Journal of Computational Mathematics
Vol.xx, No.x, 2026, 1–19.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2510-m2024-0216

CHARACTERIZING THE RATE OF CONVERGENCE OF THE AUGMENTED LAGRANGE METHOD FOR NONLINEAR PROGRAMMING*

Yule Zhang
School of Science, Dalian Maritime University, Dalian 116085, China
Email: ylzhang@dlmu.edu.cn
Jihong Zhang
School of Science, Shenyang Ligong University, Shenyang 110159, China
Email: zjh7815040x@163.com
Jia Wu[1)]
Institute of Operations Research and Control Theory, School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China
Email: wujia@dlut.edu.cn

Abstract

The rate of convergence of the augmented Lagrangian method for solving nonlinear programming is studied under the Jacobian uniqueness conditions. It is demonstrated that, for a given multiplier vector (μ, λ), the rate of convergence of the augmented Lagrangian method is linear with respect to $\|(\mu, \lambda) - (\mu^*, \lambda^*)\|$ and the ratio constant is proportional to $1/c$ when the ratio $\|(\mu, \lambda) - (\mu^*, \lambda^*)\|/c$ is small enough, where c is the penalty parameter that exceeds a threshold $c^* > 0$ and (μ^*, λ^*) is the multiplier corresponding to a local minimum point. Importantly, the ratio constant of the Q-linear convergence of the sequence of multiplier vectors is estimated by the second-order derivative of the value function of the nonlinear optimization problem. This characterization gives an explicit expression for the rate constant of the Q-linear convergence of the sequence of multiplier vectors.

Mathematics subject classification: 90C30.
Key words: Nonlinear programming, Jacobian uniqueness conditions, Augmented Lagrangian method, Rate of convergence, Value function.

1. Introduction

Consider the nonlinear programming problem of the form

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & h(x) = 0, \\ & g(x) \leq 0, \end{aligned} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^q$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ are twice continuously differentiable in a neighborhood of a feasible point $\overline{x} \in \mathbb{R}^n$. Let the Lagrange function for problem (1.1) be

$$L(x, \mu, \lambda) = f(x) + \mu^\top h(x) + \lambda^\top g(x), \quad (x, \mu, \lambda) \in \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}^p.$$

* Received October 4, 2024 / Revised version received April 28, 2025 / Accepted October 27, 2025 / Published online March 19, 2026 /
1) Corresponding author

The augmented Lagrangian method (ALM) was initiated by Hestenes [5] and Powell [7] for solving optimization problems with equality constraints and was generalized by Rockafellar [9] to inequality constrained optimization problems. For convex programming, convergence properties of the augmented Lagrangian method were systematically studied by Rockafellar, see [9–11].

Let Π_K denote the projection operator onto a convex cone K, then the corresponding augmented Lagrange function is defined by

$$L_c(x,\mu,\lambda)=f(x)+\mu^\top h(x)+\frac{c}{2}\|h(x)\|^2+\frac{1}{2c}\left(\left\|\Pi_{\mathbb{R}^p_+}\left(\lambda+cg(x)\right)\right\|^2-\|\lambda\|^2\right).$$

The augmented Lagrange method for solving problem (1.1) can be expressed of the following form:

Step 0. Given $c_0>0, x^0\in\mathbb{R}^n, \mu^0\in\mathbb{R}^q$ and $\lambda^0\in\mathbb{R}^p_+, k=0$.

Step 1. If

$$\left\|\nabla_x L(x^k,\mu^k,\lambda^k)\right\|+\|h(x^k)\|+\left\|\lambda^k-\Pi_{\mathbb{R}^p_+}\left(\lambda^k+g(x^k)\right)\right\|=0,$$

then stop and (x^k,μ^k,λ^k) is a Karush-Kuhn-Tucker (KKT) pair.

Step 2. Solve the following problem:

$$x^{k+1}\in\operatorname{argmin} L_{c_k}(x,\mu^k,\lambda^k)$$

and compute

$$\mu^{k+1}=\mu^k+c_kh(x^{k+1}),\quad \lambda^{k+1}=\Pi_{\mathbb{R}^p_+}\left(\lambda^k+c_kg(x^{k+1})\right).$$

Step 3. Update c_{k+1}, set $k+1$ to k, and go to Step 1.

The study about local convergence properties of the augmented Lagrangian method for nonlinear programming is relatively complete. For the equality constrained problem, Powell [7] proved that the augmented Lagrangian method converges locally at a linear rate to a local minimum point when the linear independence constraint qualification and the second-order sufficient condition are satisfied. This result was stated in Bertsekas [1, Proposition 2.4] followed by an implicit function theorem based proof. Moreover, [1, Proposition 2.7] gave an important result about the linear rate of convergence in terms of the Hessian of the primal functional (namely value function popularly used in literatures). However, for nonlinear programming with both equality and inequality problems, the Jacobian uniqueness conditions, namely the conditions in assumption (S^+) in [1, p. 161] are used to analyze the augmented Lagrange method. Bertsekas [1, p. 162] pointed out that, under the Jacobian uniqueness conditions, the results about augmented Lagrange method for equality constrained optimization problems can be extended to studying problem (1.1). In this paper, we give detailed analysis of convergence properties of the augmented Lagrange method for solving problem (1.1), including demonstrating the theorem about the rate of convergence, and estimating the ratio constant of the linear convergence under the Jacobian uniqueness conditions.

We should point out that, without assuming the strict complementarity condition, Conn *et al.* [2], Contesse-Becker [3], and Ito and Kunisch [6] derived linear convergence rate for the augmented Lagrangian method in a weak formation compared with the result stated in Section 3.

The paper is organized as follows. In Section 2, we develop some properties of the augmented Lagrangian function under the Jacobian uniqueness conditions for problem (1.1), which will be used to prove results about the rate of convergence of ALM. In Section 3, we demonstrate the linear rate of convergence of ALM. In Section 4, the asymptotical convergence rate of Lagrange multipliers is analyzed, which estimates the ratio constant of the Q-linear convergence of the sequence of multipliers generated by ALM.

2. Properties of the Augmented Lagrangian

Let $\overline{x}$ be a feasible point of problem (1.1) around which f, h and g are twice differentiable. We need the following conditions, which are named as Jacobian uniqueness conditions in [1, pp. 161–162].

Definition 2.1. *We say Jacobian uniqueness conditions are satisfied at* $(\overline{x},\overline{\mu},\overline{\lambda})\in\mathbb{R}^n\times\mathbb{R}^q\times\mathbb{R}^p$, *if*

(i) The point $\overline{x}$ *is a stationary point and* $(\overline{\mu},\overline{\lambda})$ *is its corresponding multiplier, namely*

$$\nabla_x L(\overline{x},\overline{\mu},\overline{\lambda}) = 0, \quad h(\overline{x}) = 0, \quad 0 \geq g(\overline{x}) \perp \overline{\lambda} \geq 0.$$

(ii) The linear independence constraint qualification holds at $\overline{x}$, *namely the set of vectors*

$$\{\nabla h_1(\overline{x}), \ldots, \nabla h_q(\overline{x})\} \cup \{\nabla g_j(\overline{x}) : j \in I(\overline{x})\}$$

are independent, where $I(\overline{x}) = \{i : g_i(\overline{x}) = 0, i = 1, \ldots, p\}$.

(iii) The strict complementarity condition holds, namely $\overline{\lambda} - g(\overline{x}) > 0$.

(iv) The second-order sufficiency optimality conditions holds at $(\overline{x},\overline{\mu},\overline{\lambda})$, *namely for any* $d \in \mathcal{C}(\overline{x}), d \neq 0$,

$$d^\top \nabla^2_{xx} L(\overline{x},\overline{\mu},\overline{\lambda}) d > 0,$$

where $\mathcal{C}(\overline{x})$ *is the critical cone of problem* (1.1) *at* $\overline{x}$ *defined by*

$$\mathcal{C}(\overline{x}) = \left\{ d \in \mathbb{R}^n : \mathcal{J}h(\overline{x})d = 0,\ \mathcal{J}g(\overline{x})d \in T_{\mathbb{R}^p_-}\big(g(\overline{x})\big),\ \nabla f(\overline{x})^\top d \leq 0 \right\}.$$

In this section, we will present some important properties of the Jacobian uniqueness conditions of problem (1.1) and properties of the augmented Lagrangian function under this set of conditions. These properties are crucial for studying the rate of augmented Lagrange method.

