cap \operatorname{dom}\zeta$. Let $\{Q_1, \dots, Q_q\}$ be the set of all facets of $\mathcal{F}$. Let W_i, U_i, and ξ_i be defined in (13), (14), and Lemma 2 for the facet $Q_i = \{x \in \mathcal{F} : f_l(x) = 0,\ l \in I_i\}$. Define

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

which is evidently a closed set. By Lemma 2, 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 1 hold.

(i) For any $u \in \mathrm{ri}D$, by [18, Theorem 23.4 and Corollary 23.5.3], there exists an optimal solution x^* to problem (10), which together with a Lagrangian multiplier $\bar{\lambda}^*$ satisfies the KKT conditions:

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

Because $u \neq 0$ by Assumption 1, x^* must lie on some facets of $\mathcal{F}$. 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 $u = \sum_{i \in I_j} \lambda_i^* \nabla f_i(x^*)$, which together with $\lambda^* \geq 0$ and $x^* \in Q_j$ implies $u \in D_j$. This shows that $\mathrm{ri}D \subseteq \cup_{i=1}^q D_i$. Thereby, $D \subseteq \cup_{i=1}^q D_i$ since each D_i is closed.

For any $u \in D_j$, let $\bar{x} \in Q_j$ and $\bar{\lambda} \geq 0$ represent u as in (18). $\bar{\lambda} \neq 0$, since $u \neq 0$ by Assumption 1. This implies that $\sum_{i \in I_j} \bar{\lambda}_i \nabla^2 f_i(\bar{x}) \succ 0$, since all $\nabla^2 f_i$ are positive definite. Thus $(\bar{x}, \bar{\lambda}) \in W_j$ and $u \in U_j$. This shows $D_j \subseteq U_j$.

(ii) By Lemma 3, $\zeta_i \in C^2(U_i)$ for $i = 1, \ldots, q$.

(iii) For any $u \in D \cap D_j$, 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, we have

$$\zeta_j(u) = u^T \xi_{jx}(u) = u^T \bar{x}.$$

Thus $\zeta(u) = \zeta_j(u)$ for any $u \in D \cap D_j$.

The above shows that ζ is piecewise C^2 on $D(= \Omega \cap \mathrm{dom}\zeta)$. By virtue of Proposition 2, φ is piecewise C^2 on its domain. □

Remark 4. 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 an $\mathcal{F}$ which is defined by a linear constraint and a strictly convex constraint with a positive definite Hessian, and gives a negative answer to the above question.

Example 1. Let

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

where

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

and

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

Since $\mathcal{F}$ is bounded, $\mathcal{F}^b = \mathbb{R}^3$ and $\mathrm{dom}\varphi = \mathbb{R}^2$. One can verify that Assumptions 1, 2, and 3 are satisfied. It is easy to see that

$$\varphi(v) = \sup\{(v_1, v_2, 1)^T x \mid x \in \mathcal{F}\},$$

and that the maximizer is $x = (v_1/\|v\|, v_2/\|v\|, 0)^T$ if $v \neq 0$ and is any point on $\mathcal{F} \cap \{(x_1, x_2, x_3) \mid x_3 = 0\}$ if $v = 0$. It follows that

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

Obviously, φ is smooth at any point $v \neq 0$. So for any nonzero $v \in \mathbb{R}^2$, the gradient and the Hessian of φ can be written as

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

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

It is evident that $\nabla^2\varphi(v)$ is unbounded as $v \to 0$ ($v \neq 0$), (either $\frac{\partial^2\varphi(v)}{\partial v_1^2} \to \infty$ if $|v_1| \le |v_2|$, or $\frac{\partial^2\varphi(v)}{\partial v_2^2} \to \infty$ if $|v_1| \ge |v_2|$).

To show φ is not piecewise C^2 on its domain $\mathbb{R}^2$, let (E_j, V_j, φ_j) be any piece representing φ in a neighborhood of $v = 0$, namely, $0 \in E_j \subset V_j$ and φ_j is a function on V_j satisfying $\varphi_j(v) = \varphi(v)$ for all $v \in E_j$. Since $\nabla^2\varphi_j(v) = \nabla^2\varphi(v)$ wherever φ is twice differentiable, we have $\nabla^2\varphi_j(v) \to \infty$ as $v \to 0$ ($E_j \ni v \neq 0$). Since the origin is an interior point of V_j, $\varphi_j \notin C^2(V_j)$. Therefore, φ is not a piecewise C^2 function on its domain.

4. Semismoothness of the gradient of η and its conjugate. In this section, we will study the semismoothness of the gradient of the Moreau–Yosida regularization of φ as discussed in section 3, where $f(x) = c^T x$. We will also investigate the properties of the conjugate of η and explore its relations with the original problem.

4.1. Semismoothness of the gradient of η. Our study on the semismoothness of g is based on the theory established by Mifflin, Qi, and Sun [13]. In their paper, they assume that φ is piecewise C^2 convex on the whole space $\mathbb{R}^m$, i.e., $\mathrm{dom}\varphi = \mathbb{R}^m$. We follow this assumption in this section. By Lemma 1, $\mathrm{dom}\varphi = \mathbb{R}^m$ if and only if $\Omega \subset \mathcal{F}^b$. Therefore, we make the following assumption in this section to replace Assumption 3.

Assumption 4. $\mathcal{F} \neq \emptyset$ and $\Omega \subset \mathcal{F}^b$.

