

# THE RATE OF CONVERGENCE OF AUGMENTED LAGRANGE METHOD FOR A COMPOSITE OPTIMIZATION PROBLEM

LIWEI ZHANG\*, JIHONG ZHANG†, AND YULE ZHANG‡

**Abstract.** In this paper we analyze the rate of local convergence of the augmented Lagrange method for solving optimization problems with equality constraints and the objective function expressed as the sum of a convex function and a twice continuously differentiable function. The presence of the non-smoothness of the convex function in the objective requires extensive tools such as the second-order variational analysis on the Moreau-Yosida regularization of a convex function and matrix techniques for estimating the generalized Hessian of the dual function defined by the augmented Lagrange. With two conditions, we prove that, the rate of convergence for the augmented Lagrange method is linear and the ratio constant is proportional to  $1/c$ , where  $c$  is the penalty parameter that exceeds a threshold  $\bar{c} > 0$ . As an illustrative example, for nonlinear semidefinite programming problem, we show that the assumptions used in Sun et al. [13], for the rate of convergence of the augmented Lagrange method, are sufficient to the two conditions adopted in this paper.

**Key words.** composite optimization problem, augmented Lagrange method, the rate of convergence, Moreau-Yosida regularization, variational analysis

**AMS subject classifications.** 49K40, 90C31, 49J53

**1. Introduction.** We consider the following composite optimization problem

$$(1) \quad \begin{aligned} \min \quad & f(x) + g(x) \\ \text{s.t.} \quad & h(x) = 0, \end{aligned}$$

where  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is proper lower semicontinuous convex function,  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  is a twice continuously differentiable real-valued function, and  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a twice continuously differentiable mapping. Optimization model (1) with linear constraints covers a large number of important practical optimization problems from compressive sensing and statistical learning. Besides this, any smooth conic optimization problem can be transformed as a such model. Assume a conic optimization problem is of the following form:

$$(2) \quad \begin{aligned} \min \quad & m(z) \\ \text{s.t.} \quad & q(z) \in \mathcal{K}, \end{aligned}$$

where  $m : \mathbb{R}^p \rightarrow \mathbb{R}$  is a smooth function,  $\mathcal{K} \subset Y$  is a closed convex cone,  $Y$  is a finite-dimensional Hilbert space and  $q : \mathbb{R}^p \rightarrow Y$  is a smooth mapping. If we define  $x = (z, y) \in \mathbb{R}^p \times Y$ ,  $g(x) = m(z)$ ,  $h(x) = q(z) - y$  and  $f(x) = \delta_{\mathcal{K}}(y)$ , then Problem (2) is reformulated as model (1).

The augmented Lagrangian method was initiated by Hestenes [4] and Powell [8] for solving the smooth equality constrained problem and was generalized by Rockafellar [9] to the following nonlinear programming problem

$$(3) \quad \min f(x) \quad \text{s.t.} \quad h(x) = 0, \quad g(x) \geq 0,$$

---

\*School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, P.R. China (lwzhang@dlut.edu.cn). The research of this author was supported by the National Natural Science Foundation of China under project No. 91330206 and No. 11571059.

†School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, P.R. China (jihongzhang@mail.dlut.edu.cn).

‡School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, P.R. China (yulezhang@mail.dlut.edu.cn).

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^q$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$  are twice continuously differentiable. Problem (3) is a special case of (2) with  $Y := \mathbb{R}^q \times \mathbb{R}^p$  and  $K = \{0_q\} \times \mathbb{R}_+^p$ .

For the equality constrained optimization problem, Powell [8] proved that if the linear independence constraint qualification and the second-order sufficient condition are satisfied, then the augmented Lagrangian method can converge locally at a linear rate proportional to  $1/c$  when the penalty parameter  $c > 0$  is large enough. For convex programming, Rockafellar [9] established a saddle point theorem in terms of the augmented Lagrangian and Rockafellar [10] proved the global convergence of the augmented Lagrangian method for any  $c > 0$ . In Chapter 2 of [1], Bertsekas demonstrated the linear rate of convergence of the augmented Lagrangian method for the equality constrained problem with the ratio constant proportional to  $1/c$ . In Chapter 3 of [1], Bertsekas explained that the result for equality constrained problem can be applied to the nonlinear programming problem (3) when the strict complementarity condition holds. Importantly, without assuming the strict complementarity condition, Conn et al. [5] derived linear convergence rate of the augmented Lagrangian method for nonlinear programming problem.

For the nonlinear semidefinite optimization problem

$$(4) \quad \min f(x) \quad \text{s.t.} \quad h(x) = 0, \quad g(x) \in \mathbb{S}_+^p,$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^q$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{S}^p$  are twice continuously differentiable. Problem (4) is a special case of (2) with  $Y := \mathbb{R}^q \times \mathbb{S}^p$  and  $K = \{0_q\} \times \mathbb{S}_+^p$ . Sun et al. [13] studied, without assuming the strict complementarity, the rate of convergence of the augmented Lagrangian method for solving the nonlinear semidefinite programming problem. The analysis in [13] is heavily dependent on the variational properties of the projection operator in the symmetric matrix space.

In this paper, with the help of Moreau-Yosida regularization of a convex function, we apply second-order analysis of  $C^{1,1}$  functions to analyze the rate of convergence of the augmented Lagrange method for Problem (1).

The paper is organized as follows. In Section 2, we propose a set of second-order sufficient optimality conditions for the problem minimizing a  $C^{1,1}$  function subject to twice continuously differentiable equalities. In Section 3, we establish a set of second-order sufficient optimality conditions for Problem (1) and give the so-called strong second-order sufficient optimality conditions from which properties of the dual function based on the augmented Lagrangian are developed. Section 4 develops a general theory on the rate of convergence of the augmented Lagrangian method for Problem (1). In Section 5, we take the nonlinear semidefinite optimization problem as an example to show the assumptions used in [13] are sufficient to our conditions required in the analysis for the rate of convergence, and moreover we draw a conclusion for this paper as well.

**2. Preliminaries.** Consider the equality constrained optimization problem

$$(5) \quad \begin{aligned} \min \quad & \varphi_0(x) \\ \text{s.t.} \quad & h(x) = 0, \end{aligned}$$

where  $\varphi_0 : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  is an extended real-valued function and  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a mapping. The Lagrangian of Problem (5) is

$$L(x, \mu) = \varphi_0(x) + \mu^T h(x).$$

Now we give and prove the second-order necessary optimality conditions for Problem (5) when  $\varphi_0$  is a  $\mathbb{C}^{1,1}$  function and  $h$  is twice differentiable mapping around a local minimizer.

Let  $C = \{x \in \mathfrak{R}^n : h(x) = 0\}$  be the feasible set of Problem (5). For  $\bar{x} \in C$ , define the critical cone of Problem (5) at  $\bar{x}$  by

$$\mathcal{C}(\bar{x}) = \{d \in \mathfrak{R}^n : \mathcal{J}h(\bar{x})d = 0, \nabla\varphi_0(\bar{x})^T d \leq 0\}.$$

LEMMA 1. *If  $\bar{x}$  is a local minimizer of Problem (5),  $\varphi_0$  is a  $\mathbb{C}^{1,1}$  function and  $h$  is twice differentiable in a neighborhood of  $\bar{x}$ . Assume that  $\nabla\varphi_0$  is directionally differentiable in a neighborhood of  $\bar{x}$  and  $\mathcal{J}h(x)$  is of full rank in row. Then there exists a unique Lagrange multiplier  $\bar{\mu}$  such that*

$$\nabla_x L(\bar{x}, \bar{\mu}) = 0.$$

And for any  $d \in \mathcal{C}(\bar{x})$ ,

$$L''_x(\bar{x}, \bar{\mu}; d, d) \geq 0,$$

where  $L''_x(\bar{x}, \bar{\mu}; d, d) = d^T [\nabla\varphi_0]'(\bar{x}; d) + \sum_{i=1}^m d^T \nabla^2 h_i(\bar{x}) d$ .

*Proof.* Since  $\mathcal{J}h(x)$  is of full rank in row, we have  $T_C(\bar{x}) = \ker \mathcal{J}h(\bar{x})$  and

$$T_C^2(\bar{x}, d) = \{w : \mathcal{J}h(\bar{x})w + D^2h(\bar{x})(d, d) = 0\}$$

for  $d \in T_C(\bar{x})$ , where  $T_C^2(\bar{x}, d)$  is the second-order outer set of  $C$  at  $\bar{x}$  with respect to  $d$ , for its definition, see [2] or [14]. Then

$$\varphi_0''(\bar{x}; d, w) \geq 0, \quad \forall w \in T_C^2(\bar{x}; d)$$

for  $d \in \mathcal{C}(\bar{x})$ . Nothing that  $\varphi_0''(\bar{x}; d, w) = \nabla\varphi_0(\bar{x})^T w + d^T (\nabla\varphi_0)'(\bar{x}; d)$ , we have that the optimal value of the problem

$$\begin{aligned} \min \quad & \nabla\varphi_0(\bar{x})^T w + d^T (\nabla\varphi_0)'(\bar{x}; d) \\ \text{s.t.} \quad & \mathcal{J}h(\bar{x})w + D^2h(\bar{x})(d, d) = 0 \end{aligned}$$

is non-negative for each  $d \in \mathcal{C}(\bar{x})$ . The result follows from the duality theory directly.  $\square$

LEMMA 2. *Let  $\bar{x} \in C$  and  $\bar{\lambda} \in \mathfrak{R}^m$  satisfy  $\nabla_x L(\bar{x}, \bar{\lambda}) = 0$ . Suppose that, for any  $d \in \mathcal{C}(\bar{x}) \setminus \{0_n\}$ ,*

$$(6) \quad d^T [W + \sum_{i=1}^m \bar{\lambda}_i \nabla^2 h_i(\bar{x})] d > 0, \quad \forall W \in \partial(\nabla\varphi_0)(\bar{x}).$$

*Then the second-order growth condition holds at  $\bar{x}$ , namely there exist  $c_0 > 0$  and  $\varepsilon_0 > 0$  such that*

$$\varphi_0(x) \geq \varphi_0(\bar{x}) + c_0 \|x - \bar{x}\|^2, \quad \forall x \in C \cap \mathbb{B}_{\varepsilon_0}(\bar{x}).$$