For a KKT pair $(\overline{x},\overline{\mu},\overline{\lambda})$, assume that (iii) holds for $\alpha = I(\overline{x})$ and $\gamma = \{1, \ldots, p\} \setminus \alpha$, then

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

and

$$g_\alpha(\overline{x}) = 0, \quad g_\gamma(\overline{x}) < 0, \quad \lambda_\alpha > 0, \quad \lambda_\gamma = 0.$$

Under these conditions, for any $d \in \mathcal{C}(\overline{x})$, we have that $\mathcal{J}g_\alpha(\overline{x})d \leq 0$ and

$$\mathcal{J}_x L(\overline{x},\overline{\mu},\overline{\lambda})d = \mathcal{J}f(\overline{x})d + \overline{\mu}^\top \mathcal{J}h(\overline{x})d + \overline{\lambda}^\top \mathcal{J}g(\overline{x})d = 0.$$

Then the critical cone $C(\overline{x})$ is reduced to the following subspace:

$$C(\overline{x}) = \big\{ d \in \mathbb{R}^n : \mathcal{J}h(\overline{x})d = 0,\ \mathcal{J}g_\alpha(\overline{x})d = 0 \big\}. \tag{2.1}$$

If (iv) of Jacobian uniqueness conditions holds, then there exists $\beta_0 > 0$ such that

$$d^\top \nabla^2_{xx} L(\overline{x}, \overline{\mu}, \overline{\lambda}) d \geq \beta_0 \|d\|^2, \quad \forall\, d \in C(\overline{x}). \tag{2.2}$$

Define

$$\Psi(x, \mu, \lambda) = \begin{bmatrix} \nabla_x L(x, \mu, \lambda) \\ h(x) \\ \Pi_{\mathbb{R}^p_+}\big(g(x) + \lambda\big) - \lambda \end{bmatrix}. \tag{2.3}$$

Then the Jacobian of $\Psi(x, \mu, \lambda)$, denoted by $K_0(x, \mu, \lambda)$, is expressed as

$$K_0(x, \mu, \lambda) = \begin{bmatrix} \nabla^2_{xx} L(x, \mu, \lambda) & \mathcal{J}h(x)^\top & \mathcal{J}g(x)^\top \\ \mathcal{J}h(x) & 0 & 0 \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(g(x) + \lambda\big)\mathcal{J}g(x) & 0 & -I_p + \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(g(x) + \lambda\big) \end{bmatrix}. \tag{2.4}$$

It is easy to prove that $K_0(\overline{x}, \overline{\mu}, \overline{\lambda})$ is nonsingular under the Jacobian uniqueness conditions.

Lemma 2.1. *Let $\overline{x} \in \mathbb{R}^n$ be a point around which f, h and g are twice continuously differentiable. Let $(\overline{\mu}, \overline{\lambda}) \in \mathbb{R}^q \times \mathbb{R}^p$ be the multiplier such that the Jacobian uniqueness conditions hold at $(\overline{x}, \overline{\mu}, \overline{\lambda})$. Then $K_0(\overline{x}, \overline{\mu}, \overline{\lambda})$ is nonsingular.*

Proof. Consider the equation

$$K_0(\overline{x}, \overline{\mu}, \overline{\lambda}) \begin{pmatrix} d_x \\ d_\mu \\ d_\lambda \end{pmatrix} = 0,$$

where $d_x \in \mathbb{R}^n, d_\mu \in \mathbb{R}^q, d_\lambda \in \mathbb{R}^p$. This equation is equivalent to

$$\nabla^2_{xx} L(\overline{x}, \overline{\mu}, \overline{\lambda}) d_x + \mathcal{J}h(\overline{x})^\top d_\mu + \mathcal{J}g(\overline{x})^\top d_\lambda = 0, \tag{2.5a}$$

$$\mathcal{J}h(\overline{x}) d_x = 0, \tag{2.5b}$$

$$\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(g(\overline{x}) + \overline{\lambda}\big)\mathcal{J}g(\overline{x}) d_x - d_\lambda + \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(g(\overline{x}) + \overline{\lambda}\big) d_\lambda = 0. \tag{2.5c}$$

From the equality (2.5c) and the strict complementarity condition, we have $[d_\lambda]_\gamma = 0$, and $\mathcal{J}g_\alpha(\overline{x}) d_x = 0$. Under the assumptions of Jacobian uniqueness conditions, we know from (2.1) that $d_x \in C(\overline{x})$. Then, multiplying $d_x^\top$ to the Eq. (2.5a), we obtain that $d_x^\top \nabla^2_{xx} L(\overline{x}, \overline{\mu}, \overline{\lambda}) d_x = 0$. It follows from the second-order sufficiency optimality conditions that $d_x = 0$, which together with the equality (2.5a) and the linear independence constraint qualification implies $d_\mu = 0$, $d_\lambda = 0$. So we obtain that $K_0(\overline{x}, \overline{\mu}, \overline{\lambda})$ is nonsingular. □

The following result shows that $x \to L_c(x, \mu, \lambda)$ is strictly convex in a neighborhood of $\overline{x}$ when (μ, λ) is sufficiently close to $(\overline{\mu}, \overline{\lambda})$ and c is large enough. It should be pointed out that even without strict complementarity, similar results can been established, see, for example, [12, Proposition 5]. Here we state its conclusion under the Jacobian uniqueness conditions for consistency.

Proposition 2.1. *Let $(\overline{x}, \overline{\mu}, \overline{\lambda})$ be a KKT point of problem (1.1) at which the Jacobian uniqueness conditions are satisfied. Then there exists positive numbers $c_0^* > 0$ and $\delta_0 > 0$ such that $\nabla^2_{xx} L_c(x, \mu, \lambda)$ is positively definite when $(x, \mu, \lambda) \in \mathbf{B}_{\delta_0}(\overline{x}, \overline{\mu}, \overline{\lambda})$ and $c \geq c_0^*$.*

Proof. It is easy to check

$$\nabla_x L_c(x,\mu,\lambda) = \nabla f(x) + \mathcal{J}h(x)^\top\big(\mu + ch(x)\big) + \mathcal{J}g(x)^\top \Pi_{\mathbb{R}^p_+}\big(\lambda + cg(x)\big).$$

If $\lambda + cg(x) \neq 0$, then $\Pi_{\mathbb{R}^p_+}$ is differentiable at $\lambda + cg(x)$. In this case,

$$\begin{aligned}\nabla^2_{xx} L_c(x,\mu,\lambda) &= \nabla^2_{xx} L\big(x, \mu + ch(x), \Pi_{\mathbb{R}^p_+}\big(\lambda + cg(x)\big)\big)\\ &\quad + c\mathcal{J}h(x)^\top \mathcal{J}h(x) + c\mathcal{J}g(x)^\top \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(\lambda + cg(x)\big)\mathcal{J}g(x).\end{aligned}$$

Then we obtain for any $d \in \mathbb{R}^n$,

$$\begin{aligned}d^\top \nabla^2_{xx} L_c(\overline{x},\overline{\mu},\overline{\lambda})d &= d^\top \nabla^2_{xx} L(\overline{x},\overline{\mu},\overline{\lambda})d + c\|\mathcal{J}h(\overline{x})d\|^2\\ &\quad + c\big\langle \mathcal{J}g(\overline{x})d, \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(\overline{\lambda} + cg(\overline{x})\big)\mathcal{J}g(\overline{x})d\big\rangle\\ &= d^\top \nabla^2_{xx} L(\overline{x},\overline{\mu},\overline{\lambda})d + c\|\mathcal{J}h(\overline{x})d\|^2 + c\|\mathcal{J}g_\alpha(\overline{x})d\|^2.\end{aligned}$$

From (2.2) and the expression $\mathcal{C}(\overline{x})$ of (2.1), noting $\big(\mathcal{J}h(\overline{x})^\top \quad \mathcal{J}g_\alpha(\overline{x})^\top\big)$ has full column rank, by [4], we have that there exists a positive number $\overline{c}^*_0 > 0$ such that

$$d^\top \nabla^2_{xx} L_c(\overline{x},\overline{\mu},\overline{\lambda})d \geq \frac{\beta_0}{2}\|d\|^2, \quad \forall\, c \geq \overline{c}^*_0.$$

Therefore there exist positive numbers $c^*_0 > \overline{c}^*_0$ and $\delta_0 > 0$ such that $\nabla^2_{xx} L_c(x,\mu,\lambda)$ is positively definite when$(x,\mu,\lambda) \in \mathbf{B}_{\delta_0}(\overline{x},\overline{\mu},\overline{\lambda})$ and $c \geq c^*_0$. □

Suppose that $Z(x,\lambda,t) = g(x) + (t+1)\lambda \neq 0$ such that $\Pi_{\mathbb{R}^p_+}$ is differentiable at $Z(x,\lambda,t)$, in this case we define a matrix of the form

$$K(x,\mu,\lambda,t) = \begin{bmatrix} \nabla^2_{xx} L(x,\mu,\lambda) & \mathcal{J}h(x)^\top & \mathcal{J}g(x)^\top \\ \mathcal{J}h(x) & -tI_q & 0 \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z(x,\lambda,t)\big)\mathcal{J}g(x) & 0 & -(t+1)I_p + \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z(x,\lambda,t)\big)\end{bmatrix}. \quad (2.6)$$

The following proposition shows that $K(\overline{x},\overline{\mu},\overline{\lambda},t)$ is nonsingular and has a bounded inverse when $t > 0$ is small enough.