To show the semismoothness of g, we shall first show that φ defined in (3) satisfies AIPCQ in the following two cases: (i) all f_i are affine, and (ii) all f_i possess positive definite Hessian matrices. Suppose that $\varphi(v)$ is a piecewise smooth function with the representation $\{(E_i, V_i, \varphi_i)\}_{i\in I}$. For any $v \in \mathbb{R}^m$, define

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

LEMMA 4. *Suppose for every $i \in \hat{I}$ that f_i is an affine function on $\mathbb{R}^n$. Then for the piecewise affine function $\varphi(v)$ defined by* (9), *the AIPCQ holds at every v on $\mathbb{R}^m$.*

Proof. Suppose that φ is represented by $\{\varphi_i\}_{i\in I}$, where $\varphi_i(v) = \beta_i^T v - \alpha_i$. For any $w \in \mathbb{R}^m$ and any index set $K \subseteq I(w)$,

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

is a set of constant vectors. Therefore the AIPCQ holds at any $v \in \mathbb{R}^m$. □

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

Because the value of $\varphi_j(v)$, $v \in V_j \setminus E_j$, does not affect the representation of φ, it therefore can be set to any value. For simplicity, in what follows, we assume that

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

(see also Remark 2).

LEMMA 5. *Suppose that the conditions of Proposition* 4 *are satisfied. Let $\{(E_j, V_j, \varphi_j)\}_{j\in I}$ be a representation of φ. Then, for any $v \in \mathrm{dom}\varphi$ and any i, $j \in I$, $\nabla\varphi_i(v) = \nabla\varphi_j(v)$ if $\varphi_i(v) = \varphi_j(v) = \varphi(v)$.*

Proof. It suffices to show that for any $u \in \Omega \cap \mathrm{dom}\zeta$ and any i, $j \in I$, $\nabla\zeta_i(u) = \nabla\zeta_j(u)$ if $\zeta_i(u) = \zeta_j(u) = \zeta(u)$, where $\{(D_j, U_j, \zeta_j)\}_{j\in I}$ is the corresponding representation of ζ. Let $u \in U_i$ and $\xi_i : U_i \to W_i$ be defined as in Lemma 2. If we can show that $\zeta_i(u) = \zeta(u)$ implies that $\xi_{ix}(u)$ is indeed the unique maximizer x^* of problem (9) for the given u, then the fact $\nabla\zeta_i(u) = \xi_{ix}(u)$ (Lemma 3) leads readily to $\nabla\zeta_i(u) = x^* = \nabla\zeta_j(u)$, provided that $\zeta_i(u) = \zeta_j(u) = \zeta(u)$.

Now, if $\zeta_i(u) = \zeta(u)$, then (20) (applying to ζ) and $u \in U_i$ imply $u \in D_i$. By definition of (18), $x = \xi_{ix}(u) \in Q_i \subset \mathcal{F}$; i.e., $\xi_{ix}(u)$ is a feasible solution of problem (9). So, $u^T\xi_{ix}(u) = \zeta_i(u) = \zeta(u)$ implies that $\xi_{ix}(u)$ is a unique optimal solution x^* of problem (9) for a given u. □

The above lemma actually holds true for φ with $\mathrm{dom}\varphi \neq \mathbb{R}^m$. This lemma will be used to prove Lemma 6. In addition to it, we obtain a property of the function φ as a by-product, namely, φ is indeed differentiable on the relative interior of $\mathrm{dom}\varphi$, because the subdifferential $\partial\varphi(v)$ at any point $v \in \mathrm{ri}(\mathrm{dom}\varphi)$ is a singleton.

LEMMA 6. *Suppose that the conditions of Proposition* 4 *are satisfied. Then, for the piecewise C^2 function φ, the AIPCQ holds at each $v \in \mathbb{R}^m$.*

Proof. Let $v \in \mathbb{R}^m$ and $K \subseteq I(v)$. If $|K| = 1$, the vectors in the set in (8) 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 5. Thus the set of vectors in (7) can never be linearly independent. This means that the conditions for AIPCQ hold automatically. □

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 stated below.

PROPOSITION 5. *Let φ be defined by* (9). *Suppose that Assumptions* 1, 2, *and* 4 *are satisfied. Suppose that f_i, $i \in \hat{I}$, are either all affine or all possess positive definite Hessian matrices. In 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 $\mathbb{R}^m$.*

Proof. The proof follows directly from Propositions 1, 3, and 4, and Lemmas 4 and 6. ∎

Remark 5. The above proposition shows that g is semismooth if constraints defining $\mathcal{F}$ either are all linear or all possess positive definite Hessian matrices. In Example 1 of section 3, we found that, for some simple 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, the gradient g of the Moreau–Yosida regularization of this function φ is semismooth, as shown below.

Example 2 (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

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

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 $\|v\| \ge 1$,

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

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

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

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

$$J(p(v)) = \begin{pmatrix} 1 - v_2^2/\|v\|^3 & v_1v_2/\|v\|^3 \\ v_1v_2/\|v\|^3 & 1 - v_1^2/\|v\|^3 \end{pmatrix}. \tag{22}$$

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 need only to investigate the semismoothness of p on the region where the two sets meet, namely, $\{v \in \mathbb{R}^2 : \|v\| = 1\}$. 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 the definition of semismoothness [16], it suffices to show that

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

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

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, $v' \in \text{int}\hat{V}_1$ or $v' \in \text{int}\hat{V}_2$. 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

$$\begin{aligned} &\lim_{h'\to h,t\to 0+}[(1 - {v_2'}^2/\|v'\|^3)h_1' + (v_1'v_2'/\|v'\|^3)h_2'] \\ &= (1 - \bar{v}_2^2/\|\bar{v}\|^3)h_1 + (\bar{v}_1\bar{v}_2/\|\bar{v}\|^3)h_2 = (1 - \bar{v}_2^2)h_1 + \bar{v}_1\bar{v}_2h_2 \\ &= \bar{v}_1^2h_1 + \bar{v}_1\bar{v}_2h_2 = \bar{v}_1(h_1\bar{v}_1 + h_2\bar{v}_2) = 0, \end{aligned}$$

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 (24), 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 the generalized Jacobian, V is a convex combination of the Jacobians in (21) and (22) (with v replaced by v'). Thus, Vh' tends to the same limit, namely 0, 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 $\mathbb{R}^2$. Therefore, g is semismooth on $\mathbb{R}^2$ as well.