*Proof.* By contradiction. Suppose that there are  $\varepsilon_k \searrow 0$  and  $x^k \rightarrow \bar{x}$  with  $h(x^k) = 0$  such that

$$\varphi_0(x^k) < \varphi_0(\bar{x}) + \varepsilon_k \|x^k - \bar{x}\|^2.$$

Let  $t^k = \|x^k - \bar{x}\|$  and  $d^k = t_k^{-1}(x^k - \bar{x})$ , then  $x^k = \bar{x} + t_k d^k$ . Suppose that  $\{d^k\}$  has an accumulation point  $d$ , namely there exists  $\{k_i\} \subset \mathbb{N}$  such that  $d^{k_i} \rightarrow d$ . It is easy to prove that  $\|d\| = 1$ ,  $\mathcal{J}h(\bar{x})d = 0$  and  $\nabla\varphi_0(\bar{x})^T d \leq 0$ , which implies  $d \in \mathcal{C}(\bar{x}) \setminus \{0\}$ . By using the Taylor expansion, we have

$$0 = h(x^{k_i}) = h(\bar{x} + t_{k_i} d^{k_i}) = h(\bar{x}) + t_{k_i} \mathcal{J}h(\bar{x}) d^{k_i} + \frac{t_{k_i}^2}{2} D^2 h(\bar{x})(d^{k_i}, d^{k_i}) + o(t_{k_i}^2)$$

where  $D^2 h(\bar{x})(w, w) = (w^T \nabla^2 h_1(\bar{x}) w, \dots, w^T \nabla^2 h_m(\bar{x}) w)^T$ , and

$$\begin{aligned} \varepsilon_{k_i} t_{k_i}^2 + \varphi_0(\bar{x}) &> \varphi_0(\bar{x} + t_{k_i} d^{k_i}) \\ &= \varphi_0(\bar{x}) + t_{k_i} \nabla\varphi_0(\bar{x})^T d^{k_i} + \frac{t_{k_i}^2}{2} d^{k_i T} [\nabla\varphi_0]'(\bar{x}; d^{k_i}) + o(t_{k_i}^2). \end{aligned}$$

Therefore, we obtain

$$(7) \quad 0 = t_{k_i} \mathcal{J}h(\bar{x}) d^{k_i} + \frac{t_{k_i}^2}{2} D^2 h(\bar{x})(d^{k_i}, d^{k_i}) + o(t_{k_i}^2)$$

and

$$(8) \quad \varepsilon_{k_i} t_{k_i}^2 > t_{k_i} \nabla\varphi_0(\bar{x})^T d^{k_i} + \frac{t_{k_i}^2}{2} d^{k_i T} [\nabla\varphi_0]'(\bar{x}; d^{k_i}) + o(t_{k_i}^2).$$

Premultiplying  $\mu^T$  to the both sides of (7) and adding it to both sides of (8), we obtain

$$\varepsilon_{k_i} t_{k_i}^2 > t_{k_i} [\nabla_x L(\bar{x}; \bar{\mu})]^T d^{k_i} + \frac{t_{k_i}^2}{2} [d^{k_i T} (\nabla\varphi_0)'(\bar{x}; d^{k_i}) + \sum_{j=1}^m \bar{\mu}_j d^{k_i T} \nabla^2 h_j(\bar{x}) d^{k_i}] + o(t_{k_i}^2).$$

Nothing that  $\nabla_x L(\bar{x}, \bar{\mu}) = 0$ , we have from the above that

$$\varepsilon_{k_i} > \frac{1}{2} [d^{k_i T} (\nabla\varphi_0)'(\bar{x}; d^{k_i}) + \sum_{j=1}^m \bar{\mu}_j d^{k_i T} \nabla^2 h_j(\bar{x}) d^{k_i}] + \frac{o(t_{k_i}^2)}{t_{k_i}^2}$$

Noting the continuity  $d' \rightarrow (\nabla\varphi_0)'(\bar{x}; d')$ , taking the limit of the above inequality, one obtain

$$0 \geq d^T (\nabla\varphi_0)'(\bar{x}; d) + \sum_{j=1}^m \bar{\mu}_j d^T \nabla^2 h_j(\bar{x}) d.$$

There exists  $W \in \partial(\nabla\varphi_0)(\bar{x})$  satisfying  $(\nabla\varphi_0)'(\bar{x}; d) = Wd$ , so that

$$0 \geq d^T [W + \sum_{j=1}^m \bar{\mu}_j \nabla^2 h_j(\bar{x})] d,$$

contradicting with (6).  $\square$

**3. Second-order sufficient optimality.** Now we focus on the composite optimization problem (1) in which  $f : \mathfrak{R}^n \rightarrow \bar{\mathfrak{R}}$  is proper lower semicontinuous convex function, and  $g$  and  $h$  are twice continuously differentiable. We define the Moreau-Yosida envelope of the function  $f$  by

$$e_c f(x) = \inf_{w \in \mathfrak{R}^n} \left\{ f(w) + \frac{c}{2} \|w - x\|^2 \right\},$$

and the proximal mapping  $\mathcal{P}_c f$  by

$$\mathcal{P}_c f(x) = \operatorname{argmin}_{w \in \mathbb{R}^n} \left\{ f(w) + \frac{c}{2} \|w - x\|^2 \right\}.$$

Then it follows from Chapter 2 of [14] that  $e_c f$  is continuously differentiable and convex, with

$$\nabla e_c f(x) = c[x - \mathcal{P}_c f(x)].$$

Therefore, we obtain

$$(9) \quad \partial[\nabla e_c f](x) = \{c[I - V] : V \in \partial \mathcal{P}_c f(x)\}.$$

Let  $\Phi$  be the feasible set for Problem (1), i.e.,

$$\Phi = \{x \in \mathbb{R}^n : h(x) = 0\}.$$

For  $\bar{x} \in \Phi$ , we define the critical cone of Problem (1) at  $\bar{x}$  by

$$\mathcal{C}(\bar{x}) = \{d \in \ker \mathcal{J}h(\bar{x}) : f'(\bar{x}, d) + \nabla g(\bar{x})^T d \leq 0\}.$$

Let  $\bar{x} \in \Phi$  be a given point and  $g$  and  $h$  be twice continuously differentiable in a neighborhood of  $\bar{x}$ . We propose the following three conditions:

- (I)** (Stationary Condition) There exists a vector  $\bar{\mu} \in \mathbb{R}^m$  such that  $0 \in \partial f(\bar{x}) + \nabla g(\bar{x}) + \mathcal{J}h(\bar{x})^T \bar{\mu}$ .
- (II)** (Second-order Sufficiency Optimality Condition) There exists  $c_0 > 0$  such that

$$d^T [c_0 c^{-1} W + \nabla^2 g(\bar{x}) + \sum_{i=1}^m \bar{\mu}_j \nabla^2 h_i(\bar{x})] d > 0, \quad \forall W \in \partial(\nabla e_c f)(\bar{x} - \bar{\lambda}/c)$$

when  $c \geq c_0$  and  $d \in \mathcal{C}(\bar{x}) \setminus \{0\}$ , where  $\bar{\lambda} = \nabla g(\bar{x}) + \mathcal{J}h(\bar{x})^T \bar{\mu}$ .

- (III)** (Strong Second-order Sufficiency Optimality Condition) There exists  $c_0 > 0$  such that

$$d^T [c_0 c^{-1} W + \nabla^2 g(\bar{x}) + \sum_{i=1}^m \bar{\mu}_j \nabla^2 h_i(\bar{x})] d > 0, \quad \forall W \in \partial(\nabla e_c f)(\bar{x} - \bar{\lambda}/c)$$

when  $c \geq c_0$  and  $d \in \ker \mathcal{J}h(\bar{x}) \setminus \{0\}$ , where  $\bar{\lambda} = \nabla g(\bar{x}) + \mathcal{J}h(\bar{x})^T \bar{\mu}$ .

For  $d \in \mathcal{C}(\bar{x})$ , noting the definition of  $\bar{\lambda}$ , we have  $d \in \ker \mathcal{J}h(\bar{x})$  and  $f'(\bar{x}, d) + \bar{\lambda}^T d \leq 0$ . Since the epi-derivative of  $f$  at  $\bar{x}$  along  $d$ , denoted by  $f^\downarrow(\bar{x}, d)$  (see [2]), satisfies

$$\delta^*(d | \partial f(\bar{x})) = f^\downarrow(\bar{x}, d) \leq f'(\bar{x}, d),$$

and  $0 \in \partial f(\bar{x}) + \bar{\lambda}$ , we have for  $d \in \mathcal{C}(\bar{x})$  that

$$0 \leq \delta^*(d | \partial f(\bar{x}) + \bar{\lambda}) \leq 0,$$

which is equivalent to  $\delta^*(d | \partial f(\bar{x}) + \bar{\lambda}) = 0$ .

**LEMMA 3.** *Let  $\bar{x} \in \mathbb{R}^n$  be a given point. Suppose that Conditions (I) and (II) are satisfied. Then the second-order growth condition for Problem (1) holds at  $\bar{x}$ .*

*Proof.* Nothing that  $0 \in \partial f(\bar{x}) + \bar{\lambda}$  implies that

$$\nabla e_c f(\bar{x} - \bar{\lambda}/c) + \nabla g(\bar{x}) + \mathcal{J}h(\bar{x})^T \bar{\mu} = 0,$$

we have that  $(\bar{x}, \bar{\mu})$  is the KKT point to the following differentiable optimization problem

$$(10) \quad \min e_c f(x - \bar{\lambda}/c) + g(x) \quad \text{s.t.} \quad h(x) = 0.$$

Obviously, because of the positive semi-definiteness of  $W$ , Condition (II) implies that

$$d^T [W + \nabla^2 g(\bar{x}) + \sum_{i=1}^m \bar{\mu}_j \nabla^2 h_i(\bar{x})] d > 0, \quad \forall W \in \partial(\nabla e_c f)(\bar{x} - \bar{\lambda}/c)$$

when  $c \geq c_0$  and  $d \in \mathcal{C}(\bar{x}) \setminus \{0\}$ . Therefore (II) implies the second-order sufficient optimality condition, stated in Lemma 2, for Problem (10). It follows from Lemma 2 that the second-order growth condition for Problem (10) holds at  $\bar{x}$ . There exist  $c_0 > 0, \varepsilon_0 > 0$  satisfying

$$(11) \quad e_c f(x - \bar{\lambda}/c) + g(x) \geq e_c f(\bar{x} - \bar{\lambda}/c) + g(\bar{x}) + c_0 \|x - \bar{x}\|^2, \quad \forall x \in \Phi \cap \mathbb{B}(\bar{x}, \varepsilon_0).$$

Let

$$\bar{z}_c = \arg \min f(z) + \frac{c}{2} \|z - \bar{x} + \bar{\lambda}/c\|^2,$$

then

$$0 \in \partial f(\bar{z}_c) + c[\bar{z}_c - \bar{x} + \bar{\lambda}/c],$$

and  $\bar{z}_c = \bar{x}$  is the unique solution to  $\min f(z) + \frac{c}{2} \|z - \bar{x} + \bar{\lambda}/c\|^2$ . Therefore, we obtain

$$(12) \quad e_c f(\bar{x} - \bar{\lambda}/c) = f(\bar{x}) + \frac{1}{2c} \|\bar{\lambda}\|^2.$$

On the other hand,

$$(13) \quad \begin{aligned} e_c f(x - \bar{\lambda}/c) &= \inf_z [f(z) + \frac{c}{2} \|z - x + \bar{\lambda}/c\|^2] \\ &\leq f(x) + \frac{1}{2c} \|\bar{\lambda}\|^2. \end{aligned}$$

Combining (12), (13) with (11), we have

$$f(x) + g(x) + \frac{1}{2c} \|\bar{\lambda}\|^2 \geq f(\bar{x}) + g(\bar{x}) + \frac{1}{2c} \|\bar{\lambda}\|^2 + c_0 \|x - \bar{x}\|^2, \quad \forall x \in \Phi \cap \mathbb{B}(\bar{x}, \varepsilon_0),$$

which is just the second-order growth condition of Problem (1) at  $\bar{x}$ .  $\square$

We can rewrite Problem(1) as

$$(14) \quad \begin{aligned} \min \quad & f(x) + g(z) \\ \text{s.t.} \quad & h(z) = 0, \\ & x - z = 0. \end{aligned}$$

The augmented lagrangian for (14) is defined by

$$L_c(x, z, \mu, \lambda) = f(x) + g(z) + \langle \mu, h(z) \rangle + \langle \lambda, x - z \rangle + \frac{c}{2} \|h(z)\|^2 + \frac{c}{2} \|x - z\|^2.$$

Define

$$\begin{aligned}\theta_c(\mu, \lambda) &= \inf_{x,z} L_c(x, z, \mu, \lambda) = \inf_z \inf_x L_c(x, z, \mu, \lambda), \\ x_c(z, \lambda) &= \arg \min_{x \in \mathfrak{R}^n} \{L_c(x, z, \mu, \lambda)\}.\end{aligned}$$

Then we have  $x_c(z, \lambda) = (\mathcal{P}_c f)(z - \lambda/c)$ . and

$$\theta_c(\mu, \lambda) = \min_z \varphi_c(z, \mu, \lambda),$$

where

$$(15) \quad \varphi_c(z, \mu, \lambda) = (e_c f)(z - \lambda/c) + g(z) - \frac{\|\lambda\|^2}{2c} + \mu^T h(z) + \frac{c}{2} \|h(z)\|^2.$$

Let us denote  $y = (\mu, \lambda)$  and  $\bar{y} = (\bar{\mu}, \bar{\lambda})$ , then we have the following result.