Proposition 2.2. *Let $(\overline{x},\overline{\mu},\overline{\lambda})$ be a KKT point of problem (1.1) at which Jacobian uniqueness conditions are satisfied and c^*_0 be given in Proposition 2.1. Then there exists a positive number $c^*_1 > c^*_0$ large enough such that $K(\overline{x},\overline{\mu},\overline{\lambda},t)$ is nonsingular and*

$$\|K(\overline{x},\overline{\mu},\overline{\lambda},t)^{-1}\| \leq \beta_0$$

for some positive constant $\beta_0 > 0$, if $t \in [0, t^*_1]$, where $t^*_1 = [c^*_1]^{-1}$.

Proof. Since $K(\overline{x},\overline{\mu},\overline{\lambda},0) = K_0(\overline{x},\overline{\mu},\overline{\lambda})$, we have from Lemma 2.1 that $K(\overline{x},\overline{\mu},\overline{\lambda},0)$ is nonsingular. Now we consider the case where $t > 0$. Consider the equation

$$K(\overline{x},\overline{\mu},\overline{\lambda},t)\begin{pmatrix} d_x \\ d_\mu \\ d_\lambda \end{pmatrix} = 0,$$

where $d_x \in \mathbb{R}^n, d_\mu \in \mathbb{R}^q, d_\lambda \in \mathbb{R}^p$. This equation is equivalent to

$$\nabla^2_{xx} L(\overline{x}, \overline{\mu}, \overline{\lambda}) d_x + \mathcal{J} h(\overline{x})^\top d_\mu + \mathcal{J} g(\overline{x})^\top d_\lambda = 0, \tag{2.7a}$$

$$\mathcal{J} h(\overline{x}) d_x - t d_\mu = 0, \tag{2.7b}$$

$$\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(g(\overline{x}) + (1+t)\overline{\lambda}\big)\mathcal{J} g(\overline{x}) d_x - (1+t) d_\lambda + \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(g(\overline{x}) + (1+t)\overline{\lambda}\big) d_\lambda = 0. \tag{2.7c}$$

From the equality (2.7b), we have

$$d_\mu = t^{-1} \mathcal{J} h(\overline{x}) d_x.$$

From the equality (2.7c), we have that

$$[d_\lambda]_\alpha = t^{-1} \mathcal{J} g_\alpha(\overline{x}) d_x, \quad [d_\lambda]_\gamma = 0.$$

Then, multiplying $d_x^\top$ to the equality (2.7a), we obtain

$$\begin{aligned} 0 &= d_x^\top \nabla^2_{xx} L(\overline{x}, \overline{\mu}, \overline{Y}) d_x + d_x^\top \mathcal{J} h(\overline{x})^\top d_\mu + \langle \mathcal{J} g(\overline{x}) d_x, d_\lambda \rangle \\ &= d_x^\top \nabla^2_{xx} L(\overline{x}, \overline{\mu}, \overline{Y}) d_x + t^{-1} \|\mathcal{J} h(\overline{x}) d_x\|^2 + t^{-1} \|\mathcal{J} g_\alpha(\overline{x}) d_x\|^2 \\ &= d_x^\top \nabla^2_{xx} L_{t^{-1}}(\overline{x}, \overline{\mu}, \overline{Y}) d_x, \end{aligned}$$

which implies $d_x = 0$ from Proposition 2.1 when $t < [c_0^*]^{-1}$. Therefore, we obtain $d_\mu = 0$, $d_\lambda = 0$ and $K(\overline{x}, \overline{\mu}, \overline{Y}, t)$ is nonsingular when $t < [c_0^*]^{-1}$.

Noting, for $\overline{Z} = g(\overline{x}) + \overline{\lambda}$, we have $Z(\overline{x}, \overline{\lambda}, t) = \overline{Z} + t\overline{\lambda}$ and $Z(\overline{x}, \overline{\lambda}, 0) = \overline{Z}$. Therefore, we get

$$\begin{aligned} &K(\overline{x}, \overline{\mu}, \overline{\lambda}, t) - K(\overline{x}, \overline{\mu}, \overline{\lambda}, 0) \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & -tI_q & 0 \\ \Big[\mathcal{J}\Pi_{\mathbb{R}^p_+}(\overline{Z} + t\overline{\lambda}) - \mathcal{J}\Pi_{\mathbb{R}^p_+}(\overline{Z})\Big] \mathcal{J} g(\overline{x}) & 0 & -tI_p + \Big[\mathcal{J}\Pi_{\mathbb{R}^p_+}(\overline{Z} + t\overline{\lambda}) - \mathcal{J}\Pi_{\mathbb{R}^p_+}(\overline{Z})\Big] \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & -tI_q & 0 \\ 0 & 0 & -tI_p \end{bmatrix}. \end{aligned} \tag{2.8}$$

Thus, we have, for $t \in [0, [c_0^*]^{-1})$, that

$$\|K(\overline{x}, \overline{\mu}, \overline{\lambda}, t) - K(\overline{x}, \overline{\mu}, \overline{\lambda}, 0)\| \le t.$$

Therefore there exists a positive number $c_1^* > c_0^*$ large enough, for $t_1^* = [c_1^*]^{-1}$, if $t \in [0, t_1^*]$, then $K(\overline{x}, \overline{\mu}, \overline{\lambda}, t)$ is nonsingular and

$$\|K(\overline{x}, \overline{\mu}, \overline{\lambda}, t)^{-1}\| \le \beta_0$$

for some positive constant $\beta_0 > 0$. The proof is complete. □

The following result gives an estimate of the norm of $K(x, \mu, \lambda, t)^{-1}$ when (x, μ, λ) is in a neighborhood of $(\overline{x}, \overline{\mu}, \overline{\lambda})$ and small $t > 0$, which is crucial in proving the main result of the paper.

Corollary 2.1. *Let $(\overline{x}, \overline{\mu}, \overline{\lambda})$ be a KKT point of problem* (1.1) *at which Jacobian uniqueness conditions are satisfied. Then there exist $\beta_1 \ge \beta_0, \delta_1 \in (0, \delta_0)$, and $c_2^* \ge c_1^*$ for $t_2^* = [c_2^*]^{-1}$, $K(x, \mu, \lambda, t)$ is nonsingular and*

$$\|K(x, \mu, \lambda, t)^{-1}\| \le \beta_1,$$

if $(x, \mu, \lambda) \in \mathbf{B}_{\delta_1}(\overline{x}, \overline{\mu}, \overline{\lambda})$ and $t \in [0, t_2^*)$.

Proof. The result comes from the continuity of $K(x, \mu, \lambda, t)$. □

3. The Rate of Convergence of the Augmented Lagrange Method

In this section, we consider the local convergence rate of the augmented Lagrange method for the nonlinear programming problem when the Jacobian uniqueness conditions are satisfied at $(\overline{x}, \overline{\mu}, \overline{\lambda})$. For $\delta > 0$, define

$$D(c^*, \delta) = \{(\mu, \lambda, c) : \|(\mu, \lambda) - (\overline{\mu}, \overline{\lambda})\| \leq \delta c,\ c \geq c^*\}.$$

Theorem 3.1. *Under Jacobian uniqueness conditions, there exist $\delta > 0, c^* > 0, \epsilon > 0$ and $\beta > 0$ such that for any $(\mu, \lambda, c) \in D(c^*, \delta)$, problem*

$$\begin{array}{ll} \min & L_c(x, \mu, \lambda) \\ \text{s.t.} & x \in \mathbf{B}_\epsilon(\overline{x}) \end{array} \tag{3.1}$$

has a unique solution, denoted by $x(\mu, \lambda, c)$, which is differentiable on $\operatorname{int} D(c^*, \delta)$. Furthermore, for all $(\mu, \lambda, c) \in \operatorname{int} D(c^*, \delta)$,

$$\begin{aligned} \|x(\mu, \lambda, c) - \overline{x}\| &\leq \frac{\beta}{c}\|(\mu, \lambda) - (\overline{\mu}, \overline{\lambda})\|, \\ \|\widetilde{\mu}(\mu, \lambda, c) - \overline{\mu}\| &\leq \frac{\beta}{c}\|(\mu, \lambda) - (\overline{\mu}, \overline{\lambda})\|, \\ \|\widetilde{\lambda}(\mu, \lambda, c) - \overline{\lambda}\| &\leq \frac{\beta}{c}\|(\mu, \lambda) - (\overline{\mu}, \overline{\lambda})\|. \end{aligned} \tag{3.2}$$

where

$$\widetilde{\mu}(\mu, \lambda, c) = \mu + ch\big(x(\mu, \lambda, c)\big), \quad \widetilde{\lambda}(\mu, \lambda, c) = \Pi_{\mathbb{R}^p_+}\big(\lambda + cg\big(x(\mu, \lambda, c)\big)\big).$$