4.2. Conjugate of the Moreau–Yosida regularization. In this subsection, we investigate the relationship between the original problem with the linear objective and the Fenchel conjugate of Moreau–Yosida regularization of its Lagrangian-dual function.

First, recall the notion of Fenchel conjugate. Let ϕ be a real-valued convex function on $\mathbb{R}^l$. The *Fenchel conjugate*, denoted by ϕ^*, of ϕ is defined by (see [18])

$$\phi^*(x) := \sup\{\langle x^*, x\rangle - \phi(x) \mid x^* \in \mathbb{R}^l\} \quad \forall\, x \in \mathbb{R}^l.$$

Note that η, the Moreau–Yosida regularization of φ defined in (9), can be rewritten as

$$\eta(v) = (\pi_1\Box\pi_2)(v) := \inf\{\pi_1(v - w) + \pi_2(w) : w \in \mathbb{R}^m\}, \quad v \in \mathbb{R}^m, \tag{27}$$

where "$\Box$" denotes the *infimal convolution* operation [18], $\pi_1(v) := \frac{1}{2}\|v\|_M^2, \pi_2(v) := \varphi(v)$, as defined in (9). Evidently, both π_1 and π_2 are proper convex functions; then by [18, Theorem 5.4], η is a convex function.

Using the conjugate operator, it is not hard to derive that

$$\pi_1^*(v) = \tfrac{1}{2}\|v\|_{M^{-1}}^2 \quad \forall\, v \in \mathbb{R}^m.$$

Hence, we have $\mathrm{dom}\pi_1^* = \mathbb{R}^m$. Thereby, it follows from [8, Corollary 2.1.3] that

$$\eta^*(v) = \pi_1^*(v) + \pi_2^*(v) \quad \forall\, v \in \mathbb{R}^m.$$

Next, we study the conjugate of π_2. To ease notation, we define a mapping $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^n$ by

$$\mathcal{A}(v) = A^T v - c.$$

Then we have

$$\delta^*(A^T v - c \mid \mathcal{F}) = \zeta \circ \mathcal{A}(v), \quad v \in \mathbb{R}^m,$$

where ζ is defined in (10). Since $\mathrm{dom}\zeta = \mathcal{F}^b$, so $\zeta \circ \mathcal{A}$ is a closed convex function on $\mathbb{R}^m$ under Assumption 3. Thus, by [18, Theorem 16.3], it follows that

$$(\zeta \circ \mathcal{A})^*(v) = \mathrm{cl} \inf_{x \in \mathbb{R}^n} \{\zeta^*(x) - \langle -c, x\rangle \mid Ax = v\}.$$

Since $\mathcal{F}$ is closed, we then have

$$\begin{aligned}
(\zeta \circ \mathcal{A})^*(v) &= \mathrm{cl} \inf_{x \in \mathbb{R}^n} \{(\delta^*(x \mid \mathcal{F}))^* + \langle c, x\rangle \mid Ax = v\} \\
&= \mathrm{cl} \inf_{x \in \mathbb{R}^n} \{\delta(x \mid \mathcal{F}) + \langle c, x\rangle \mid Ax = v\} \\
&= \mathrm{cl} \inf\{\langle c, x\rangle \mid Ax = v, x \in \mathcal{F}\}.
\end{aligned}$$

On the other hand, by definition of conjugate, we have

$$\begin{aligned}
\pi_2^*(v) &= \sup\{\langle v + a, v'\rangle - \sup\{\langle A^T v' - c, x\rangle \mid x \in \mathcal{F}\} \mid v' \in \mathrm{dom}\pi_2\} \\
&= \sup\{\langle v + a, v' - \zeta \circ \mathcal{A}(v')\rangle \mid v' \in \mathrm{dom}\pi_2\} \\
&= \sup\{\langle v + a, v' - \zeta \circ \mathcal{A}(v')\rangle \mid v' \in \mathrm{dom}(\zeta \circ \mathcal{A})\} \\
&= (\zeta \circ \mathcal{A})^*(v + a).
\end{aligned}$$

Thus, we obtain the conjugate of π_2 as follows:

$$\pi_2^*(v) = \mathrm{cl} \inf\{\langle c, x\rangle \mid Ax = v + a, x \in \mathcal{F}\}.$$

We now derive an interesting result on the conjugate of Moreau–Yosida regularization of the Lagrangian-dual function as follows.

PROPOSITION 6. *Assume $\Omega \cap \mathcal{F}^b \neq \emptyset$. Then, for any $v \in \mathbb{R}^m$,*

$$\begin{aligned}
&\mathrm{cl} \inf\{\langle c, x\rangle \mid Ax - a = v,\ x \in \mathcal{F}\} \\
&= \eta^*(v) - \tfrac{1}{2}\|v\|_{M^{-1}}^2.
\end{aligned}$$

From Proposition 6, we can see that the optimal value function of the underlying parametric optimization problem can be represented by the conjugate function of the regularized dual function of the (unperturbed) original problem, together with a quadratic function in terms of the perturbation parameter v. Note that the expression is taken under the *closure* and *infimal* operations on the set of objective values due to the fact that the minimum of the set of objective values of the corresponding feasible points might not exist in general.

Next we investigate under which situations these two operations can be replaced by the usual minimization operator so as to simplify the analysis on conventional minimization problems. We need the following assumption in the rest of this subsection.

Assumption 5. $\Omega \cap \mathrm{ri}\mathcal{F}^b \neq \emptyset$.