LEMMA 4. *Let Conditions (I) and (III) be satisfied. Then*

- (i) *For  $\bar{\lambda} = \nabla g(\bar{x}) + \mathcal{J}h(\bar{x})^T \bar{\mu}$ , ones has  $\nabla_z \varphi_c(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0$ .*
- (ii) *There exists  $\bar{c} \geq c_0$  such that every element in  $\partial_z^2 \varphi_c(\bar{x}, \bar{y})$  is positive definite when  $c \geq \bar{c}$ .*

*Proof.* From the inclusion  $0 \in f(\bar{x}) + \nabla g(\bar{x}) + \mathcal{J}h(\bar{x})^T \bar{\mu}$  or  $0 \in \partial f(\bar{x}) + \bar{\lambda}$ , we obtain that  $\bar{x} \in \arg \min [f(x) + \bar{\lambda}^T x]$ . Let  $\hat{f}(x) = f(x) + \bar{\lambda}^T x$ , then  $\bar{x} \in \arg \min [e_c \hat{f}](x)$ . Since

$$e_c \hat{f}(x) = \inf_z [\hat{f}(z) + \frac{c}{2} \|z - x\|^2] = e_c f(x - \bar{\lambda}/c) + \bar{\lambda}^T x - \frac{\|\bar{\lambda}\|^2}{2c},$$

one has that  $0 = \nabla e_c \hat{f}(\bar{x})$  or

$$\nabla e_c f(\bar{x} - \bar{\lambda}/c) + \bar{\lambda} = 0,$$

which implies  $\nabla_z \varphi_c(\bar{x}, \bar{\mu}, \bar{\lambda}) = 0$ . Nothing that

$$\partial_z^2 \varphi_c(z, y) = \partial(\nabla e_c f)(z - \lambda/c) + \nabla^2 g(z) + \sum_{i=1}^m (\mu_i + c h_i(z)) \nabla^2 h_i(z) + c \mathcal{J}h(z)^T \mathcal{J}h(z),$$

one has

$$\partial_z^2 \varphi_c(\bar{x}, \bar{y}) = \partial(\nabla e_c f)(\bar{x} - \bar{\lambda}/c) + \nabla^2 g(\bar{x}) + \sum_{i=1}^m \bar{\mu}_i \nabla^2 h_i(\bar{x}) + c \mathcal{J}h(\bar{x})^T \mathcal{J}h(\bar{x}).$$

Noting that each element  $W \in \partial_z^2 \varphi_c(\bar{x}, \bar{y})$  is self-adjoint and positively semi-definite, we have  $c_0 c^{-1} W \preceq W$  for  $c \geq c_0$  and that the conclusion (ii) comes from condition (III) by Debru theorem [6].  $\square$

**4. Augmented Lagrange method.** In this paper, we focus on Problem (1), in which  $f : \mathfrak{R}^n \rightarrow \bar{\mathfrak{R}}$  is a proper l.s.c. convex function,  $g : \mathfrak{R}^n \rightarrow \bar{\mathfrak{R}}$  is twice continuously differentiable function and  $h : \mathfrak{R}^n \rightarrow \mathfrak{R}^m$  is a twice continuously differentiable mapping around a feasible point  $\bar{x} \in \mathfrak{R}^n$ . By introducing an auxiliary variable  $z$ , we can express Problem (1) as (14).

The augmented Lagrange method can be described as follows: given  $c_0 > 0$ , choose  $y^0 = (\mu^0, \lambda^0) \in \mathfrak{R}^m \times \mathfrak{R}^n$ , as the initial multiplier. At the  $k$ -th iteration, solve

$$(16) \quad \min L_{c_k}(x, z, y^k)$$

to obtain  $(x^k, z^k)$ . Update  $y$  by

$$(17) \quad y^{k+1} = y^k + c_k \begin{bmatrix} h(z^k) \\ x^k - z^k \end{bmatrix}.$$

Update  $c$  by

$$c_{k+1} = c_k \text{ OR } c_{k+1} = \kappa c_k,$$

where  $\kappa > 1$  is a positive constant.

Note that  $L_c$  is convex in  $x$  and it is hopeful to be convex in  $z$  too when  $c$  is large enough. Especially, for many specific non-smooth functions  $f$ ,  $x_c(z, \mu, \lambda)$  can be easily obtained (even has an explicit expression) and we only need to minimize  $\varphi_c(z, \mu, \lambda)$  to obtain solution  $z^k$ . The equivalent expression (14) for the original problem (1) provides a proper way for using the augmented lagrangian method.

Define

$$z_c(y) = \arg \min\{\varphi_c(z, \mu, \lambda) : z \in \mathfrak{R}^n\},$$

where  $\varphi_c(z, \mu, \lambda)$  is defined by (15). Then  $z_c(y)$  satisfies the equality

$$(18) \quad \nabla(e_c f)(z_c(y) - \lambda/c) + \nabla g(z) + \mathcal{J}h(z_c(y))^T(\mu + h(z_c(y))) = 0$$

and  $\partial_z \varphi_c(z, y)$  has the expression

$$\partial_z \varphi_c(z, y) = \partial_z(\nabla e_c f)(z - \lambda/c) + \nabla^2 g(z) + \sum_{i=1}^m (\mu_i + ch_i(z)) \nabla^2 h_i(z) + c \mathcal{J}h(z)^T \mathcal{J}h(z).$$

For  $W \in \partial(\nabla e_c f)(z_c(y) - \lambda/c)$ , define

$$\mathcal{A}_c(y, W) = W + \nabla^2 g(z_c(y)) + \sum_{i=1}^m [\mu_i + ch_i(z_c(y))] \nabla^2 h_i(z_c(y)) + c \mathcal{J}h(z_c(y))^T \mathcal{J}h(z_c(y)).$$

In view of Lemma 4, we can easily obtain the following lemma:

LEMMA 5. *Let Conditions (I) and (III) be satisfied. Then there exists a positive constant  $\bar{c} \geq c_0$  such that the following assertions hold:*

- (i) *when  $c \geq \bar{c}$ ,  $\mathcal{A}_c(\bar{y}, W) \succ 0$ ;*
- (ii) *when  $\bar{c}$  is large enough, there exist  $\delta_1 > 0$  such that*

$$\mathcal{A}_c(y, W) \succeq \eta_1 I, \forall c \geq \bar{c}, y \in \mathbb{B}_{\delta_1}(\bar{y}).$$

The mapping  $z_c(y)$  is directionally differentiable with

$$(19) \quad [\nabla(e_c f)]'(z_c(y) - \lambda/c; z'_c(y; \Delta y) - \Delta \lambda/c) + [\nabla^2 g(z_c(y)) + c \mathcal{J}h(z_c(y))^T \mathcal{J}h(z_c(y)) + \sum_{i=1}^m (\mu_i + ch_i(z_c(y))) \nabla^2 h_i(z_c(y))] z'_c(y; \Delta y) + \mathcal{J}h(z_c(y))^T \Delta \mu = 0.$$

There exists an element  $W \in \partial_B(\nabla e_c f)(z_c(y) - \lambda/c)$  such that

$$[\nabla(e_c f)]'(z_c(y) - \lambda/c; z'_c(y; \Delta y) - \Delta \lambda/c) = W(z'_c(y; \Delta y) - \Delta \lambda/c).$$

Combing this with (19) we obtain

$$-\frac{1}{c} W \Delta \lambda + \mathcal{J}h(z_c(y))^T \Delta \mu + \mathcal{A}_c(y, w) z'_c(y; \Delta y) = 0,$$



yielding

$$(20) \quad z'_c(y; \Delta y) = \mathcal{A}_c(y, W)^{-1}[-\mathcal{J}h(z_c(y))^T \Delta \mu + \frac{1}{c}W \Delta \lambda].$$

Summarizing the above discussions, we obtain the following proposition.

**PROPOSITION 6.** *Suppose that Conditions (I) and (III) are satisfied and  $c \geq \bar{c}$ . Then there are two positive numbers  $\varepsilon_1$  and  $\delta_0$  (depending on  $c$ ) with  $\delta_1 > \delta_0 > 0$  and a locally Lipschitz continuous mapping  $z_c(y)$  defined on  $\mathbb{B}_{\delta_0}(\bar{y})$ , satisfying*

$$\{z_c(y)\} = \arg \min\{\varphi_c(z, y) : z \in \mathbb{B}_{\varepsilon_1}(\bar{x})\}.$$

And the following properties hold:

- (a) *The mapping  $z_c(\cdot)$  is semismooth at each point in  $\mathbb{B}_{\delta_0}(\bar{y})$ .*
- (b) *If  $\nabla^2 g, \nabla^2 h_i, i = 1, \dots, m$  are continuous around  $\bar{x}$  and  $(\nabla e_c f)(\cdot)$  is strongly semismooth everywhere, then  $z_c(\cdot)$  is strongly semismooth in  $\mathbb{B}_{\delta_0}(\bar{y})$ .*
- (c) *For each  $z \in \mathbb{B}_{\varepsilon}(\bar{x})$  and  $y \in \mathbb{B}_{\delta_0}(\bar{y})$ , every element of  $\Pi_z \partial_B(\nabla_z \varphi_c(z, y))$  is positive definite.*
- (d) *For each  $y \in \mathbb{B}_{\delta_0}(\bar{y})$ ,  $z_c(y)$  is the unique solution to*

$$\min \varphi_c(z, y) \quad \text{s.t.} \quad z \in \mathbb{B}_{\varepsilon_1}(\bar{x}).$$

Define

$$(21) \quad \theta_c : \mathfrak{R}^m \times \mathfrak{R}^n \rightarrow \mathfrak{R} \quad \text{by} \quad \theta_c(y) = \min_{z \in \mathbb{B}_{\varepsilon_1}(\bar{x})} \varphi_c(z, y).$$