Proof. If x is a local minimizer of $L_c(\cdot, \mu, \lambda)$, then, in view of the definition of $(\widetilde{\mu}, \widetilde{\lambda})$, we get

$$\begin{aligned} &\nabla f(x) + \mathcal{J}h(x)^\top \widetilde{\mu} + \mathcal{J}g(x)^\top \widetilde{\lambda} = 0, \\ &h(x) + \frac{1}{c}(\mu - \widetilde{\mu}) = 0, \\ &\Pi_{\mathbb{R}^p_+}\left(g(x) + \frac{1}{c}\lambda\right) - \frac{1}{c}\widetilde{\lambda} = 0. \end{aligned} \tag{3.3}$$

Define

$$\eta = \frac{1}{c}[\mu - \overline{\mu}], \quad \xi = \frac{1}{c}[\lambda - \overline{\lambda}], \quad t = \frac{1}{c},$$

noting

$$\Pi_{\mathbb{R}^p_+}\left(g(x) + \frac{1}{c}\lambda\right) - \frac{1}{c}\widetilde{\lambda} = 0 \iff \widetilde{\lambda} + \frac{1}{c}\widetilde{\lambda} = \Pi_{\mathbb{R}^p_+}\left(g(x) + \frac{1}{c}\lambda + \widetilde{\lambda}\right),$$

then (3.3) is equivalently expressed as $F(x, \widetilde{\mu}, \widetilde{\lambda}; \eta, \xi, t) = 0$, where

$$F(x, \widetilde{\mu}, \widetilde{\lambda}; \eta, \xi, t) = \begin{bmatrix} \nabla f(x) + \mathcal{J}h(x)^\top \widetilde{\mu} + \mathcal{J}g(x)^\top \widetilde{\lambda} \\ h(x) + \eta + t\overline{\mu} - t\widetilde{\mu} \\ \Pi_{\mathbb{R}^p_+}\big(g(x) + \widetilde{\lambda} + \xi + t\overline{\lambda}\big) - (1 + t)\widetilde{\lambda} \end{bmatrix}.$$

Obviously we have

$$F(\overline{x}, \overline{\mu}, \overline{\lambda}; 0, 0, t) = 0, \quad \forall\, t \in [0, t_2^*/2],$$

and

$$\begin{aligned}&\mathcal{J}_{(x,\widetilde{\mu},\widetilde{\lambda})}F(\overline{x},\overline{\mu},\overline{\lambda};0,0,t)\\&=\begin{bmatrix}\nabla^2_{xx}L(\overline{x},\overline{\mu},\overline{\lambda}) & \mathcal{J}h(\overline{x})^\top & \mathcal{J}g(\overline{x})^*\\ \mathcal{J}h(\overline{x}) & -tI & 0\\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(g(\overline{x})+(1+t)\overline{\lambda}\big)\mathcal{J}g(\overline{x}) & 0 & -(1+t)I_p+\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(g(\overline{x})+(1+t)\overline{\lambda}\big)\end{bmatrix}.\end{aligned}$$

Obviously, from the definition $K(x,\mu,\lambda,t)$ in (2.6), we have

$$\mathcal{J}_{(x,\widetilde{\mu},\widetilde{\lambda})}F(\overline{x},\overline{\mu},\overline{\lambda};0,0,t)=K(\overline{x},\overline{\mu},\overline{\lambda},t).$$

Then from Proposition 2.1, we have that $\mathcal{J}_{(x,\widetilde{\mu},\widetilde{\lambda})}F(\overline{x},\overline{\mu},\overline{\lambda};0,0,t)$ is nonsingular when $t\in[0,t_2^*)$.

Define $t^*=t_2^*/2$ and $c^*=[t^*]^{-1}$ and

$$\Omega=\{0\}\times\{0\}\times[0,t^*]\subset\mathbb{R}^n\times\mathbb{R}^p\times\mathbb{R},$$

it follows from of [1, Implicit Function Theorem 2, p. 12] that there exists $\delta\in(0,t_2^*/2)$ with $\delta<\delta_1, 0<\epsilon<\delta_1$ and mapping

$$\big(\widehat{x}(\cdot),\widehat{\mu}(\cdot),\widehat{\lambda}(\cdot)\big):\mathbf{B}(\Omega,\delta)\ \to\ \mathbf{B}_\epsilon\big((\overline{x},\overline{\mu},\overline{\lambda})\big),$$

which is differentiable on $\operatorname{int}\mathbf{B}(\Omega,\delta)$ and satisfies

$$\begin{aligned}&(\overline{x},\overline{\mu},\overline{\lambda})=\big(\widehat{x}(0,0,t),\widehat{\mu}(0,0,t),\widehat{\lambda}(0,0,t)\big),\\&F\big(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t);\eta,\xi,t\big)=0,\quad\forall\,(\eta,\xi,t)\in\mathbf{B}(\Omega,\delta).\end{aligned}\tag{3.4}$$

From Proposition 2.1 and Corollary 2.1, we may choose $\delta>0$ and $\epsilon>0$ small enough such that constraint nondegeneracy condition holds at $\widehat{x}(\eta,\xi,t)$, $\nabla^2_{xx}L_{t^{-1}}(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t))$ is positively definite and

$$\big\|K\big(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t),t\big)^{-1}\big\|\le\beta_1,\quad\forall\,(\eta,\xi,t)\in\mathbf{B}(\Omega,\delta).$$

Differentiating the three equations in (3.4) with respect to (η,ξ,t), we obtain

$$\begin{aligned}&K\big(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t),t\big)\mathcal{J}_{(\eta,\xi,t)}\begin{bmatrix}\widehat{x}(\eta,\xi,t)\\ \widehat{\mu}(\eta,\xi,t)\\ \widehat{\lambda}(\eta,\xi,t)\end{bmatrix}\\&=\begin{bmatrix}0 & 0 & 0\\ -I_q & 0 & \widehat{\mu}(\eta,\xi,t)-\overline{\mu}\\ 0 & -\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(\widehat{Z}_t(\eta,\xi,t)\big) & \widehat{\lambda}-\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(\widehat{Z}_t(\eta,\xi,t)\big)\overline{\lambda}\end{bmatrix},\end{aligned}\tag{3.5}$$

where

$$\widehat{Z}_t(\eta,\xi,t)=g\big(\widehat{x}(\eta,\xi,t)\big)+\widehat{\lambda}(\eta,\xi,t)+\xi+t\overline{\lambda}.$$

It follows from (3.5) for

$$z(s)=(s\eta,sU,st),\quad\widehat{Z}(s)=g\big(\widehat{x}\big(z(s)\big)\big)+\widehat{\lambda}\big(z(s)\big)+\xi+s\overline{\lambda}$$

with $t \in [0, t_2^*/2]$ that

$$\begin{aligned}
\begin{bmatrix} \widehat{x}(\eta,\lambda,t)-\overline{x} \\ \widehat{\mu}(\eta,\lambda,t)-\overline{\mu} \\ \widehat{\lambda}(\eta,\lambda,t)-\overline{\lambda} \end{bmatrix} &= \begin{bmatrix} \widehat{x}(\eta,\lambda,t)-\widehat{x}(0,0,0) \\ \widehat{\mu}(\eta,\lambda,t)-\widehat{\mu}(0,0,0) \\ \widehat{\lambda}(\eta,\lambda,t)-\widehat{\lambda}(0,0,0) \end{bmatrix} \\
&= \int_0^1 K\left(\widehat{x}\big(z(s)\big), \widehat{\mu}\big(z(s)\big), \widehat{\lambda}\big(z(s)\big), s\right)^{-1} \\
&\quad \times \begin{bmatrix} 0 & 0 & 0 \\ -I_q & 0 & \widehat{\mu}\big(z(s)\big)-\overline{\mu} \\ 0 & -\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z(s)\big) & \widehat{\lambda}-\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z(s)\big)\overline{\lambda} \end{bmatrix} \begin{bmatrix} \eta \\ \xi \\ t \end{bmatrix} \mathrm{d}s \\
&= \int_0^1 K\left(\widehat{x}\big(z(s)\big), \widehat{\mu}\big(z(s)\big), \widehat{\lambda}\big(z(s)\big), s\right)^{-1} \\
&\quad \times \begin{bmatrix} 0 \\ -\eta + \big(\widehat{\mu}\big(z(s)\big)-\overline{\mu}\big)\, t \\ -\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z(s)\big)\xi + \big(\widehat{\lambda}\big(z(s)\big)-\overline{\lambda}\big)t \end{bmatrix} \mathrm{d}s \\
&\quad + \int_0^1 K\left(\widehat{x}\big(z(s)\big), \widehat{\mu}\big(z(s)\big), \widehat{\lambda}\big(z(s)\big), s\right)^{-1} \begin{bmatrix} 0 \\ 0 \\ t\mathcal{J}\Pi_{\mathbb{R}^p_-}\big(Z(s)\big)\overline{\lambda} \end{bmatrix} \mathrm{d}s. \qquad (3.6)
\end{aligned}$$