Note that under Assumption 5 and by virtue of [8, Theorem 2.2.3], we have

$$(\zeta \circ \mathcal{A})^*(v) = \min\{\langle c, x\rangle \mid Ax = v,\, x \in \mathcal{F}\} \quad \forall\, v \in \mathrm{dom}(\zeta \circ \mathcal{A})^*.$$

Let (P_v) denote the perturbed problem

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \langle c, x\rangle \\ \text{s.t.} \quad & Ax - a = v, \\ & x \in \mathcal{F}, \end{aligned}$$

where v serves as the perturbation parameter. We refer to the original problem (1) where the objective function is taken as an affine function, denoted by (P_0), to the unperturbed problem. We denote the optimal value function of (P_v) by $f_{\mathrm{val}}(v)$. Accordingly, $f_{\mathrm{val}}(0)$ denotes the optimal value of the original problem (1) or (P_0).

Then, we derive the following result immediately by virtue of Proposition 6.

PROPOSITION 7. *Suppose that Assumption* 5 *holds. Then*

$$f_{\mathrm{val}}(v) = \eta^*(v) - \tfrac{1}{2}\|v\|^2_{M^{-1}}$$

for any $v \in dom(\zeta \circ \mathcal{A})^* - a$.

Note that the above result enhances Proposition 6. It provides a new and interesting characteristic of convex conjugates in perturbation analysis. Note that the result is valid only if the parameter v belongs to the set $\mathrm{dom}(\zeta \circ \mathcal{A})^* - a$. Also, this result has a potential role in studying sensitivity analysis and some stochastic programs, both theoretically and numerically.

The next immediate question is about the nonemptiness of the domain of $(\zeta \circ \mathcal{A})^*$. Consider the case when the original problem (P_0) is bounded below; by definition, it follows that

$$\begin{aligned} &(\zeta \circ \mathcal{A})^*(a) \\ &= \min\{\langle c, x\rangle \mid Ax = a,\ x \in \mathcal{F}\} < \infty. \end{aligned}$$

Thus, $a \in \mathrm{dom}(\zeta \circ \mathcal{A})^*$. This implies that $\mathrm{dom}(\zeta \circ \mathcal{A})^*$ is nonempty, and so is $\mathrm{dom}(\zeta \circ \mathcal{A})^* - a$.

Before ending this section, we derive the following result based on the above arguments.

PROPOSITION 8. *Suppose that the original problem, namely,*

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \langle c, x\rangle \\ \text{s.t.} \quad & Ax = a, \\ & f_i(x) \le 0, \quad i \in \hat{I} = \{1, \ldots, \theta\}, \end{aligned}$$

is bounded below. Then, dom$(h \circ \mathcal{A})^* \neq \emptyset$ *and* $a \in$ *dom*$(h \circ \mathcal{A})^*$.

Furthermore, let $\{v^k\}$ *be a sequence in dom*$(h \circ \mathcal{A})^* - a$ *satisfying* $v^k \to 0$ *as* $k \to \infty$*; then*

$$\lim_{v^k \to 0} \eta^*(v^k) = \lim_{v^k \to 0} \left(f_{\text{val}}(v^k) + \tfrac{1}{2}\|v^k\|^2_{M^{-1}}\right)$$
$$= f_{\text{val}}(0) = \min\{\langle c, x\rangle \mid Ax - a = 0,\ x \in \mathcal{F}\}.$$

Remark 6. Note that Assumption 4 used in section 4.1 is obviously stronger than Assumption 5. In other words, the former implies the latter, but not vice versa. Hence, the results obtained in Propositions 6–8 will be valid under Assumption 4. In Proposition 8, we assume that problem (P_0) is bounded below. This assumption is natural and reasonable in optimization. Proposition 8 tells us that the optimal value of the unperturbed optimization problem (the original problem) can be achieved by solving a sequence of the conjugates which corresponds to the perturbed problems, in which affine equality constraints are perturbed on the right-hand side, and setting the perturbation parameters driven to zero. This result helps us to better understand the conjugate and Lagrange dual, and it might serve to study multistage stochastic nonlinear convex programs.

Also, this kind of perturbation problem is closely related to the perturbation problems discussed in [3]. In [9], Magnanti showed the equivalence between Fenchel dual and Lagrangian dual problems where the convex conjugate was employed. We believe that the results established in this subsection complement his theory to some extent. In addition, note that η is originally obtained from the Moreau–Yosida regularization by relaxing the original problem using the Lagrangian dual. Its conjugate η^*, as shown in Propositions 6–8, is related to the parametric (or perturbed) problem of the original problem. From this observation, we see that the perturbation analysis and Lagrangian dual are closely linked under the conjugate operation and Moreau–Yosida regularization. Besides the usual optimization methods, it also provides another possible option for solving some optimization problems, i.e., by solving the induced conjugate.

5. General convex objectives functions. In this section, we investigate the piecewise smoothness and semismoothness of the Lagrangian-dual function φ and the gradient g for the case of the general convex objective functions in (1). We will also provide an alternative way to study the semismoothness of the gradient g based on the structure of the epigraph of φ.

5.1. Convex objective functions with positive definite Hessian. We now discuss the case for the general convex objective functions in (1). Consider the following Lagrangian-dual function φ in (3):

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

When analyzing the piecewise smoothness of φ in section 3, we frequently use the fact that the optimal solutions of problem (9) lie on the boundary (or facets) of the set $\mathcal{F}$. This fact is guaranteed by Assumption 1, namely $u = A^T v - c \neq 0$, for the linear objective function $f(x) = c^T x$. For nonlinear objective functions, Assumption 1 cannot be made. Thus, multiple optimal solutions of problem (3) may appear in the interior of $\mathcal{F}$, and the piecewise C^2 smoothness of φ may probably be destroyed. This conjecture is confirmed by the following example where $\nabla^2\varphi$ is unbounded in some area, and thus φ is not piecewise C^2.