Obviously  $\theta_c(y)$  is concave and for each  $y \in \mathbb{B}_{\delta_0}(\bar{y})$ ,

$$\vartheta_c(y) = \varphi_c(z_c(y), y).$$

For any  $y \in \mathbb{B}_{\delta_0}(\bar{y})$ ,  $y = (\mu, \lambda) \in \mathfrak{R}^m \times \mathfrak{R}^n$ , let  $x_c(y) = (P_c f)(z_c(y) - \lambda/c)$ ,

$$\begin{pmatrix} \mu_c(y) \\ \lambda_c(y) \end{pmatrix} = \begin{pmatrix} \mu + ch(z_c(y)) \\ \lambda + c[x_c(y) - z_c(y)] \end{pmatrix}.$$

**PROPOSITION 7.** *Suppose Conditions (I) and (III) hold and  $c \geq \bar{c}$ . Then  $\vartheta_c(y)$  defined by (21), is continuously differentiable over  $\mathbb{B}_{\delta_0}(\bar{y})$ , and*

$$(22) \quad \nabla \vartheta_c(y) = \begin{bmatrix} h(z_c(y)) \\ -\frac{1}{c} \nabla e_c f(z_c(y) - \lambda/c) - \frac{\lambda}{c} \end{bmatrix}.$$

*Proof.* Let  $y = (\mu, \lambda) \in \mathbb{B}_{\delta_0}(\bar{y})$ . From the definition of  $\vartheta_c(y)$ , we have from Theorem 2.6.6 of [3], for any  $\Delta y = (\Delta \mu, \Delta \lambda) \in \mathfrak{R}^m \times \mathfrak{R}^n$ , that

$$\begin{aligned} \partial \vartheta(y)(\Delta y) &= D_z \varphi_c(z_c(y), y) z'_c(y; \Delta y) + \langle \nabla e_c f(z_c(y) - \lambda/c), -\Delta \lambda/c \rangle \\ &\quad - \frac{1}{c} \langle \lambda, \Delta \lambda \rangle + h(z_c(y))^T \Delta \mu, \end{aligned}$$

which implies from  $D_z \varphi_c(z_c(y), y) = 0$  that

$$\partial \vartheta(y)(\Delta y) = \langle h(z_c(y)), \Delta \mu \rangle - \frac{1}{c} \langle \nabla e_c f(z_c(y) - \lambda/c) + \lambda, \Delta \lambda \rangle.$$

Therefore,  $\vartheta_c(\cdot)$  is differentiable of  $y \in \mathbb{B}_{\delta_0}(y)$  and  $\nabla \vartheta_c(y)$  has the expression (22).  $\square$

Define

$$(23) \quad \bar{\mathcal{V}}_c(\Delta y) = \left\{ - \begin{pmatrix} \mathcal{J}h(\bar{x}) \\ -W/c \end{pmatrix} \mathcal{A}_c(\bar{y}, W)^{-1} [\mathcal{J}h(\bar{x})^T - W/c] \begin{bmatrix} \Delta\mu \\ \Delta\lambda \end{bmatrix} \right. \\ \left. + \begin{bmatrix} 0 \\ (1/c^2 W - 1/cI) \Delta\lambda \end{bmatrix} : W \in \partial_B(\nabla e_c f)(\bar{x} - \bar{\lambda}/c) \right\}.$$

PROPOSITION 8. *Suppose that Conditions (I) and (III) are satisfied. Let  $c \geq \bar{c}$ . Then for any  $\Delta y := (\Delta\mu, \Delta\lambda) \in \mathfrak{R}^m \times \mathfrak{R}^n$ ,*

$$(24) \quad \partial_B(\nabla\vartheta_c)(\bar{y})(\Delta y) \subseteq \bar{\mathcal{V}}_c(\Delta y).$$

*Proof.* Let  $\Delta y = (\Delta\mu, \Delta\lambda) \in \mathfrak{R}^m \times \mathfrak{R}^n$  be arbitrary. From Proposition 7, we know that  $\nabla\vartheta_c(\cdot)$  is semismooth at any point  $y \in \mathbb{B}_{\delta_0}(\bar{y})$ . Let  $\mathcal{D}_{\nabla\vartheta_c}$  denote the set of all Fréchet-differentiable points of  $\nabla\vartheta_c(\cdot)$  in  $\mathbb{B}_{\delta_0}(\bar{y})$ . Then, for any  $y = (\lambda, \mu) \in \mathcal{D}_{\nabla\vartheta_c}$ , we have

$$(25) \quad \nabla^2\vartheta_c(y)(\Delta y) = \begin{bmatrix} \mathcal{J}h(z_c(y))z'_c(y; \Delta y) \\ -\frac{1}{c}[\nabla e_c f]'(z_c(y) - \lambda/c; z'_c(y; \Delta y) - \Delta\lambda/c) - \frac{\Delta\lambda}{c} \end{bmatrix}.$$

Differentiating (18) with respect to  $y$ , one has

$$0 = [\nabla e_c f]'(z_c(y) - \lambda/c; z'_c(y; \Delta y) - \Delta\lambda/c) + \mathcal{J}h(z_c(y))^T \Delta\mu \\ + \left[ \nabla^2 g(z_c(y)) + \sum_{i=1}^m (\mu_i + ch_i(z_c(y))) \nabla^2 h_i(z_c(y)) + c \mathcal{J}h(z_c(y))^T \mathcal{J}h(z_c(y)) \right] z'_c(y; \Delta y).$$

There exists an element  $W \in \partial_B(\nabla e_c f)(z_c(y) - \lambda/c)$  such that

$$[\nabla e_c f]'(z_c(y) - \lambda/c; z'_c(y; \Delta y) - \Delta\lambda/c) = W(z'_c(y; \Delta y) - \Delta\lambda/c)$$

so that

$$z'_c(y; \Delta y) = \mathcal{A}(y, W)^{-1} [-\mathcal{J}h(z_c(y))^T \Delta\mu + \frac{1}{c} W \Delta\lambda].$$

Therefore, we obtain from (25) that

$$\nabla^2\vartheta_c(y)(\Delta y) \\ = \begin{bmatrix} \mathcal{J}h(z_c(y))z'_c(y; \Delta y) \\ -\frac{1}{c}W(z'_c(y; \Delta y) - \frac{\Delta\lambda}{c}) - \frac{\Delta\lambda}{c} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{c} \left[ \frac{1}{c}W - I \right] \Delta\lambda \end{bmatrix} \\ + \begin{bmatrix} \mathcal{J}h(z_c(y)) \\ -\frac{1}{c}W \end{bmatrix} \mathcal{A}(y, W)^{-1} \begin{bmatrix} -\mathcal{J}h(z_c(y)) & W \\ & c \end{bmatrix} \begin{bmatrix} \Delta\mu \\ \Delta\lambda \end{bmatrix}.$$

Therefore we obtain (24) from the definition of B-subdifferential.  $\square$

For  $W \in \partial(\nabla e_c f)(z_c(y) - \lambda/c)$ , let us define

$$\mathcal{B}_{c',c}(y, W) = c'c^{-1}W + \nabla^2 g(z_c(y)) + \sum_{i=1}^m [\mu_i + ch_i(z_c(y))] \nabla^2 h_i(z_c(y)) + c' \mathcal{J}h(z_c(y))^T \mathcal{J}h(z_c(y))$$

for  $c_0 \leq c' < c$ . Under Conditions (I) and (III), noting that  $c^{-1}W \in I - \partial P_{c,f}(z_c(y) - \lambda/c)$  for  $W \in \partial(\nabla e_{c,f})(z_c(y) - \lambda/c)$ , we have from Lemma 5 that we may choose  $c_0 > 0$  to satisfy

$$(26) \quad \underline{\eta} \|d\|^2 \leq \langle d, \mathcal{B}_{c_0,c}(y, W)d \rangle \leq \bar{\eta} \|d\|^2, \forall W \in \partial_B(\nabla e_{c,f})(z_c(y) - \lambda/c), \forall y \in \mathbb{B}_{\delta_1}(\bar{y}).$$

where  $\bar{\eta} > \underline{\eta} > 0$  are constants. Then we rewrite  $\mathcal{A}_c(\bar{y}, W)$  as

$$(27) \quad \mathcal{A}_c(\bar{y}, W) = \mathcal{B}_{c_0,c}(\bar{y}, W) + \hat{c}c^{-1}W + \hat{c}A^T A$$

where  $A = \mathcal{J}h(\bar{x})$ , which is assumed to have the SVD as follows

$$A = U[\Sigma \quad 0]R^T.$$

For the purpose of estimating the norm of  $z'_c(\bar{y}; \Delta y)$ , we establish several lemmas as follows.

LEMMA 9. Let  $n_1 = n - m$  and

$$(28) \quad \mathcal{G}_{c_0,c}(\bar{y}, W) = \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I_{n_1} \end{bmatrix} R^T \mathcal{B}_{c_0,c}(\bar{y}, W) R \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I_{n_1} \end{bmatrix}.$$

Then for  $\forall y \in \mathbb{B}_{\delta_1}(\bar{y})$  and  $\forall W \in \partial(\nabla e_{c,f})(z_c(\bar{y}) - \lambda/c)$ ,

$$\mathcal{G}_{c_0,c}(\bar{y}, W) \succeq \underline{\eta} \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I_{n_1} \end{bmatrix} \succeq \underline{\sigma} \underline{\eta} I_n$$

and

$$\mathcal{G}_{c_0,c}(\bar{y}, W) \preceq \bar{\eta} \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I_{n_1} \end{bmatrix} \preceq \bar{\sigma} \bar{\eta} I_n,$$

where  $\underline{\sigma} = \min\{1, \min[\Sigma_{ii}^{-1}, i = 1, \dots, m]\}$  and  $\bar{\sigma} = \max\{1, \max[\Sigma_{ii}^{-1}, i = 1, \dots, m]\}$ .

LEMMA 10. Let

$$(29) \quad \mathcal{H}_{c_0,c}(\bar{y}, W) = \mathcal{G}_{c_0,c}(\bar{y}, W)^{-1} = \begin{bmatrix} H_1(W) & H_2(W)^T \\ H_2(W) & H_3(W) \end{bmatrix}.$$

Then

$$\|H_1(W)\|_2 \leq (\underline{\sigma} \underline{\eta})^{-1}, \|H_1(W)^{-1}\|_2 \leq \bar{\sigma} \bar{\eta}, \|H_2(W)H_1(W)^{-1}\|_2 \leq (\underline{\sigma} \underline{\eta})^{-1} \bar{\sigma} \bar{\eta},$$

where the matrix norm  $\|\cdot\|_2$  is the spectral norm.

LEMMA 11. With the notations in (29), we have

$$\begin{aligned} & \left( (\mathcal{G}(\bar{y}, W) + \hat{c} \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}) \right)^{-1} = \mathcal{H}_{c_0}(\bar{y}, W) - \\ & - \hat{c} \begin{bmatrix} H_1(W)(I_m + \hat{c}H_1(W))^{-1}H_1(W) & H_1(W)(I_m + \hat{c}H_1(W))^{-1}H_2(W) \\ H_2(W)^T(I_m + \hat{c}H_1(W))^{-1}H_1(W) & H_2(W)^T(I_m + \hat{c}H_1(W))^{-1}H_2(W) \end{bmatrix} \end{aligned}$$

*Proof.* It comes from the Sherman-Morrison-Woodbury formula (cf. Section 2.1 of [7]).  $\square$

Now we estimate the norm of  $z'_c(\bar{y}; \Delta y)$ .