Noting that

$$\left\| K\big(\widehat{x}\big(z(s)\big), \widehat{\mu}\big(z(s)\big), \widehat{\lambda}\big(z(s)\big), s\big)^{-1} \right\| \le \beta_1$$

for $(\eta, \xi, t) \in \mathbf{B}(\Omega, \delta)$ and $s \in [0, 1]$, we obtain from (3.6) and

$$\left\| \mathcal{J}\Pi_{\mathbb{R}^p_+}\Big(g\big(\widehat{x}\big(z(s)\big)\big) + \widehat{\lambda}\big(z(s)\big) + \xi + s\overline{\lambda}\Big) \right\| \le 1,$$

that

$$\begin{aligned}
&\|\widehat{x}(\eta,\lambda,t)-\overline{x}\|^2 + \|\widehat{\mu}(\eta,\lambda,t)-\overline{\mu}\|^2 + \|\widehat{\lambda}(\eta,\lambda,t)-\overline{\lambda}\|^2 \\
&\le 4\beta_1^2 \int_0^1 \Big[\|\eta\|^2 + \big\|\widehat{\mu}\big(z(s)\big)-\overline{\mu}\big\|^2 t^2 + \|\xi\|^2 + \big\|\widehat{\lambda}\big(z(s)\big)-\overline{\lambda}\big\|^2 t^2\Big] \mathrm{d}s \\
&\quad + 2\beta_1^2 \int_0^1 \Big[\|\mathcal{J}\Pi_{\mathbb{R}^p_-}\Big(g\big(\widehat{x}\big(z(s)\big)\big) + \widehat{\lambda}\big(z(s)\big) + \xi + s\overline{\lambda}\Big)\overline{\lambda}\|^2 t^2\Big] \mathrm{d}s. \qquad (3.7)
\end{aligned}$$

Noting that $\Pi_{\mathbb{R}^p_-}$ is twice continuously differentiable at $g(\overline{x}) + (s+1)\overline{\lambda}$, we have

$$\begin{aligned}
&\mathcal{J}\Pi_{\mathbb{R}^p_-}\Big(g\big(\widehat{x}\big(z(s)\big)\big) + \widehat{\lambda}\big(z(s)\big) + \xi + s\overline{\lambda}\Big)\overline{\lambda} \\
&= \mathcal{J}\Pi_{\mathbb{R}^p_-}\big(g(\overline{x}) + (s+1)\overline{\lambda}\big)\overline{\lambda} \\
&\quad + \overline{\lambda}^\top \nabla^2\Pi_{\mathbb{R}^p_-}\big(g(\overline{x}) + (s+1)\overline{\lambda}\big)\Big[g\big(\widehat{x}\big(z(s)\big)\big) + \xi - g(\overline{x}) + \widehat{\lambda}\big(z(s)\big) - \overline{\lambda}\Big] \\
&\quad + \mathrm{o}\Big(\big\|g\big(\widehat{x}\big(z(s)\big)\big) + \xi - g(\overline{x}) + \widehat{\lambda}\big(z(s)\big) - \overline{\lambda}\big\|\Big).
\end{aligned}$$

It is easy to check the equality $\mathcal{J}\Pi_{\mathbb{R}^p_-}(g(\overline{x}) + (s+1)\overline{\lambda})\overline{\lambda} = 0$. Then, when $\delta > 0$ is chosen small enough, there exists a positive constant $\kappa_0 > 0$ such that

$$\left\| \mathcal{J}\Pi_{\mathbb{R}^p_-}\Big(g\big(\widehat{x}\big(z(s)\big)\big) + \widehat{\lambda}\big(z(s)\big) + \xi + s\overline{\lambda}\Big)\overline{\lambda} \right\|$$

$$\leq 2\left|\overline{\lambda}^\top \nabla^2 \Pi_{\mathbb{R}^p_-}\left(g(\overline{x})+(s+1)\overline{\lambda}\right)\left[g\big(\widehat{x}(z(s))\big)+\xi-g(\overline{x})+\widehat{\lambda}\big(z(s)\big)-\overline{\lambda}\right]\right|$$
$$\leq \kappa_0\left[\|\xi\|+\|\widehat{x}\big(z(s)\big)-\overline{x}\|+\|\widehat{\lambda}\big(z(s)\big)-\overline{\lambda}\|\right],$$

when $(\eta,\xi,t)\in \mathbf{B}(\Omega,\delta)$ and $s\in[0,1]$.

Combine this estimate with (3.7), we obtain

$$\begin{aligned}&\|\widehat{x}(\eta,\lambda,t)-\overline{x}\|^2+\|\widehat{\mu}(\eta,\lambda,t)-\overline{\mu}\|^2+\|\widehat{\lambda}(\eta,\lambda,t)-\overline{\lambda}\|^2\\ \leq{}& 4\beta_1^2\|\eta\|^2+4\beta_1^2\big[1+\kappa_0^2t^2\big]\|\xi\|^2\\ &+4\beta_1^2t^2\int_0^1\left[\big\|\widehat{\mu}\big(z(s)\big)-\overline{\mu}\big\|^2+\big(1+2\kappa_0^2\big)\big\|\widehat{\lambda}\big(z(s)\big)-\overline{\lambda}\big\|^2+2\kappa_0^2\big\|\widehat{x}\big(z(s)\big)-\overline{x}\big\|^2\right]\mathrm{d}s. \end{aligned}\tag{3.8}$$

Substituting $(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t)),\widehat{\lambda}(\eta,\xi,t))$ by $(\widehat{x}(z(s)),\widehat{\mu}(z(s)),\widehat{\lambda}(z(s)))$ in (3.8) yields

$$\begin{aligned}&\big\|\widehat{x}\big(z(s)\big)-\overline{x}\big\|^2+\big\|\widehat{\mu}\big(z(s)\big)-\overline{\mu}\big\|^2+\big\|\widehat{\lambda}\big(z(s)\big)-\overline{\lambda}\big\|^2\\ \leq{}& 4\beta_1^2\|\eta\|^2+4\beta_1^2\big[1+\kappa_0^2t^2\big]\|\xi\|^2\\ &+4\beta_1^2t^2\int_0^1\left[\big\|\widehat{\mu}\big(z(s)\big)-\overline{\mu}\big\|^2+\big(1+2\kappa_0^2\big)\big\|\widehat{\lambda}\big(z(s)\big)-\overline{\lambda}\big\|^2+2\kappa_0^2\big\|\widehat{x}\big(z(s)\big)-\overline{x}\big\|^2\right]\mathrm{d}s. \end{aligned}\tag{3.9}$$

From the arbitrariness of $s\in[0,1]$ in (3.9), we obtain

$$\begin{aligned}&\max_{0\leq s\leq 1}\left\{\big\|\widehat{x}\big(z(s)\big)-\overline{x}\big\|^2+\big\|\widehat{\mu}\big(z(s)\big)-\overline{\mu}\big\|^2+\big\|\widehat{\lambda}\big(z(s)\big)-\overline{\lambda}\big\|^2\right\}\\ \leq{}& 4\beta_1^2\|\eta\|^2+4\beta_1^2\big[1+\kappa_0^2t^2\big]\|\xi\|^2\\ &+4\beta_1^2\big(1+2\kappa_0^2\big)t^2\cdot\max_{0\leq s\leq 1}\left\{\big\|\widehat{x}\big(z(s)\big)-\overline{x}\big\|^2+\big\|\widehat{\mu}\big(z(s)\big)-\overline{\mu}\big\|^2+\big\|\widehat{\lambda}\big(z(s)\big)-\overline{\lambda}\big\|^2\right\},\end{aligned}$$

which implies

$$\begin{aligned}&\|\widehat{x}(\eta,\lambda,t)-\overline{x}\|^2+\|\widehat{\mu}(\eta,\lambda,t)-\overline{\mu}\|^2+\|\widehat{\lambda}(\eta,\lambda,t)-\overline{\lambda}\|^2\\ \leq{}&\frac{4\beta_1^2\big[1+\kappa_0^2t^2\big]}{1-4\beta_1^2\big(1+2\kappa_0^2\big)t^2}[\|\eta\|+\|\xi\|]^2,\quad \forall\,(\eta,\xi,t)\in\mathbf{B}(\Omega,\delta)\end{aligned}$$

or

$$\left\|\begin{pmatrix}\widehat{x}(\eta,\xi,t)-\overline{x}\\ \widehat{\mu}(\eta,\xi,t)-\overline{\mu}\\ \widehat{\lambda}(\eta,\xi,t)-\overline{\lambda}\end{pmatrix}\right\|\leq\frac{2\beta_1\sqrt{1+\kappa_0^2t^2}}{\sqrt{1-4\beta_1^2\big(1+2\kappa_0^2\big)t^2}}[\|\eta\|+\|\xi\|],\quad\forall\,(\eta,\xi,t)\in\mathbf{B}(\Omega,\delta).\tag{3.10}$$