Example 3. Let

$$f(x) = \begin{cases} 0 & \text{if} \quad \rho \le 1, \\ (\rho - 1)^4 & \text{if} \quad \rho > 1, \end{cases}$$

where $x \in \mathbb{R}^2$ and $\rho = \sqrt{x_1^2 + x_2^2}$, and let

$$A = \begin{pmatrix} 1 & 1 \end{pmatrix}, \quad a = 0, \quad \mathcal{F} = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \le 4\}.$$

Obviously, $f(x)$ is convex and twice continuously differentiable on $\mathbb{R}^2$. After some manipulations we obtain

$$\varphi(v) = \sqrt{2}|v| + (3/4)|v|^{4/3}$$

for v in a neighborhood of zero, namely $\mathcal{N} = \{v \in \mathbb{R} : |v| < 2\sqrt{2}\}$. Since $\mathcal{F}$ is bounded, the effective domain of φ is the whole space $\mathbb{R}$. On $\mathbb{R} \setminus \mathcal{N}$, the function φ has a different form. For our purpose, the investigation of φ within $\mathcal{N}$ suffices. Thus we do not elaborate φ outside $\mathcal{N}$. For any $0 \ne v \in \mathcal{N}$, the first- and second-order derivatives of φ are

$$\varphi'(v) = \sqrt{2}\,\mathrm{sign}(v) + v^{1/3}, \qquad \varphi''(v) = (1/3)v^{-2/3}.$$

Now for any nonzero $v \to 0$, we have $\varphi''(v) \to \infty$. Using the same arguments as in Example 1, we can see that φ cannot be piecewise C^2 in the neighborhood $\mathcal{N}$.

This example shows that we cannot extend the results in sections 3 and 4 to problems with arbitrary convex objective functions. However, if the objective function $f(x)$ possesses a positive definite Hessian, we can obtain results similar to those in sections 3 and 4. Also, in this case, the constraints need not be strictly convex.

PROPOSITION 9. *Let φ be defined by* (3), *where f and f_i, $i \in \hat{I}$, are C^2 convex functions on $\mathbb{R}^n$. Suppose that the Hessian of f is positive definite, and for any facet Q of $\mathcal{F}$ with the index set I_Q and for any $x \in Q$, $\{\nabla f_i(x)\}_{i \in I_Q}$ are linearly independent. Suppose also that $\mathcal{F}$ is nonempty and bounded. Then the Lagrangian-dual function φ is piecewise C^2, and the gradient g of the Moreau–Yosida regularization η is piecewise smooth, and thereby semismooth, on $\mathbb{R}^m$.*

Proof. Similar to the analysis in section 3, we shall construct a piece corresponding to each facet of $\mathcal{F}$. There is one major difference we should highlight. For the problem with a nonlinear objective function, maximizers of the problem (3) can lie on the boundary as well as in the interior of $\mathcal{F}$, while in the case of linear objective functions, Assumption 1 prohibits interior maximizers. Thus, in the present case, an additional piece corresponding to the interior of $\mathcal{F}$ is needed.

Here we define the function ζ slightly differently from the approach in (10):

$$\zeta(u) = \sup\{u^T x - f(x) \mid x \in \mathcal{F}\}. \tag{28}$$

Then $\varphi(v) = \zeta(A^T v) - a^T v$. For each facet Q (on the boundary) of $\mathcal{F}$, we still construct a piece by a slightly different definition:

$$W := \left\{ (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^{|I_Q|} : f_i(x) = 0,\, i \in I_Q,\, \nabla^2 f(x) + \sum_{i \in I_Q} \lambda_i \nabla^2 f_i(x) \succ 0 \right\},$$

$$U := \left\{ u = \nabla f(x) + \sum_{i \in I_Q} \lambda_i \nabla f_i(x) \in \mathbb{R}^n : (x, \lambda) \in W \right\},$$

and

$$\Gamma(x,\lambda) := \begin{pmatrix} \nabla f(x) + \sum_{i\in I_Q} \lambda_i \nabla f_i(x) \\ \tilde{f}(x) \end{pmatrix}.$$

Then the result of Lemma 2 can be analogously proved, and a piece can be constructed.

In addition, a piece corresponding the interior of $\mathcal{F}$ will be constructed as follows. Since $\nabla^2 f(x) \succ 0$, for any $u \in \mathbb{R}^n$,

$$\nabla f(x) = u$$

has a unique solution, denoted by $\xi_{0x}(u)$. In other words, $\xi_{0x}(u)$ is the unique maximizer of the unconstrained problem

$$\max_{x\in\mathbb{R}^n} \{u^T x - f(x)\}.$$

Now, this piece is defined by $U_0 = \mathbb{R}^n$, $\zeta_0(u) = u^T\xi_{0x}(u) - f(\xi_{0x}(u))$, and $D_0 = \mathrm{cl}D_{\mathrm{int}}$, where $D_{\mathrm{int}} = \{u \mid \xi_{0x}(u) \in \mathrm{int}\mathcal{F}\}$. For any $u \in D_{\mathrm{int}}$, since the unique maximizer $\xi_{0x}(u)$ of the objective function $u^T x - f(x)$ is in the interior of the set $\mathcal{F}$, $\xi_{0x}(u)$ is the optimal solution to the constrained problem (28), too. Thus

$$\zeta(u) = u^T\xi_{0x}(u) - f(\xi_{0x}(u)) = \zeta_0(u). \tag{29}$$

Since ζ and ζ_0 are continuous, thus $\zeta(u) = \zeta_0(u)$ also holds for all $u \in D_0$. It is also easy to verify that

$$\nabla\zeta_0(u) = \xi_{0x}(u) \qquad \forall\, u \in U_0. \tag{30}$$

Now an analogue of the proof of Proposition 4 is valid to prove the piecewise-C^2 smoothness of ζ with the representation $\{(D_0, U_0, \zeta_0), (D_1, U_1, \zeta_2), \ldots, (D_q, U_q, \zeta_q)\}$. (The only difference is that now the nonnegative vector $\bar{\lambda}$ need not be nonzero since $u \neq 0$ is not assumed. Still, $\nabla^2 f(x) + \sum \bar{\lambda}_i \nabla^2 f_i(x) \succ 0$ because $\nabla^2 f(x) \succ 0$. This implies that the Jacobian of Γ is invertible.) Therefore, φ is piecewise C^2 on its domain.