PROPOSITION 12. *Suppose that Conditions (I) and (III) are satisfied. Then there exists  $\rho_0 > 0$  such that*

$$(30) \quad \|z'_c(\bar{y}; \Delta y)\| \leq \rho_0 \|\Delta y\|/c, \quad c \geq 2\bar{c}.$$

*Proof.* From the expression

$$z'_c(\bar{y}; \Delta y) = \mathcal{A}_c(\bar{y}, W)^{-1}[-A^T \Delta \mu + \frac{1}{c} W \Delta \lambda],$$

we obtain

$$(31) \quad \begin{aligned} \langle z'_c(\bar{y}; \Delta y), z'_c(\bar{y}; \Delta y) \rangle &\leq 2 \langle A^T \Delta \mu, \mathcal{A}_c(\bar{y}, W)^{-2} A^T \Delta \mu \rangle \\ &\quad + \frac{2}{c^2} \langle W \Delta \lambda, \mathcal{A}_c(\bar{y}, W)^{-2} W \Delta \lambda \rangle. \end{aligned}$$

From the expression

$$\mathcal{A}_c(\bar{y}, W) = \mathcal{B}_{c_0, c}(\bar{y}, W) + \hat{c}c^{-1}W + \hat{c}A^T A,$$

we obtain that

$$\begin{aligned} &\langle A^T \Delta \mu, \mathcal{A}_c(\bar{y}, W)^{-2} A^T \Delta \mu \rangle \\ &\leq \langle A^T \Delta \mu, [\mathcal{B}_{c_0, c}(\bar{y}, W) + \hat{c}A^T A]^{-2} A^T \Delta \mu \rangle \\ &= \|[\mathcal{B}_{c_0, c}(\bar{y}, W) + \hat{c}A^T A]^{-1} A^T \Delta \mu\|^2 \\ &= \left\| [\mathcal{B}_{c_0, c}(\bar{y}, W) + \hat{c}R \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} U^T U [\Sigma \quad 0] R^T]^{-1} R \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} U^T \Delta \mu \right\|^2 \\ &= \left\| R \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I_{n_1} \end{bmatrix} \begin{bmatrix} (H_1(W)^{-1} + \hat{c}I_m)^{-1} \\ H_2(W)^T H_1(W)^{-1} (H_1(W)^{-1} + \hat{c}I_m)^{-1} \end{bmatrix} U^T \Delta \mu \right\|^2 \\ &= \left\| \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & I_{n_1} \end{bmatrix} \right\|^2 \left\| \begin{bmatrix} I_m \\ H_2(W)^T H_1(W)^{-1} \end{bmatrix} \right\|^2 \|(H_1(W)^{-1} + \hat{c}I_m)^{-1}\|^2 \|\Delta \mu\|^2. \end{aligned}$$

Then, by defining

$$\sigma_0 = \max\{\max_{1 \leq i \leq m} \Sigma_{ii}^{-2}, 1\} [1 + (\underline{\sigma}\eta)^{-1} \overline{\sigma}\eta]^2,$$

we obtain

$$(32) \quad \begin{aligned} &\langle A^T \Delta \mu, \mathcal{A}_c(\bar{y}, W)^{-2} A^T \Delta \mu \rangle \\ &\leq \sigma_0 \|\hat{c}^{-1} I_m - \hat{c}^{-1} (\hat{c} I_{n_2} + H_1(W))^{-1} H_1(W)\|^2 \|\Delta \mu\|^2 \\ &\leq \sigma_0 [\hat{c}^{-1} \|\Delta \mu\| (\hat{c}^{-1} \|\hat{c} I_{n_2} + H_1^{-1}(W))^{-1}\|_\alpha \|H_1^{-1}(W)\|_2 \|\Delta \mu\|^2] \\ &\leq \sigma_0 \hat{c}^{-2} (1 + \|H_1(W)\|_2 \|H_1^{-1}(W)\|_2)^2 \|\Delta \mu\|^2 \\ &\leq \sigma_1 \hat{c}^{-2} \|\Delta \mu\|^2, \end{aligned}$$

where  $\sigma_1 = \sigma_0 (\underline{\sigma}\eta)^{-1} \overline{\sigma}\eta)^2$ .

Since  $\mathcal{B}_{c_0,c}(\bar{y}, W)$  satisfies (26), we can easily obtain

$$\begin{aligned}
& \langle W\Delta\lambda, \mathcal{A}_c(\bar{y}, W)^{-2}W\Delta\lambda \rangle \\
& \leq \langle W\Delta\lambda, [\mathcal{B}_{c_0,c}(\bar{y}, W) + \widehat{c}c^{-1}W]^{-2}W\Delta\lambda \rangle \\
& = \|[\mathcal{B}_{c_0,c}(\bar{y}, W) + \widehat{c}c^{-1}W]^{-1}W\Delta\lambda\|^2 \\
(33) \quad & = \|\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}(I + \widehat{c}c^{-1}\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2})^{-1} \times \\
& \quad \times \mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}\mathcal{B}_{c_0,c}(\bar{y}, W)^{1/2}\Delta\lambda\|^2 \\
& \leq \underline{\eta}^{-1}\bar{\eta}\|(I + \widehat{c}c^{-1}\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2})^{-1} \times \\
& \quad \times \mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}\|^2\|\Delta\lambda\|^2.
\end{aligned}$$

Noting that  $c^{-1}W$  is positively semi-definite with  $\|c^{-1}W\| \leq 1$  and  $\mathcal{B}_{c_0,c}(\bar{y}, W)$  satisfies (26), we have that the eigenvalues of  $c^{-1}\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}$  are all in  $[0, \underline{\eta}^{-1/2}\bar{\eta}^{1/2}]$ . Assume that  $c^{-1}\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}$  has the following spectral decomposition

$$c^{-1}\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2} = Q_W \begin{bmatrix} \Lambda(W)_\alpha & 0 \\ 0 & 0_\beta \end{bmatrix} Q_W^T,$$

where  $\alpha = \{i : \lambda_i(c^{-1}\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}) > 0\}$  and  $\beta = \{1, \dots, n\} \setminus \alpha$ , and we use  $\lambda(W)_i$  denote the  $i$ -th eigenvalue of  $c^{-1}\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}$ . Since

$$\begin{aligned}
& (I + \widehat{c}c^{-1}\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2})^{-1}\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2} \\
& = Q_W \begin{bmatrix} (I_\alpha + \widehat{c}\Lambda(W)_\alpha)^{-1} & 0 \\ 0 & I_\beta \end{bmatrix} \begin{bmatrix} c\Lambda(W)_\alpha & 0 \\ 0 & 0_\beta \end{bmatrix} Q_W^T \\
& = Q_W \begin{bmatrix} c(I_\alpha + \widehat{c}\Lambda(W)_\alpha)^{-1}\Lambda(W)_\alpha & 0 \\ 0 & 0_\beta \end{bmatrix} Q_W^T,
\end{aligned}$$

we obtain

$$\begin{aligned}
(34) \quad & \|(I + \widehat{c}c^{-1}\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2})^{-1}\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}W\mathcal{B}_{c_0,c}(\bar{y}, W)^{-1/2}\| \\
& \leq \max_{i \in \alpha} \frac{c\lambda(W)_i}{1 + \widehat{c}\lambda(W)_i} < \frac{c}{\widehat{c}} \leq 2
\end{aligned}$$

when  $c \geq 2\bar{c} \geq 2c_0$ . Combining (33) and (34), we obtain for  $c \geq 2\bar{c}$  that

$$(35) \quad \langle W\Delta\lambda, \mathcal{A}_c(\bar{y}, W)^{-2}W\Delta\lambda \rangle \leq 2\underline{\eta}^{-1}\bar{\eta}\|\Delta\lambda\|^2$$

Let  $\sigma_2 = \underline{\eta}^{-1}\bar{\eta}$  and  $\rho_0 = 2\sqrt{\max(\sigma_1, \sigma_2)}$ . Then for  $c \geq 2\bar{c}$ , from (31), (32) and (35), we have

$$\begin{aligned}
\|z'_c(\bar{y}; \Delta y)\|^2 & \leq 2\sigma_1\widehat{c}^{-2}\|\Delta\mu\|^2 + 4\sigma_2c^{-2}\|\Delta\lambda\|^2 \\
& \leq \frac{4\sigma_1}{c^2}\|\Delta\mu\|^2 + \frac{4\sigma_2}{c^2}\|\Delta\lambda\|^2 \leq \frac{\rho_0^2}{c^2}\|\Delta y\|^2
\end{aligned}$$

The proof is completed.  $\square$

PROPOSITION 13. *Under Condition(I) and Condition(III), there exists  $\mu_0 > 0$  such that for  $c \geq 2\bar{c}$ ,*

$$(36) \quad \langle V(\Delta y) + c^{-1}\Delta y, \Delta y \rangle \in \mu_0c^{-2}[-1, 1]\|\Delta y\|^2, \forall V(\Delta y) \in \bar{\mathcal{V}}_c(\Delta y).$$

*Proof.* For  $\forall V(\Delta y) \in \bar{\mathcal{V}}_c(\Delta y)$ , there exists an element  $W \in \partial_B(\nabla ecf)(\bar{x} - \bar{\lambda}/c)$  such that

$$V(\Delta y) = - \begin{bmatrix} \mathcal{J}h(\bar{x}) \\ -W/c \end{bmatrix} \mathcal{A}_c(\bar{y}, W)^{-1} [\mathcal{J}h(\bar{x})^T - W/c] \Delta y + \begin{bmatrix} 0 \\ (\frac{1}{c^2}W - \frac{1}{c}I)\Delta\lambda \end{bmatrix}.$$

So we have

$$V(\Delta y) + \frac{\Delta y}{c} = - \begin{bmatrix} \mathcal{J}h(\bar{x}) \\ -W/c \end{bmatrix} \mathcal{A}_c(\bar{y}, W)^{-1} [\mathcal{J}h(\bar{x})^T - W/c] \begin{bmatrix} \Delta\mu \\ \Delta\lambda \end{bmatrix} + \begin{bmatrix} \frac{\Delta\mu}{c} \\ \frac{W}{c^2}\Delta\lambda \end{bmatrix}.$$