Define

$$x(\mu,\lambda,c)=\widehat{x}(\eta,\xi,t),\quad \widetilde{\mu}(\mu,\lambda,c)=\widehat{\mu}(\eta,\xi,t),\quad \widetilde{\lambda}(\mu,\xi,c)=\widehat{\lambda}(\eta,\xi,t),\quad\forall\,(\eta,\xi,t)\in\mathbf{B}(\Omega,\delta).$$

From the definitions of $D(c_*,\delta)$ and Ω, we have that

$$(\mu,\lambda,c)\in D(c_*,\delta)\implies(\eta,\xi,t)\in\mathbf{B}(\Omega,\delta),\quad(\eta,\xi,t)=\left(\frac{\mu-\overline{\mu}}{c},\frac{Y-\overline{\lambda}}{c},\frac{1}{c}\right).$$

It follows from (3.4) that

$$\big(x(\overline{\mu},\overline{\lambda},c),\widetilde{\mu}(\overline{\mu},\overline{\lambda},c),\widetilde{\lambda}(\overline{\mu},\overline{\lambda},c)\big)=(\overline{x},\overline{\mu},\overline{\lambda}),$$

and

$$\nabla_x L_c\big(x(\mu,\lambda,c),\mu,\lambda\big) = \nabla_x L\big(x(\mu,\lambda,c),\widetilde{\mu}(\mu,\lambda,c),\widetilde{\lambda}(\mu,\lambda,c)\big) = 0,$$
$$\widetilde{\mu}(\mu,\lambda,c) = \mu + ch\big(x(\mu,\lambda,c)\big),\widetilde{\lambda}(\mu,\lambda,c) = \Pi_{\mathbb{R}^p_+}\big(\lambda + cg\big(x(\mu,\lambda,c)\big)\big).$$

Noting that $(x(\mu,\lambda,c),\widetilde{\mu}(\mu,\lambda,c),\widetilde{\lambda}(\mu,\lambda,c)) \in \mathbf{B}_\epsilon(\overline{x},\overline{\mu},\overline{\lambda})$ and $\epsilon < \delta_1 \le \delta_0$ and $c \ge c^* \ge c_0^*$ we have from Proposition 2.1 that

$$\nabla^2_{xx} L_c\big(x(\mu,\lambda,c),\mu,\lambda\big) \succ 0.$$

Thus, $x(\mu,\xi,c)$ is the unique solution of problem (3.1) and differentiable on $\operatorname{int} D(c^*,\delta)$. Without loss of generality, suppose

$$c^* > \sqrt{\kappa_0^2 + 8\beta_1^2\big(1 + 2\kappa_0^2\big)},$$

and define $\beta = 4\beta_1$. The for any $(\mu,\lambda,c) \in D(c^*,\delta)$, we obtain from (3.10) that

$$\left\|\begin{pmatrix} x(\mu,\lambda,c) - \overline{x} \\ \widetilde{\mu}(\mu,\lambda,c) - \overline{\mu} \\ \widetilde{\lambda}(\mu,\lambda,c) - \overline{\lambda}\end{pmatrix}\right\| \le \frac{\beta}{c}\left[\|\mu - \overline{\mu}\| + \|\lambda - \overline{\lambda}\|\right], \tag{3.11}$$

which implies the estimates (3.2). □

4. Asymptotical Superlinear Convergence Rate of Multipliers

In Theorem 3.1, the rate of convergence of the augmented Lagrange method is characterized by (3.2), which involves a constant β. In this section, we estimate β by the eigenvalues of the second-order derivative of the value function of problem (1.1).

Let $(\overline{x},\overline{\mu},\overline{\lambda})$ be a Kurash-Kuhn-Tucker point of problem (1.1). Consider the following system of equations in (x,μ,λ,u):

$$\begin{aligned} &\nabla f(x) + \mathcal{J}h(x)^\top \mu + \mathcal{J}g(x)^\top \lambda = 0,\\ &h(x) + u_h = 0,\\ &\Pi_{\mathbb{R}^p_+}\big(c\big(g(x) + u_g\big) + \lambda\big) - \lambda = 0,\end{aligned} \tag{4.1}$$

then $(\overline{x},\overline{\mu},\overline{\lambda})$ is a solution of (4.1) for any where $c > 0$. By the standard implicit function theorem, there exist a constant $\delta > 0$ and functions $(x(\cdot),\mu(\cdot),\lambda(\cdot)) \in \mathcal{C}^1(\mathbf{B}_\delta(0))$ such that

$$x(0) = \overline{x}, \quad \mu(0) = \overline{\mu}, \quad \lambda(0) = \overline{\lambda},$$

and for $\|u\| \le \delta$, where $u = (u_h,u_g) \in \mathbb{R}^q \times \mathbb{R}^p$,

$$\begin{aligned} &\nabla f\big(x(u)\big) + \mathcal{J}h\big(x(u)\big)^\top \mu(u) + \mathcal{J}g\big(x(u)\big)^\top \lambda(u) = 0,\\ &h\big(x(u)\big) + u_h = 0,\\ &\Pi_{\mathbb{R}^p_+}\big(c\big(g\big(x(u)\big) + u_g\big) + \lambda(u)\big) - \lambda(u) = 0.\end{aligned} \tag{4.2}$$

Moreover, there exists $\varepsilon > 0$ such that

$$x(u) \in \mathbf{B}_\varepsilon(\overline{x}), \quad \mu(u) \in \mathbf{B}_\varepsilon(\overline{\mu}), \quad \lambda(u) \in \mathbf{B}_\varepsilon(\overline{\lambda})$$

for $\|u\| < \delta$. Define the following function $p : \mathbf{B}_\delta(0) \to \mathbb{R}$,

$$p(u) = f\big(x(u)\big), \quad u \in \mathbf{B}_\delta(0). \tag{4.3}$$

In view of Jacobian uniqueness conditions, δ and ε can be taken sufficiently small so that $x(u)$ is actually a local minimum point of the following perturbed problem:

$$\min_{x\in\mathbb{R}^n} \{f(x) : h(x) + u_h = 0,\ g(x) + u_g \leq 0\}. \tag{4.4}$$

Thus, an equivalent definition of p is given by

$$p(u) = f\big(x(u)\big) = \min_{x\in\mathbf{B}_\varepsilon(\overline{x})} \{f(x) : h(x) + u_h = 0,\ g(x) + u_g \leq 0\}, \quad u \in \mathbf{B}_\delta(0). \tag{4.5}$$

Lemma 4.1. *Suppose that Jacobian uniqueness conditions hold at $(\overline{x}, \overline{\mu}, \overline{\lambda})$, and δ and ε can be taken sufficiently small so that $x(u)$ is a local minimum point of problem* (4.4) *in the sense of* (4.5)*. Then*

$$\nabla p(u) = \begin{pmatrix} \mu(u) \\ \lambda(u) \end{pmatrix}, \quad \forall\, u \in \mathbf{B}_\delta(0), \tag{4.6}$$

where $\mu(u), \lambda(u)$ satisfy (4.2)*.*

Proof. Let the Lagrange function of problem (4.4) be

$$\mathcal{L}(x, \mu, \lambda; u) = f(x) + \mu^\top\big(h(x) + u_h\big) + \lambda^\top\big(g(x) + u_g\big).$$

Then we may express $p(u)$ as follows:

$$\begin{aligned} p(u) &= f\big(x(u)\big) + \big\langle \mu(u), h\big(x(u)\big) + u_h\big\rangle + \big\langle \lambda(u), g\big(x(u)\big) + u_g\big\rangle \\ &= \mathcal{L}\big(x(u), \mu(u), \lambda(u); u\big). \end{aligned}$$