The proof of the piecewise smoothness of g follows from Lemmas 5 and 6 and Proposition 5. The proofs of Lemmas 5 and 6 do not directly rely on Assumption 1, and thus they can be extended without changing to the representation $\{(E_0, V_0, \varphi_0), (E_1, V_1, \varphi_1), \ldots, (E_q, V_q, \varphi_q)\}$ of the Lagrangian-dual function φ of the present problem. ∎

5.2. Piecewise smoothness under the structure of the epigraph. In this subsection, we investigate the piecewise smoothness and the semismoothness of g using a different approach. In the analysis we will employ the piecewise smoothness or the semismoothness of the metric projection mapping under the structure of the epigraph of the underlying function. Our analysis is based on the framework of [11].

Recently, Meng, Sun, and Zhao [11] investigated the Moreau–Yosida regularization of a lower semicontinuous convex function, $\gamma : Z \to \mathbb{R} \cup \{+\infty\}$, and derived the semismoothness of the solution to the Moreau–Yosida regularization under the structure of the epigraph of γ. Here, Z is a finite dimensional vector space equipped with a scalar product, and the Moreau–Yosida regularization of γ is defined in the form of

$$\begin{aligned} \hat{\gamma}_\epsilon(u) := \min &\left\{\gamma(z) + \tfrac{\epsilon}{2}\langle u - z, u - z\rangle\right\} \\ \text{s.t. } & z \in Z, \end{aligned} \tag{31}$$

where ϵ is a positive number. Let Υ be the epigraph of γ; i.e., $\Upsilon := \text{epi}(\gamma) = \{(u,t) \in Z \times \mathbb{R} \mid t \geq \gamma(u)\}$. Noticing that Υ is a closed convex set, problem (31) then can be written as

$$\begin{aligned} \min \quad & \left\{\tfrac{1}{\epsilon}t + \tfrac{1}{2}\langle u-z, u-z\rangle\right\} \\ \text{s.t.} \quad & (z,t) \in \Upsilon. \end{aligned} \tag{32}$$

For any closed convex set D of Z and $z \in Z$, let $\Pi_D(z)$ denote the metric projection of z onto D, namely,

$$\Pi_D(z) := \text{argmin}\left\{\tfrac{1}{2}\|d-z\|^2 \mid d \in D\right\}.$$

Let $(z(u), t(u))$ be the unique optimal solution of (32), where $t(u) := \gamma(z(u))$. Define the mapping H by

$$H(z,t,u) := \begin{pmatrix} z \\ t \end{pmatrix} - \Pi_\Upsilon(G(z,t,u)),$$

where $G(z,t,u) := (u^T \;\; t - 1/\epsilon)^T$. Then, it follows from [11] that

$$H(z(u), t(u), u) = 0, \quad G(z(u), t(u), u) \notin \Upsilon \quad \forall\, u \in Z.$$

The following proposition is taken from [11, Theorem 4].

PROPOSITION 10. *For $u_0 \in Z$, let $z_0 := z(u_0)$ and $t_0 := \gamma(z(u_0))$. Then, $(z(\cdot), t(\cdot))$ is semismooth at u_0 if $\Pi_\Upsilon(G(z_0,t_0,u_0))_z \in \text{int}(\text{dom}\gamma)$ and $\Pi_\Upsilon(\cdot)$ is semismooth at $G(z_0,t_0,u_0)$.*

Here we consider the case where $M = \lambda I$ in the Moreau–Yosida regularization as defined in (5), where I is the identity matrix of $\mathbb{R}^{m\times m}$ and $\lambda > 0$. For $v \in \mathbb{R}^m$, let $w(v)$ denote the unique solution of (5), $s(v) := \varphi(w(v))$, and epi(φ) denote the epigraph of φ. Evidently, $(w(v), s(v))$ is the unique solution of

$$\begin{aligned} \min \quad & \left\{\tfrac{1}{\lambda}s + \tfrac{1}{2}\langle v-w, v-w\rangle\right\} \\ \text{s.t.} \quad & (w,s) \in \text{epi}(\varphi), \end{aligned}$$

which is a reformulation of (5). Note that

$$g(v) = \nabla\eta(v) = \lambda(v - w(v)).$$

Hence to study the semismoothness of g, we need only to study the properties of $w(\cdot)$. Set

$$\Phi(w,s,v) := \begin{pmatrix} w \\ s \end{pmatrix} - \begin{pmatrix} w - v \\ 1/\lambda \end{pmatrix} = \begin{pmatrix} v \\ s - 1/\lambda \end{pmatrix}.$$

According to Proposition 10 and following the arguments as in [11], we then have the following result.

PROPOSITION 11. *For $\bar{v} \in \mathbb{R}^m$, let $\bar{w} := w(\bar{v})$, $\bar{s} := \varphi(w(\bar{v}))$. Suppose that $\Pi_{\text{epi}(\varphi)}(\Phi(\bar{w},\bar{s},\bar{v}))_w \in \text{int}(\text{dom}\varphi)$ and $\Pi_{\text{epi}(\varphi)}(\cdot)$ is semismooth at $\Phi(\bar{w},\bar{s},\bar{v})$. Then $(w(\cdot), s(\cdot))$ is semismooth at $\bar{v}$. Thereby, g is semismooth at $\bar{v}$.*

Furthermore, if φ is finite valued everywhere and $\Pi_{\text{epi}(\varphi)}(\cdot)$ is semismooth on $\mathbb{R}^m \times \mathbb{R}$, then g is semismooth on $\mathbb{R}^m$.