Noting  $\mathcal{A}_c(\bar{y}, W) = \mathcal{B}_{\bar{c},c}(\bar{y}, W) + \tilde{c}c^{-1}W + \tilde{c}A^T A$  with  $A = \mathcal{J}h(\bar{x})$  and  $\tilde{c} = c - \bar{c}$ , we have

$$(37) \quad \langle V(\Delta y) + \frac{\Delta y}{c}, \Delta y \rangle = \left\langle \begin{pmatrix} \Delta\mu \\ c^{-1/2}W^{1/2}\Delta\lambda \end{pmatrix}, G_c(\bar{y}, W) \begin{pmatrix} \Delta\mu \\ c^{-1/2}W^{1/2}\Delta\lambda \end{pmatrix} \right\rangle$$

where

$$G_c(\bar{y}, W) = c^{-1}I - \hat{A}_W[\mathcal{B}_{\bar{c},\tilde{c}}(\bar{y}, W) + \tilde{c}\hat{A}_W^T \hat{A}_W]^{-1} \hat{A}_W^T$$

and

$$\hat{A}_W = \begin{pmatrix} A \\ -c^{-1/2}W^{1/2} \end{pmatrix}.$$

Since  $\hat{A}_W \in \mathfrak{R}^{(m+n) \times n}$ , it is impossible be of full row rank. Let  $r_W = \text{rank } \hat{A}_W$ , then  $r_W \geq m$ . Without loss of generality, we assume that the first  $r_W - m$  columns of  $-c^{-1/2}W^{1/2}$  are linearly independent and

$$\text{range}[A^T \quad -c^{-1/2}W^{1/2}] = \text{range}[A^T \quad E_W],$$

where  $E_W$  denotes the matrix generated by the first  $r_W - m$  columns of  $-c^{-1/2}W^{1/2}$ . Then  $-c^{-1/2}W^{1/2}$  is expressed as

$$-c^{-1/2}W^{1/2} = [E_W \quad (A^T \quad E_W)\Theta_W]$$

for some matrix  $\Theta_W \in \mathfrak{R}^{r_W \times (n+m-r_W)}$ , and  $\hat{A}_W^T$  is expressed as

$$\hat{A}_W^T = (A^T \quad E_W)[I_{r_W} \quad \Theta_W].$$

With these notations, we have

$$(38) \quad \begin{aligned} G_c(\bar{y}, W) &= c^{-1}I - [I_{r_W} \quad \Theta_W]^T (A^T \quad E_W)^T [\mathcal{B}_{\bar{c},\tilde{c}}(\bar{y}, W) + \\ &\quad + \tilde{c}(A^T \quad E_W)(I_{r_W} + \Theta_W \Theta_W^T)(A^T \quad E_W)^T]^{-1} (A^T \quad E_W)[I_{r_W} \quad \Theta_W] \\ &= c^{-1}I - [I_{r_W} \quad \Theta_W]^T (I_{r_W} + \Theta_W \Theta_W^T)^{-1/2} (I_{r_W} + \Theta_W \Theta_W^T)^{1/2} (A^T \quad E_W)^T \times \\ &\quad \times [\mathcal{B}_{\bar{c},\tilde{c}}(\bar{y}, W) + \tilde{c}(A^T \quad E_W)(I_{r_W} + \Theta_W \Theta_W^T)(A^T \quad E_W)^T]^{-1} \times \\ &\quad \times (A^T \quad E_W)(I_{r_W} + \Theta_W \Theta_W^T)^{1/2} (I_{r_W} + \Theta_W \Theta_W^T)^{-1/2} [I_{r_W} \quad \Theta_W] \\ &= c^{-1}I - F_W^T \tilde{A}_W [\mathcal{B}_{\bar{c},\tilde{c}}(\bar{y}, W) + \tilde{c}\tilde{A}_W^T \tilde{A}_W]^{-1} \tilde{A}_W^T F_W, \end{aligned}$$

where

$$F_W = (I_{r_W} + \Theta_W \Theta_W^T)^{-1/2} [I_{r_W} \quad \Theta_W] \text{ and } \tilde{A}_W = (I_{r_W} + \Theta_W \Theta_W^T)^{1/2} (A^T \quad E_W)^T.$$

For the singular value decomposition  $\tilde{A}_W = U_W[\Sigma_W \ 0]R_W^T$ , define

$$\tilde{\mathcal{G}}_{\tilde{c},c}(\bar{y}, W) = \begin{bmatrix} \Sigma_W^{-1} & 0 \\ 0 & I_{n_1(W)} \end{bmatrix} R_W^T \mathcal{B}_{\tilde{c},c}(\bar{y}, W) R_W \begin{bmatrix} \Sigma_W^{-1} & 0 \\ 0 & I_{n_1(W)} \end{bmatrix},$$

where  $n_1(W) = n - r_W$ . Let

$$\tilde{\mathcal{H}}_{\tilde{c},c}(\bar{y}, W) = \tilde{\mathcal{G}}_{\tilde{c},c}(\bar{y}, W)^{-1} = \begin{bmatrix} \tilde{H}_1(W) & \tilde{H}_2(W)^T \\ \tilde{H}_2(W) & \tilde{H}_3(W) \end{bmatrix}$$

and

$$\Delta y_W = F_W \begin{pmatrix} \Delta \mu \\ c^{-1/2} W^{1/2} \Delta \lambda \end{pmatrix}.$$

Then we obtain

$$\begin{aligned} & \Delta y_W^T \tilde{A}_W [\mathcal{B}_{\tilde{c},\tilde{c}}(\bar{y}, W) + \tilde{c} \tilde{A}_W^T \tilde{A}_W]^{-1} \tilde{A}_W^T \Delta y_W \\ &= \langle \Delta y_W, U_W [I_{r_W} \ 0] \left( \mathcal{G}_{\tilde{c},c}(\bar{y}, W) + \tilde{c} \begin{bmatrix} I_{r_W} & 0 \\ 0 & 0 \end{bmatrix} \right)^{-1} \begin{pmatrix} I_{r_W} \\ 0 \end{pmatrix} U_W^T \Delta y_W \rangle \\ &= \langle U_W^T \Delta y_W, (\tilde{H}_1(W) - \tilde{H}_1(W)(I_{r_W} + \tilde{c} \tilde{H}_1(W))^{-1} \tilde{H}_1(W)) U_W^T \Delta y_W \rangle \\ &= \langle U_W^T \Delta y_W, (\tilde{H}_1(W)^{-1} + \tilde{c} I_{r_W})^{-1} U_W^T \Delta y_W \rangle \\ &= \langle U_W^T \Delta y_W, (\tilde{c}^{-1} I_{r_W} - \tilde{c}^{-1} \tilde{H}_1(W)^{-\frac{1}{2}} (\tilde{c} I_{r_W} + \tilde{H}_1(W)^{-1})^{-1} \tilde{H}_1(W)^{-\frac{1}{2}}) U_W^T \Delta y_W \rangle \\ &= \tilde{c}^{-1} \|\Delta y_W\|^2 - \langle U_W^T \Delta y_W, \tilde{c}^{-1} \tilde{H}_1(W)^{-\frac{1}{2}} (\tilde{c} I_{r_W} + \tilde{H}_1(W)^{-1}) \tilde{H}_1(W)^{-\frac{1}{2}} U_W^T \Delta y_W \rangle \\ &= \tilde{c}^{-1} \|\Delta y_W\|^2 - \langle \tilde{H}_1(W)^{-\frac{1}{2}} U_W^T \Delta y_W, \tilde{c}^{-1} (\tilde{c} I_{r_W} + \tilde{H}_1(W)^{-1}) \tilde{H}_1(W)^{-\frac{1}{2}} U_W^T \Delta y_W \rangle \end{aligned}$$

Suppose  $\tilde{H}_1^{-1}(W)$  has the following spectral decomposition

$$\tilde{H}_1(W)^{-1} = \mathcal{Q} \text{diag}(\xi_1, \dots, \xi_{r_W}) \mathcal{Q}^T, \Xi_1 = \text{diag}(\xi_1, \dots, \xi_{r_W}),$$

then we obtain

$$\begin{aligned} & \langle \tilde{H}_1(W)^{-1/2} U_W^T \Delta y_W, \tilde{c}^{-1} (\tilde{c} I_{r_W} + \tilde{H}_1(W))^{-1} \tilde{H}_1(W)^{-\frac{1}{2}} U_W^T \Delta y_W \rangle \\ &= \langle \Xi_1^{\frac{1}{2}} \mathcal{Q}^T U_W^T \Delta y_W, \tilde{c}^{-1} (\tilde{c} I_{r_W} + \Xi_1)^{-1} \Xi_1^{\frac{1}{2}} \mathcal{Q}^T U_W^T \Delta y_W \rangle \\ (39) \quad &= \langle \mathcal{Q}^T U_W^T \Delta y_W, \tilde{c}^{-1} \Xi_1 (\tilde{c} I_{r_W} + \Xi_1)^{-1} \Xi_1^{\frac{1}{2}} \mathcal{Q}^T U_W^T \Delta y_W \rangle \\ &= \langle \mathcal{Q}^T U_W^T \Delta y_W, \tilde{c}^{-1} \text{diag} \left( \frac{\xi_1}{\tilde{c} + \xi_1}, \dots, \frac{\xi_{r_W}}{\tilde{c} + \xi_{r_W}} \right) \mathcal{Q}^T U_W^T \Delta y_W \rangle \\ &\in \tilde{c}^{-1} \left[ \frac{\xi_1}{\tilde{c} + \xi_1}, \frac{\xi_{r_W}}{\tilde{c} + \xi_{r_W}} \right] \|\Delta y_W\|^2, \end{aligned}$$

which implies

$$\begin{aligned} & \Delta y_W^T \tilde{A}_W [\mathcal{B}_{\tilde{c},\tilde{c}}(\bar{y}, W) + \tilde{c} \tilde{A}_W^T \tilde{A}_W]^{-1} \tilde{A}_W^T \Delta y_W \\ (40) \quad & \in \left( \frac{1}{\tilde{c}} + \frac{1}{\tilde{c}} \left[ -\frac{\xi_{r_W}}{\tilde{c} + \xi_{r_W}}, -\frac{\xi_1}{\tilde{c} + \xi_1} \right] \right) \|\Delta y_W\|^2 \\ &= \left[ \frac{1}{\tilde{c} + \xi_{r_W}}, \frac{1}{\tilde{c} + \xi_1} \right] \|\Delta y_W\|^2. \end{aligned}$$

Now we estimate  $\|\Delta y_W\|$ . Obviously we have that  $F_W$  can be expressed as

$$(41) \quad F_W = (I_{r_W} \ \Theta_W)(I_{r_W} \ \Theta_W)^T)^{-1/2}(I_{r_W} \ \Theta_W)$$

where  $(I_{r_W} \ \Theta_W)$  is a matrix in full row rank. Assume that  $(I_{r_W} \ \Theta_W)$  has the following SVD:

$$(I_{r_W} \ \Theta_W) = \widehat{U}_W[\widehat{\Xi}_W \ 0]\widehat{R}_W^T$$

then we have from (41) that  $F_W$  can be expressed as

$$(42) \quad F_W = \widehat{U}_W[I_{r_W} \ 0]\widehat{R}_W^T.$$

Therefore, we have from the equality (42) that (the spectral norm)  $\|F_W\|_2 = 1$  and in turn,

$$(43) \quad \|\Delta y_W\|^2 \leq \|\Delta\mu\|^2 + \|c^{-1/2}W^{1/2}\Delta\lambda\|^2$$

Therefore, for  $c \geq 2\bar{c}$ , we have from (37), (40) and (43) that

$$(44) \quad \langle V(\Delta y) + \frac{\Delta y}{c}, \Delta y \rangle \in \left[ \frac{1}{c} - \frac{1}{\bar{c} + \xi_1}, \frac{1}{c} - \frac{1}{\bar{c} + \xi_{r_W}} \right] [\|\Delta\mu\|^2 + \|c^{-1/2}W^{1/2}\Delta\lambda\|^2].$$