Using this formula and noting

$$\nabla_x \mathcal{L}\big(x(u), \mu(u), \lambda(u); u\big) = 0, \quad \nabla_{(\mu,\lambda)} \mathcal{L}\big(x(u), \mu(u), \lambda(u); u\big) = 0,$$

we obtain

$$\begin{aligned} \nabla p(u) &= \mathcal{J}x(u)^\top \nabla_x \mathcal{L}\big(x(u), \mu(u), \lambda(u); u\big) + \mathcal{J}\mu(u)^\top \nabla_\mu \mathcal{L}\big(x(u), \mu(u), \lambda(u); u\big) \\ &\quad + \mathcal{J}\lambda(u)^\top \nabla_\lambda \mathcal{L}\big(x(u), \mu(u), \lambda(u); u\big) + \nabla_u \mathcal{L}\big(x(u), \mu(u), \lambda(u); u\big) \\ &= \nabla_u \mathcal{L}\big(x(u), \mu(u), \lambda(u); u\big) = \begin{pmatrix} \mu(u) \\ \lambda(u) \end{pmatrix}. \end{aligned}$$

The proof is complete. ☐

Lemma 4.2. *Suppose that Jacobian uniqueness conditions hold and δ and ε can be taken sufficiently small so that $x(u)$ is a local minimum point of problem* (4.4)*. Then*

$$\begin{aligned} & c\begin{bmatrix} I_q & 0 \\ 0 & I_p \end{bmatrix} + \begin{bmatrix} \mathcal{J}\mu(u) \\ \mathcal{J}\lambda(u) \end{bmatrix} \\ &= \left\{ \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{1}{c}\big(I_p - \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\big) \end{bmatrix} + \begin{bmatrix} \mathcal{J}h\big(x(u)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\mathcal{J}g\big(x(u)\big) \end{bmatrix} \right. \\ &\qquad \left. \times \nabla^2_{xx} L_c\big(x(u), \mu(u), \lambda(u)\big)^{-1} \begin{bmatrix} \mathcal{J}h\big(x(u)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\mathcal{J}g\big(x(u)\big) \end{bmatrix}^\top \right\}^{-1}, \end{aligned} \tag{4.7}$$

where $Z_c(u) = c(g(x(u)) + u_g) + \lambda(u)$ and $\mu(u), \lambda(u)$ satisfy (4.2)*.*

Proof. Differentiating (4.2), we obtain

$$\nabla^2_{xx}\mathcal{L}\big(x(u),\mu(u),\lambda(u)\big)\mathcal{J}x(u)+\mathcal{J}h\big(x(u)\big)^\top\mathcal{J}\mu(u)+\mathcal{J}g\big(x(u)\big)^\top\mathcal{J}\lambda(u)=0, \tag{4.8}$$

and

$$\begin{aligned}&\mathcal{J}h\big(x(u)\big)\mathcal{J}x(u)=\begin{bmatrix}-I_q & 0\end{bmatrix},\\ &\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(c\big(g\big(x(u)\big)+u_g\big)+\lambda(u)\big)\Big[c\mathcal{J}g\big(x(u)\big)\mathcal{J}x(u)+c\begin{bmatrix}0 & I_p\end{bmatrix}+\mathcal{J}\lambda(u)\Big]-\mathcal{J}\lambda(u)=0.\end{aligned} \tag{4.9}$$

From the definition of $Z_c(u)$, the Eqs. (4.8) and (4.9) can be written as

$$\begin{aligned}&\begin{bmatrix}\nabla^2_{xx}\mathcal{L}\big(x(u),\mu(u),\lambda(u)\big) & \mathcal{J}\big(x(u)\big)^\top & \mathcal{J}g\big(x(u)\big)^\top\\ \mathcal{J}h\big(x(u)\big) & 0 & 0\\ c\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\mathcal{J}g\big(x(u)\big) & 0 & -I_p+\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\end{bmatrix}\begin{bmatrix}\mathcal{J}x(u)\\ \mathcal{J}\mu(u)\\ \mathcal{J}\lambda(u)\end{bmatrix}\\ &=\begin{bmatrix}0 & 0\\ -I_q & 0\\ 0 & -c\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\end{bmatrix}\end{aligned}$$

or

$$\begin{aligned}&\left(\overline{\mathbb{K}}_c\big(x(u),\mu(u),\lambda(u)\big)+\begin{bmatrix}0 & 0 & 0\\ 0 & c^{-1}I_q & 0\\ 0 & 0 & c^{-1}\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\end{bmatrix}\right)\begin{bmatrix}\mathcal{J}x(u)\\ \mathcal{J}\mu(u)\\ \mathcal{J}\lambda(u)\end{bmatrix}\\ &=\begin{bmatrix}0 & 0\\ -I_q & 0\\ 0 & -\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\end{bmatrix},\end{aligned} \tag{4.10}$$

where

$$\overline{\mathbb{K}}_c\big(x(u),\mu(u),\lambda(u)\big)=\begin{bmatrix}\nabla^2_{xx}\mathcal{L}\big(x(u),\mu(u),\lambda(u)\big) & \mathcal{J}h\big(x(u)\big)^\top & \mathcal{J}g\big(x(u)\big)^\top\\ \mathcal{J}h\big(x(u)\big) & -c^{-1}I_q & 0\\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\mathcal{J}g\big(x(u)\big) & 0 & -c^{-1}I_p\end{bmatrix}.$$

Thus, Eq. (4.10) is equivalent to

$$\begin{aligned}&\overline{\mathbb{K}}_c\big(x(u),\mu(u),\lambda(u)\big)\begin{bmatrix}\mathcal{J}x(u)\\ \mathcal{J}\mu(u)\\ \mathcal{J}\lambda(u)\end{bmatrix}\\ &=-\begin{bmatrix}I & 0 & 0\\ 0 & I_q & 0\\ 0 & & \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\end{bmatrix}\left\{\begin{bmatrix}0 & 0\\ I_q & 0\\ 0 & I_p\end{bmatrix}+c^{-1}\begin{bmatrix}0\\ \mathcal{J}\mu(u)\\ \mathcal{J}\lambda(u)\end{bmatrix}\right\}.\end{aligned}$$

Therefore, we get that

$$\begin{aligned}\begin{bmatrix}\mathcal{J}x(u)\\ \mathcal{J}\mu(u)\\ \mathcal{J}\lambda(u)\end{bmatrix}&=-c^{-1}\overline{\mathbb{K}}_c\big(x(u),\mu(u),\lambda(u)\big)^{-1}\begin{bmatrix}I & 0 & 0\\ 0 & I_q & 0\\ 0 & 0 & \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\end{bmatrix}\\ &\quad\times\left\{c\begin{bmatrix}0 & 0\\ I_q & 0\\ 0 & I_p\end{bmatrix}+\begin{bmatrix}0\\ \mathcal{J}\mu(u)\\ \mathcal{J}\lambda(u)\end{bmatrix}\right\},\end{aligned}$$

which implies

$$
\begin{aligned}
& c\begin{bmatrix} I_q & 0 \\ 0 & I_p \end{bmatrix} + \begin{bmatrix} \mathcal{J}\mu(u) \\ \mathcal{J}\lambda(u) \end{bmatrix} \\
&= -c^{-1}\begin{bmatrix} 0 & I_q & 0 \\ 0 & 0 & I_p \end{bmatrix} \overline{\mathbb{K}}_c\big(x(u),\mu(u),\lambda(u)\big)^{-1} \begin{bmatrix} I & 0 & 0 \\ 0 & I_q & 0 \\ 0 & 0 & \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big) \end{bmatrix} \\
&\quad \times \left\{ c\begin{bmatrix} 0 & 0 \\ I_q & 0 \\ 0 & I_p \end{bmatrix} + \begin{bmatrix} 0 \\ \mathcal{J}\mu(u) \\ \mathcal{J}\lambda(u) \end{bmatrix} \right\} + c\begin{bmatrix} I_q & 0 \\ 0 & I_p \end{bmatrix}. \qquad (4.11)
\end{aligned}
$$

It follows from [8, p. 20] that the inverse of $\overline{\mathbb{K}}_c(x(u),\mu(u),\lambda(u))$ can be expressed as

$$
\begin{aligned}
& \overline{\mathbb{K}}_c\big(x(u),\mu(u),\lambda(u)\big)^{-1} \\
&= -\begin{bmatrix} 0 & 0 & 0 \\ 0 & cI_q & 0 \\ 0 & 0 & cI_p \end{bmatrix} + \begin{bmatrix} -I \\ -c\mathcal{J}h\big(x(u)\big) \\ -c\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\mathcal{J}g\big(x(u)\big) \end{bmatrix} \\
&\quad \times \left(\overline{\mathbb{K}}_c\big(x(u),\mu(u),\lambda(u)\big) \Big/ \begin{pmatrix} -c^{-1}I_q & 0 \\ 0 & -c^{-1}I_p \end{pmatrix}\right)^{-1} \begin{bmatrix} -I \\ -c\mathcal{J}h\big(x(u)\big) \\ -c\mathcal{J}g\big(x(u)\big) \end{bmatrix}^\top .
\end{aligned}
$$