Similar to the mapping H above, we define a mapping Ξ corresponding to the regularization (5),

$$\Xi(w,s,v) := \begin{pmatrix} w \\ s \end{pmatrix} - \Pi_{\mathrm{epi}(\varphi)}(\Phi(w,s,v)).$$

Thus, for any $v \in \mathbb{R}^m$

$$\Xi(w(v), s(v), v) = 0. \tag{33}$$

We now obtain the following result concerning the piecewise smoothness of g.

PROPOSITION 12. *Let* $\bar{v} \in \mathbb{R}^m$. *Suppose that* (i) $\Pi_{\mathrm{epi}(\varphi)}(\Phi(\bar{w},\bar{s},\bar{v}))_w \in \mathrm{int}(\mathrm{dom}\varphi)$, *and* (ii) $\Pi_{\mathrm{epi}(\varphi)}(\cdot)$ *is piecewise* C^k *on a neighborhood* $\mathcal{N}_1$ *of* $(\bar{v}, \varphi(w(\bar{v})) - 1/\lambda)$, *where* $\bar{w} = w(\bar{v})$ *and* $\bar{s} = \varphi(w(\bar{v}))$. *Then,* $(w(\cdot), s(\cdot))$ *is piecewise* C^k *on a neighborhood* $\mathcal{N}_2$ *of* $\bar{v}$. *Thereby,* g *is piecewise* C^k *on* $\mathcal{N}_2$. *In particular,* g *is semismooth on a neighborhood of* $\bar{v}$.

Proof. Define a mapping $\aleph : \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^m$ by

$$\aleph(w,s,v) = \begin{pmatrix} \Xi(w,s,v) \\ v - \bar{v} \end{pmatrix}.$$

By assumption, since $\Pi_{\mathrm{epi}(\varphi)}(\cdot)$ is piecewise C^k on $\mathcal{N}_1$, it is easy to see that $\aleph(\cdot)$ is piecewise C^k on some neighborhood of $(\bar{w},\bar{s},\bar{v})$, and

$$\aleph(\bar{w},\bar{s},\bar{v}) = 0. \tag{34}$$

Next, we show that every matrix in $\partial\aleph(\bar{w},\bar{s},\bar{v})$ is nonsingular [2]. To do so, it is not hard to see that we only need to show the nonsingularity of $\pi_{(w,s)}\partial\Xi(\bar{w},\bar{s},\bar{v})$. For any $V \in \pi_{(w,s)}\partial\Xi(\bar{w},\bar{s},\bar{v})$, it follows that there exists $W \in \partial\Pi_{\mathrm{epi}(\varphi)}(\Phi(\bar{w},\bar{s},\bar{v}))$ such that

$$V = I_{m+1} - W\left(I_{m+1} - \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}\right),$$

where W is a convex combination of some finitely many matrices in $\partial_B\Pi_{\mathrm{epi}(\varphi)}(\Phi(\bar{w}, \bar{s},\bar{v}))$. Suppose $W_i \in \partial_B\Pi_{\mathrm{epi}(\varphi)}(\Phi(\bar{w},\bar{s},\bar{v}))$ and $\lambda_i \geq 0$, $i = 1,\ldots,\nu$, satisfying $\sum_{i=1}^{\nu}\lambda_i = 1$, such that $W = \sum_{i=1}^{\nu}\lambda_i W_i$, where each W_i is in the form of $W_i = \left[\begin{smallmatrix} U_i & \alpha_i \\ \alpha_i^T & \beta_i \end{smallmatrix}\right]$ with $U_i \in \mathbb{R}^{m\times m}$, $\alpha_i \in \mathbb{R}^m$, and $\beta_i \geq 0$. Thus,

$$W = \begin{bmatrix} \sum_{i=1}^{\nu}\lambda_i U_i & \sum_{i=1}^{\nu}\lambda_i\alpha_i \\ \sum_{i=1}^{\nu}\lambda_i\alpha_i^T & \sum_{i=1}^{\nu}\lambda_i\beta_i \end{bmatrix}.$$

To ease the notation, we write $W = \left[\begin{smallmatrix} U & \alpha \\ \alpha^T & \beta \end{smallmatrix}\right]$. Then, by [11, Proposition 3], there exists $\varrho_i \in (0,1)$, $i = 1,\ldots,\nu$, such that

$$0 \leq \beta_i \leq \varrho_i < 1 \qquad \forall\, i.$$

Hence,

$$0 \leq \beta < 1. \tag{35}$$

Then, we have

$$V = I_{m+1} - \begin{bmatrix} U & \alpha \\ \alpha^T & \beta \end{bmatrix} \left(I_{m+1} - \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} \right)$$
$$= I_{m+1} - \begin{bmatrix} U & \alpha \\ \alpha^T & \beta \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$
$$= I_{m+1} - \begin{bmatrix} 0 & \alpha \\ 0 & \beta \end{bmatrix} = \begin{bmatrix} I_m & -\alpha \\ 0 & 1-\beta \end{bmatrix}.$$

This together with (35) implies that $\det V = 1 - \beta > 0$ for any $V \in \pi_{(w,s)}\partial\Xi(\bar{w}, \bar{s}, \bar{v})$. So, $\Xi(w, s, v)$ is coherently oriented with respect to w and s at $(\bar{w}, \bar{s}, \bar{v})$ [17, 20]. Thereby, $\pi_{(w,s)}\partial\Xi(\bar{w}, \bar{s}, \bar{v})$ is nonsingular, and so is $\partial\aleph(\bar{w}, \bar{s}, \bar{v})$. Then, by [15, Theorem 6], $\aleph$ is a locally Lipschitz homeomorphism near $(\bar{w}, \bar{s}, \bar{v})$, and $\operatorname{sgn}\det V = \operatorname{ind}(\aleph, (\bar{w}, \bar{s}, \bar{v})) = \pm 1$ for any $V \in \partial_B\aleph(\bar{w}, \bar{s}, \bar{v})$. Further, noticing that $\aleph(\cdot)$ is coherently oriented at $(\bar{w}, \bar{s}, \bar{v})$ and is piecewise C^k on a neighborhood of $(\bar{w}, \bar{s}, \bar{v})$, then by [17, Theorem 5], it follows that $\aleph$ is a PC^k-homeomorphism near $(\bar{w}, \bar{s}, \bar{v})$. Thus, the desired results follow immediately. This completes the proof. ∎