Since

$$(A^T \quad -c^{-1/2}W^{1/2}) = (A^T \quad E_W)(I_{r_W} \ \Theta_W)$$

we have

$$(A^T A + c^{-1}W) = (A^T \quad E_W)(I_{r_W} + \Theta_W\Theta_W^T)(A^T \quad E_W)^T$$

and so that

$$\begin{aligned} & (I_{r_W} + \Theta_W\Theta_W^T)^{1/2} \\ &= [(A^T A + E_W E_W^T)^{-1}(A^T \quad E_W)^T(A^T A + c^{-1}W)(A^T \quad E_W)(A^T A + E_W E_W^T)^{-1}]^{1/2}. \end{aligned}$$

Therefore the matrix  $\tilde{A}_W = (I_{r_W} + \Theta_W\Theta_W^T)^{1/2}(A^T \quad E_W)^T$  is a uniform bounded matrix with respect to  $W$  and the singular value matrix  $\Sigma_W$  is uniformly bounded with respect to  $W$ . From the expression

$$\tilde{\mathcal{H}}_{\bar{c},c}(\bar{y}, W) = \begin{bmatrix} \Sigma_W & 0 \\ 0 & I_{n_1(W)} \end{bmatrix} R_W^T \mathcal{B}_{\bar{c},c}(\bar{y}, W)^{-1} R_W \begin{bmatrix} \Sigma_W & 0 \\ 0 & I_{n_1(W)} \end{bmatrix},$$

and noting that  $\mathcal{B}_{c_0,c}(\bar{y}, W)$  satisfies (26), we have that  $\xi_1$  and  $\xi_{r_W}$  in (44) satisfy

$$0 \leq \xi_1 \leq \xi_{r_W} \leq \bar{\eta}_0$$

for some constant  $\bar{\eta}_0 > 0$ . Obviously we have  $\|c^{-1/2}W^{1/2}\Delta\lambda\| \leq \|\Delta\lambda\|$  and in turn,

$$(45) \quad \|\Delta\mu\|^2 + \|c^{-1/2}W^{1/2}\Delta\lambda\|^2 \leq \|\Delta y\|^2.$$

Therefore, we have from (45), (44) and the fact  $0 \leq \xi_1 \leq \xi_{r_W} \leq \bar{\eta}_0$ , that

$$(46) \quad \begin{aligned} \langle V(\Delta y) + \frac{\Delta y}{c}, \Delta y \rangle &\in \left[ \frac{1}{c} - \frac{1}{\bar{c} + \xi_1}, \frac{1}{c} - \frac{1}{\bar{c} + \xi_{r_W}} \right] \|\Delta y\|^2 \\ &\subseteq \left[ \frac{1}{c} - \frac{1}{\bar{c}}, \frac{1}{c} - \frac{1}{\bar{c} + \bar{\eta}_0} \right] \|\Delta y\|^2 \\ &= \left[ \frac{-\bar{c}}{c\bar{c}}, \frac{\bar{\eta}_0 - \bar{c}}{c(\bar{c} + \bar{\eta}_0)} \right] \|\Delta y\|^2, \end{aligned}$$

which implies that there exists  $\mu_0 > 0$  such that (36) is satisfied when for  $c \geq 2\bar{c}$ . The proof is completed.  $\square$



Now we are in a position to state our main result on the rate of convergence of the augmented Lagrangian method for the composite optimization problem.

**THEOREM 14.** *Suppose that Conditions (I) and (III) hold at  $\bar{x}$ . Define  $\rho_1 = 2\rho_0$  and  $\rho_2 = 4\mu_0$  (where  $\rho_0$  is defined by Proposition 12 and  $\mu_0$  is defined by Proposition 13). Then for any  $c \geq 2\bar{c}$ , there exist two positive numbers  $\varepsilon$  and  $\delta$  (both depending on  $c$ ) such that for any  $y \in \mathbb{B}_\delta(\bar{y})$ , the problem*

$$(47) \quad \begin{aligned} & \min L_c(x, z, y) \\ & \text{s.t. } (x, z) \in \mathbb{B}_\varepsilon(\bar{x}, \bar{x}) \end{aligned}$$

has a unique solution denoted  $(x_c(y), z_c(y))$ . And the function  $(x_c(y), z_c(y))$  is locally Lipschitz continuous on  $\mathbb{B}_\delta(\bar{y})$  and is semismooth at any point in  $\mathbb{B}_\delta(\bar{y})$ , and for any  $y \in \mathbb{B}_\delta(\bar{y})$ , we have

$$(48) \quad \|z_c(y) - \bar{x}\| \leq \rho_1 \|y - \bar{y}\|/c$$

and

$$(49) \quad \|(\mu_c(y), \lambda_c(y)) - (\bar{\mu}, \bar{\lambda})\| \leq \rho_2 \|y - \bar{y}\|/c.$$

*Proof.* Let  $c \geq 2\bar{c}$ . It follows from Proposition 6 that there exist two positive numbers  $\varepsilon_1 > 0$  and  $\delta_0 > 0$  (both depending on  $c$ ) and a locally Lipschitz continuous mapping  $z_c(\cdot)$  defined on  $\mathbb{B}_{\delta_0}(\bar{y})$ ,  $z_c(y)$  is the unique solution to

$$\min \varphi_c(z, y) : \text{s.t. } z \in \mathbb{B}_{\varepsilon_1}(\bar{x}).$$

Define  $x_c(y) = (P_c f)(z_c(y) - \lambda/c)$ , then  $x_c(y)$  is well-defined on  $\mathbb{B}_{\delta_0}(\bar{y})$  and it's a semismooth mapping on  $\mathbb{B}_{\delta_0}(\bar{y})$ . Let

$$\varepsilon = \min \left\{ \varepsilon_1, \max_{y \in \mathbb{B}_{\delta_0}(\bar{y})} \|x_c(y) - \bar{x}\| \right\},$$

then for any  $y \in \mathbb{B}_{\delta_0}(\bar{y})$ ,  $(x_c(y), z_c(y))$  is the unique solution to Problem (47).

Since  $z_c(\cdot)$  is locally Lipschitz continuous on  $\mathbb{B}_{\delta_0}(\bar{y})$  and it is directionally differentiable at  $\bar{y}$ , namely

$$\lim_{y \rightarrow \bar{y}} \frac{\|z_c(y) - z_c(\bar{y}) - (z_c)'(\bar{y}; y - \bar{y})\|}{\|y - \bar{y}\|} = 0.$$

By Proposition 6 and Proposition 7,  $\nabla \vartheta_c(\cdot)$  is semismooth at  $\bar{y}$ , and thus is also Bouligand-differentiable at  $\bar{y}$ . Then there exists  $\delta \in (0, \delta_0]$  such that for any  $y \in \mathbb{B}_\delta(\bar{y})$ ,

$$(50) \quad \|z_c(y) - z_c(\bar{y}) - (z_c)'(\bar{y}; y - \bar{y})\| \leq \rho_0 \|y - \bar{y}\|/c$$

and

$$(51) \quad \|\nabla \vartheta_c(y) - \nabla \vartheta_c(\bar{y}) - (\nabla \vartheta_c)'(\bar{y}; y - \bar{y})\| \leq \mu_0 \|y - \bar{y}\|/c.$$

Let  $y = (\mu, \lambda) \in \mathbb{B}_\delta(\bar{y})$  be an arbitrary point. For  $c \geq 2\bar{c}$ , from (30), (50) and the fact  $z_c(\bar{y}) = \bar{x}$ , we have

$$\|z_c(y) - \bar{x}\| \leq \|(z_c)'(\bar{y}; y - \bar{y})\| + \rho_0 \|y - \bar{y}\|/c = \rho_1 \|y - \bar{y}\|/c,$$

which shows the validity of (48).

Since  $\nabla\vartheta_c(\cdot)$  is semismooth at  $\bar{y}$ , there exists an element  $V \in \partial_B(\nabla\vartheta_c)(\bar{y})$  such that  $(\nabla\vartheta_c)'(\bar{y}; y - \bar{y}) = V(y - \bar{y})$ . Nothing that  $V$  is self-adjoint, we know from (36) and (24) in Proposition 7 that

$$(52) \quad \|V(y - \bar{y}) + c^{-1}(y - \bar{y})\| \leq 3\mu_0\|y - \bar{y}\|/c^2.$$

Therefore, we have from (51) and (52)

$$\begin{aligned} & \|y + c\nabla\vartheta_c(y) - \bar{y}\| \\ &= c\|\nabla\vartheta_c(y) - \nabla\vartheta_c(\bar{y}) - (\nabla\vartheta_c)'(\bar{y}; y - \bar{y}) + (\nabla\vartheta_c)'(\bar{y}; y - \bar{y}) + c^{-1}(y - \bar{y})\| \\ &= c\|\nabla\vartheta_c(y) - \nabla\vartheta_c(\bar{y}) - (\nabla\vartheta_c)'(\bar{y}; y - \bar{y})\| + c\|V(y - \bar{y}) + c^{-1}(y - \bar{y})\| \\ &\leq \mu_0\|y - \bar{y}\|/c + 3\mu_0\|y - \bar{y}\|/c = \rho_2\|y - \bar{y}\|/c, \end{aligned}$$

which, together with (22) and the definition of  $\mu_c(y)$  and  $\lambda_c(y)$ , proves (49). The proof is completed.  $\square$

**5. Discussions and conclusions.** In this section, we explain that the assumptions, (nlsdp-A1)(the constraint nondegeneracy condition) and (nlsdp-A2) (the strong second-order sufficiency optimality condition) in [13] required by the augmented Lagrange method for nonlinear semi-definite optimization problem, are sufficient conditions for Condition (III). This shows that the result of the rate of convergence in this paper is consistent with the main result in [13]. For notation convenience, we consider the following nonlinear semi-definite optimization problem:

$$(53) \quad \begin{aligned} & \min \quad m(z) \\ & \text{s.t.} \quad q(z) = 0, \\ & \quad \quad p(z) \in \mathbb{S}_+^p, \end{aligned}$$

where  $m : \mathfrak{R}^n \rightarrow \mathfrak{R}$ ,  $q : \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ ,  $p : \mathfrak{R}^n \rightarrow \mathbb{S}^p$  are twice continuously differentiable. Let  $(\bar{z}, \bar{\zeta}, \bar{\Xi}) \in \mathfrak{R}^n \times \mathfrak{R}^m \times \mathbb{S}^p$  be a Kuhn-Tucker point for Problem (53), namely

$$(54) \quad \begin{aligned} & \nabla_z L(\bar{z}, \bar{\zeta}, \bar{\Xi}) = 0, \\ & q(\bar{z}) = 0, \quad 0 \preceq p(\bar{z}) \perp \bar{\Xi} \succeq 0, \end{aligned}$$

where  $L(z, \zeta, \Xi)$  is the standard Lagrangian for Problem (53):

$$L(z, \zeta, \Xi) = m(z) + \zeta^T q(z) - \langle \Xi, p(z) \rangle.$$

Using our notations, we modify Proposition 4 in [13] as follows

**PROPOSITION 15.** *Suppose that Assumptions (nlsdp-A1) and (nlsdp-A2) are satisfied. Then there exists a positive number  $c_0$  such that for any  $c \geq c_0$  and  $W \in \partial\Pi_{\mathbb{S}_+^p}(\bar{\Xi} - cp(\bar{z}))$ ,*

$$\langle d, \mathcal{A}_c(\bar{\zeta}, \bar{\Xi}, W)d \rangle \geq \langle d, \bar{\mathcal{B}}_{c_0, c}(\bar{\zeta}, \bar{\Xi}, W)d \rangle > 0$$

for any  $d \in \text{Ker}\mathcal{J}q(\bar{z}) \setminus \{0\}$ , where

$$\bar{\mathcal{B}}_{c_0, c}(\bar{\zeta}, \bar{\Xi}, W) = \nabla_{xx}^2 L(\bar{z}, \bar{\zeta}, \bar{\Xi}) + c_0 A^T D_c A + 2c_0 C_{(\alpha, \gamma)}^T (\Theta_c)_{(\alpha, \gamma)} C_{(\alpha, \gamma)}$$

and  $A$ ,  $D_c$ ,  $C_{(\alpha, \gamma)}$  and  $(\Theta_c)_{(\alpha, \gamma)}$  are defined in [13, p.369-370].

By introducing an artificial variable  $Y \in \mathbb{S}^p$ , Problem (53) can be reformulated as

$$(55) \quad \begin{aligned} \min \quad & m(z) + \delta_{\mathbb{S}_+^p}(Y) \\ \text{s.t.} \quad & \begin{bmatrix} q(z) \\ Y - p(z) \end{bmatrix} = 0. \end{aligned}$$

Let  $x = (z, Y)$ ,  $g(x) = m(z)$ ,  $f(x) = \delta_{\mathbb{S}_+^p}(Y)$  and

$$h(x) = \begin{bmatrix} q(z) \\ Y - p(z) \end{bmatrix},$$

then Problem (55) takes the form of (1). It is easy to calculate

$$\nabla g(x) = \begin{bmatrix} \nabla m(z) \\ 0 \end{bmatrix}, \quad \partial f(x) = \begin{bmatrix} 0 \\ N_{\mathbb{S}_+^p}(Y) \end{bmatrix}, \quad Dh(x) = \begin{bmatrix} \mathcal{J}q(z) & 0 \\ -Dp(z) & \mathcal{I} \end{bmatrix},$$

where  $\mathcal{I} : \mathbb{S}^p \rightarrow \mathbb{S}^p$  is the identity operator. Let  $\mu = (\zeta; \Xi)$ . We have from (54) that

$$0 \in \partial f(\bar{x}) + \nabla g(\bar{x}) + \mathcal{J}h(\bar{x})^T \bar{\mu},$$

namely Condition (I) holds.

It is obvious that  $e_c f(x) = \frac{c}{2} \|Y - \Pi_{\mathbb{S}_+^p}(Y)\|^2$  and

$$\nabla e_c f(x) = \begin{bmatrix} 0 \\ c \Pi_{\mathbb{S}_-^p}(Y) \end{bmatrix}.$$

From the definition  $\bar{\lambda} = \nabla g(\bar{x}) + \mathcal{J}h(\bar{x})^T \bar{\mu}$ , we have

$$\bar{\lambda} = \begin{bmatrix} 0 \\ \bar{\Xi} \end{bmatrix}, \quad \bar{x} - \frac{\bar{\lambda}}{c} = \begin{bmatrix} \bar{z} \\ \bar{Y} - \frac{\bar{\Xi}}{c} \end{bmatrix}.$$

Then  $W \in \partial(\nabla e_c f) \left( \bar{x} - \frac{\bar{\lambda}}{c} \right) = \partial(\nabla e_c f) \left( \bar{z}, \bar{Y} - \frac{\bar{\Xi}}{c} \right)$  satisfies

$$(56) \quad W = \begin{bmatrix} 0 & 0 \\ 0 & V \end{bmatrix} \quad \text{with} \quad V \in \partial \Pi_{\mathbb{S}_-^p} \left( \bar{Y} - \frac{\bar{\Xi}}{c} \right).$$

When  $c \geq c_0$  and  $d \in \ker \mathcal{J}h(\bar{x}) \setminus \{0\}$ , one has that

$$\mathcal{J}q(\bar{z})d_z = 0, \quad d_Y = Dp(\bar{z})d_z.$$

Thus  $\forall W \in \partial(\nabla e_c f)(\bar{x} - \bar{\lambda}/c)$ ,

$$(57) \quad \begin{aligned} & \langle d, c_0 c^{-1} W d \rangle + \langle d, [\nabla^2 g(\bar{x}) + \sum_{i=1}^m \bar{\mu}_i \nabla^2 h_i(\bar{x})] d \rangle \\ & = \left\langle \begin{pmatrix} d_z \\ d_Y \end{pmatrix}, \begin{pmatrix} (\nabla^2 m(\bar{z}) + \sum_{j=1}^m \bar{\zeta}_j \nabla^2 q_j(\bar{z}) - D^2 p(\bar{z}) \bar{\Xi}) d_z \\ c_0 V d_Y \end{pmatrix} \right\rangle \\ & = \langle d_z, \nabla_{zz}^2 L(\bar{z}, \bar{\zeta}, \bar{\Xi}) d_z \rangle + c_0 \langle Dp(\bar{z}) d_z, V(Dp(\bar{z}) d_z) \rangle, \end{aligned}$$

where  $V$  and  $W$  satisfy (56). Let  $E = \bar{\Xi} - p(\bar{z})$  have the spectral decomposition  $E = P\Lambda P^T$  with  $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_p)$ . Let

$$\alpha = \{i : \lambda_i > 0\}, \quad \beta = \{i : \lambda_i = 0\}, \quad \gamma = \{i : \lambda_i < 0\}.$$

Then  $E$  can be expressed as the following form:

$$E = P \begin{bmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0_\beta & 0 \\ 0 & 0 & \Lambda_\gamma \end{bmatrix} P^T$$

and  $p(\bar{z}) - \frac{\bar{\Xi}}{c}$  has the following spectral decomposition

$$p(\bar{z}) - \frac{\bar{\Xi}}{c} = -P \begin{bmatrix} \frac{1}{c}\Lambda_\alpha & 0 & 0 \\ 0 & 0_\beta & 0 \\ 0 & 0 & \Lambda_\gamma \end{bmatrix} P^T.$$

Therefore, we have from Proposition 4 of [15] or [16], for any  $H \in \mathbb{S}^p$ , that

$$(58) \quad V(H) = P \begin{bmatrix} P_\alpha^T H P_\alpha & P_\alpha^T H P_\beta & \Omega_{\alpha\gamma}^c \circ P_\alpha^T H P_\gamma \\ P_\beta^T H P_\alpha & V_\beta(P_\beta^T H P_\beta) & 0 \\ \Omega_{\alpha\gamma}^{cT} \circ P_\gamma^T H P_\alpha & 0 & 0 \end{bmatrix} P^T$$

with

$$V_\beta \in \partial \Pi_{\mathbb{S}^{|\beta|}}(0), \quad \Omega_{ij}^c = \frac{\lambda_i}{\lambda_i + c|\lambda_j|}, i \in \alpha, j \in \gamma.$$

From the definition of  $(\Theta_c)_{(\alpha,\gamma)}$  in [13], we have  $\Omega_{\alpha\gamma}^c = (\Theta_c)_{(\alpha,\gamma)}$ . Therefore we obtain from (57), any  $d \in \text{Ker} \mathcal{J}h(\bar{x}) \setminus \{0\}$ , that

$$\begin{aligned} & \langle d, c_0 c^{-1} W d \rangle + \langle d, [\nabla^2 g(\bar{x}) + \sum_{i=1}^m \bar{\mu}_j \nabla^2 h_i(\bar{x})] d \rangle \\ & = \langle d_z, \bar{\mathcal{B}}_{c_0, c}(\bar{\zeta}, \bar{\Xi}, W) d_z \rangle > 0 \end{aligned}$$

in which the last inequality comes from Proposition 15 and  $d_z \in \text{Ker} \mathcal{J}q(\bar{z}) \setminus \{0\}$ . The above analysis shows that Condition (III) is satisfied when the constraint nondegeneracy condition the strong second-order sufficiency optimality condition hold for Problem (53).

At the end of this section, we give a short conclusion about the results in this paper. We consider an important class of composite optimization problems in which objective function is a summation of a proper lower semi-continuous convex function and a twice continuously differentiable function, and constraint set is defined by a set of equalities of twice continuously differentiable functions. The main difficulty is the non-smoothness of the convex function in the objective. Noticing that, even for some non-smooth convex functions, Moreau-Yosida regularization is  $C^{1,1}$  and we are able to characterize its generalized Hessian. This observation helps us to establish a set of second-order sufficiency optimality conditions in terms of generalized Hessian of Moreau-Yosida regularization of a convex function. By expressing the objective in a separable way, we find that proximal mapping (or the gradient of Moreau-Yosida regularization) can be used to develop an implementable augmented Lagrange method. Using the proposed strong second-order sufficiency optimality condition as a main

condition, we establish the rate of convergence of the augmented Lagrange method based on the second-order variational analysis of Moreau-Yosida regularization. As an example, we check that the assumptions used in [13] are sufficient to the strong second-order sufficiency optimality condition adopted in this paper, and this shows that we may use the analysis to obtain the same result for nonlinear semi-definite optimization problems.

#### REFERENCES

- [1] D.P. BERTSEKAS, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
- [2] J. F. BONNANS AND A. SHAPIRO, *Perturbation Analysis of Optimization Problems*, New York: Springer, 2000.
- [3] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983.
- [4] M.R.HESTENES, *Multiplier and gradient methods*, Journal of Optimization Theory and Applications, 4(1969),pp 303-320.
- [5] A. R. CONN, N.I.M GOULD AND PH. L. TOINT, *A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds*, SIAM Journal on Numerical Analysis, 28 (1991), pp545-572.
- [6] G. DEBREU, *Definite and semidefinite quadratic forms*, Econometrica, 20 (1952), pp295- 300.
- [7] G. GOLUB AND C.F. VAN LOAN, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 1996.
- [8] M.J.D. POWELL, *A method for nonlinear constraints in minimization problems*, In R. Fletcher, Editor, Optimization, Academic Press, New York, 1972, pp. 283C298.
- [9] R.T. ROCKAFELLAR, *A dual approach to solving nonlinear programming problems by unconstrained optimization*, Mathematical Programming, 5 (1973), pp354-373.
- [10] R.T. ROCKAFELLAR, *The multiplier method of Hestenes and Powell applied to convex programming*, Journal of Optimization Theory and Applications, 12 (1973), pp555-562.
- [11] R.T. ROCKAFELLAR, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization, 14 (1976), pp877-898.
- [12] R.T. ROCKAFELLAR, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of Operations Research, 1(1976), pp97-116.
- [13] DE-FENG SUN, JIE SUN, LI-WEI ZHANG, *The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming*, Mathematical Programming, 114 (2008), pp349-391.
- [14] R. T. ROCKAFELLAR AND R. J. -B. WETS, *Variational Analysis*, Springer, New York, 1998.
- [15] D. F. SUN, *The strong second order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications*, Math. Oper. Res. , 31 (2006), pp761-776.
- [16] D.F. SUN AND J. SUN, *Semismooth Matrix Valued Functions*, Mathematics of Operations Research, 27 (2002), pp150-169.