Remark 7. The condition $\Pi_{\text{epi}(\varphi)}(\Phi(\bar{w}, \bar{s}, \bar{v}))_w \in \text{int}(\text{dom}\varphi)$ in Proposition 12 holds automatically if φ is finite valued everywhere. The obtained results complement and enrich the framework of piecewise smooth functions [20, 17], and also enhance the recent results on the Moreau–Yosida regularization [11].

6. Conclusion. The Lagrangian dual is widely used for large-scale problems. A significant feature of the Lagrangian-dual function φ is the piecewise smoothness, which is studied in this paper and employed in the analysis of the Moreau–Yosida regularization of φ. We investigate the semismoothness of the gradient g of the Moreau–Yosida regularization of φ, which plays a key role in the superlinear or quadratic convergence analysis of generalized Newton methods for solving nonsmooth equations. As to problem (1) with the linear objective function, we show that the Lagrangian-dual function φ is piecewise C^2 and the gradient g is piecewise smooth and thereby semismooth if the inequality constraints in (1) either are all affine or all possess positive definite Hessian matrices. An example with an affine constraint and a strictly convex constraint is constructed. We find that the Lagrangian-dual function of this problem is *not* piecewise C^2, and that the gradient g of its Moreau–Yosida regularization *is* still semismooth. However, whether or not g is semismooth for general mixed affine and strictly convex constraints is still left unanswered. We also investigate problem (1) with a convex objective function. We show with an example that φ may not be piecewise C^2 for the problem with a general convex objective function. For problem (1) with an objective function which possesses a positive definite Hessian, φ and g can again be shown to be piecewise C^2 and semismooth, respectively. We have also provided an alternative way to study the semismoothness/piecewise smoothness of g under the structure of the epigraph of the Lagrangian dual function using the projection operator. For problem (1) with a linear objective, we have also established an interesting result characterizing the relations between the original problem and the Fenchel conjugate of the regularization of the Lagrangian dual problem. For future research, we will examine under which conditions the projection mapping over the epigraph of the Lagrangian-dual function φ is piecewise smooth or semismooth.

Acknowledgment. We would like to thank the two anonymous referees for their constructive suggestions, which helped improve the presentation of this paper.

REFERENCES

[1] X. Chen, Z. Nashed, and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1200–1216.
[2] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983.
[3] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer-Verlag, New York, 1998.
[4] A. L. Dontchev, H.-D. Qi, and L. Qi, *Convergence of Newton method for convex best interpolation*, Numer. Math., 87 (2001), pp. 435–456.
[5] A. L. Dontchev, H.-D. Qi, and L. Qi, *Quadratic convergence of Newton method for convex interpolation and smoothing*, Constr. Approx., 19 (2003), pp. 1230–1143.
[6] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.
[7] M. Fukushima and L. Qi, *A globally and superlinearly convergent algorithm for nonsmooth convex minimization*, SIAM J. Optim., 6 (1996), pp. 1106–1120.
[8] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin, 1993.
[9] T. L. Magnanti, *Fenchel and Lagrange dual are equivalent*, Math. Program., 7 (1974), pp. 253–258.
[10] F. Meng and Y. Hao, *The property of piecewise smoothness of Moreau-Yosida approximation for a piecewise C^2 convex function*, Adv. Math. (China), 30 (2001), pp. 354–358.
[11] F. Meng, D. Sun, and G. Zhao, *Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization*, Math. Program., 104 (2005), pp. 561–581.
[12] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 959–972.
[13] R. Mifflin, L. Qi, and D. Sun, *Properties of the Moreau-Yosida regularization of a piecewise C^2 convex function*, Math. Program., 84 (1999), pp. 269–281.
[14] J. J. Moreau, *Proximite et dualite dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.
[15] J.-S. Pang, D. Sun, and J. Sun, *Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems*, Math. Oper. Res., 28 (2003), pp. 39–63.
[16] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367.
[17] D. Ralph and S. Scholtes, *Sensitivity analysis of composite piecewise smooth equations*, Math. Program., 76 (1997), pp. 593–612.
[18] R. T. Rockafellar, *Convex Analysis*, Princeton, NJ, 1970.
[19] A. Ruszczyński and A. Shapiro, *Stochastic Programming, Handbook in Operations Research and Management Science*, A. Ruszczyński and A. Shapiro, eds., Elsevier Science, Amsterdam, 2003.
[20] S. Scholtes, *Introduction to Piecewise Smooth Equations*, Habilitation, University of Karlsruhe, Karlsruhe, Germany, 1994.
[21] D. Sun and J. Han, *On a conjecture in Moreau-Yosida regularization of a nonsmooth convex function*, Chinese Sci. Bull., 42 (1997), pp. 1140–1143.
[22] C. Y. Wang, X. Q. Yang, and X. M. Yang, *Nonlinear Lagrange duality theorems and penalty function methods in continuous optimization*, J. Global Optim., 27 (2003), pp. 473–484.
[23] K. Yosida, *Functional Analysis*, Springer-Verlag, Berlin, 1964.
[24] G. Zhao, *A Lagrangian dual method with self-concordant barrier for multi-stage stochastic convex nonlinear programming*, Math. Program., 102 (2005), pp. 1–24.