It is easy to check

$$
\overline{\mathbb{K}}_c\big(x(u),\mu(u),\lambda(u)\big) \Big/ \begin{pmatrix} -c^{-1}I_q & 0 \\ 0 & -c^{-1}I_p \end{pmatrix} = \nabla^2_{xx}L_c\big(x(u),\mu(u),\lambda(u)\big),
$$

which implies

$$
\begin{aligned}
& \overline{\mathbb{K}}_c\big(x(u),\mu(u),\lambda(u)\big)^{-1} \\
&= \begin{bmatrix} W & cW\begin{bmatrix} \mathcal{J}h\big(x(u)\big) \\ \mathcal{J}g\big(x(u)\big) \end{bmatrix}^\top \\ c\begin{bmatrix} \mathcal{J}h\big(x(u)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\mathcal{J}g\big(x(u)\big) \end{bmatrix} W & \Delta_c(u) \end{bmatrix}, \qquad (4.12)
\end{aligned}
$$

where

$$
\begin{aligned}
& \Delta_c(u) = -cI + c^2\begin{bmatrix} \mathcal{J}h\big(x(u)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\mathcal{J}g\big(x(u)\big) \end{bmatrix} W \begin{bmatrix} \mathcal{J}h\big(x(u)\big)^\top & \mathcal{J}g\big(x(u)\big)^\top \end{bmatrix}, \\
& W = \nabla^2_{xx}L_c^{-1}\big(x(u),\mu(u),\lambda(u)\big).
\end{aligned}
$$

Therefore, we have from (4.11) and (4.12) that

$$
\begin{aligned}
c\begin{bmatrix} I_q & 0 \\ 0 & I_p \end{bmatrix} + \begin{bmatrix} \mathcal{J}\mu(u) \\ \mathcal{J}\lambda(u) \end{bmatrix} = \Bigg\{ & \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{c}\big(I_p - \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\big) \end{bmatrix} + \begin{bmatrix} \mathcal{J}h\big(x(u)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\mathcal{J}g\big(x(u)\big) \end{bmatrix} \\
& \times\nabla^2_{xx}L_c\big(x(u),\mu(u),\lambda(u)\big)^{-1} \begin{bmatrix} \mathcal{J}h\big(x(u)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(u)\big)\mathcal{J}g\big(x(u)\big) \end{bmatrix}^\top \Bigg\}^{-1},
\end{aligned}
$$

namely, the equality (4.7) holds. □

Corollary 4.1. *Let Jacobian uniqueness conditions be satisfied at* $(\overline{x}, \overline{\mu}, \overline{\lambda})$. *Then*

$$\nabla^2 p(0) = -c\begin{bmatrix} I_q & 0 \\ 0 & I_p \end{bmatrix} + \left\{ \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{c}\mathcal{J}\Pi_{\mathbb{R}^p_-}(Z_c^*) \end{bmatrix} + \begin{bmatrix} \mathcal{J}h(\overline{x}) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}(Z_c^*)\mathcal{J}g(\overline{x}) \end{bmatrix} \times \nabla^2_{xx} L_c(\overline{x}, \overline{\mu}, \overline{\lambda})^{-1} \begin{bmatrix} \mathcal{J}h(\overline{x}) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}(Z_c^*)\mathcal{J}g(\overline{x}) \end{bmatrix}^\top \right\}^{-1}, \tag{4.13}$$

where $Z_c^* = Z_c(0) = \overline{\lambda} + cg(\overline{x})$.

Proof. The equality (4.7) is satisfied for all u with $\|u\| < \delta$ and all c large enough. For $u = 0$, we obtain

$$c\begin{bmatrix} I_q & 0 \\ 0 & I_p \end{bmatrix} + \begin{bmatrix} \mathcal{J}\mu(0) \\ \mathcal{J}\lambda(0) \end{bmatrix} = \left\{ \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{c}\big(I_p - \mathcal{J}\Pi_{\mathbb{R}^p_+}(Z_c(0))\big) \end{bmatrix} + \begin{bmatrix} \mathcal{J}h(\overline{x}) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(0)\big)\mathcal{J}g(\overline{x}) \end{bmatrix} \times \nabla^2_{xx} L_c(\overline{x}, \overline{\mu}, \overline{\lambda})^{-1} \begin{bmatrix} \mathcal{J}h(\overline{x}) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(0)\big)\mathcal{J}g(\overline{x}) \end{bmatrix}^\top \right\}^{-1},$$

which implies (4.13) from (4.6). □

Now we are in a position, by using the above properties, to analyze the rate of convergence of multipliers generated by the augmented Lagrangian method. First of all, we give an equivalent expression for

$$\begin{pmatrix} \widetilde{\mu}(\mu, \lambda, c) - \overline{\mu} \\ \widetilde{\lambda}(\mu, \lambda, c) - \overline{\lambda} \end{pmatrix},$$

which is a key property for achieving the asymptotical superlinear rate of convergence of multipliers.

Theorem 4.1. *Let Jacobian uniqueness conditions be satisfied at* $(\overline{x}, \overline{\mu}, \overline{\lambda})$. *Let* $c^* > 0$, δ *and* ε *be given by Theorem* 3.1. *Then for all* $(\mu, \lambda, c) \in D(c^*, \delta)$,

$$\begin{pmatrix} \widetilde{\mu}(\mu, \lambda, c) - \overline{\mu} \\ \widetilde{\lambda}(\mu, \lambda, c) - \overline{\lambda} \end{pmatrix} = \int_0^1 \Upsilon_c\big(\overline{\mu} + s(\mu - \overline{\mu}), \overline{\lambda} + s(\lambda - \overline{\lambda})\big) \begin{pmatrix} \mu - \overline{\mu} \\ Y - \overline{\lambda} \end{pmatrix} \mathrm{d}s, \tag{4.14}$$

where $\Upsilon_c(\mu, \lambda)$ *is defined by*

$$\Upsilon_c(\mu, \lambda) = \begin{bmatrix} I_q & 0 \\ 0 & \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(\mu, \lambda)\big) \end{bmatrix} - c\begin{bmatrix} \mathcal{J}h\big(x(\mu, \lambda, c)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(\mu, \lambda)\big)\mathcal{J}g\big(x(\mu, \lambda, c)\big) \end{bmatrix} \times \nabla^2_{xx} L_c\big(x(\mu, \lambda, c), \widetilde{\mu}(\mu, \lambda, c), \widetilde{\lambda}(\mu, \lambda, c)\big)^{-1} \begin{bmatrix} \mathcal{J}h\big(x(\mu, \lambda, c)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_c(\mu, \lambda)\big)\mathcal{J}g\big(x(\mu, \lambda, c)\big) \end{bmatrix}^\top, \tag{4.15}$$

and $Z_c(\mu, \lambda) = cg(x(\mu, \lambda, c)) + \lambda$.

Proof. Define

$$F_o(x, \widetilde{\mu}, \widetilde{\lambda}; \eta, \xi, t) = \begin{bmatrix} \nabla f(x) + \mathcal{J}h(x)^\top \widetilde{\mu} + \mathcal{J}g(x)^\top \widetilde{\lambda} \\ h(x) + \eta + t\overline{\mu} - t\widetilde{\mu} \\ \Pi_{\mathbb{R}^p_+}\big(g(x) + \xi + t\overline{\lambda}\big) - t\widetilde{\lambda} \end{bmatrix}.$$

Noting that $F_o(x,\widetilde{\mu},\widetilde{\lambda};\eta,\xi,t)=0$ is equivalent to $F(x,\widetilde{\mu},\widetilde{\lambda};\eta,\xi,t)$=0, we have

$$F_o\big(\widehat{x}(\eta,\xi,t),\widetilde{\mu}(\eta,\xi,t),\widetilde{\lambda}(\eta,\xi,t);\eta,\xi,t\big)=0. \tag{4.16}$$

Differentiating the last three equations in (4.16) with respect to (η,ξ,t), we obtain

$$\begin{aligned}&\mathcal{J}_{(x,\widetilde{\mu},\widetilde{\lambda})}F_o\big(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t);\eta,\xi,t\big)\mathcal{J}_{(\eta,\xi,t)}\begin{bmatrix}\widehat{x}(\eta,\xi,t)\\ \widehat{\mu}(\eta,\xi,t)\\ \widehat{\lambda}(\eta,\xi,t)\end{bmatrix}\\ &=\begin{bmatrix}0 & 0 & 0\\ -I_q & 0 & \widehat{\mu}(\eta,\xi,t)-\overline{\mu}\\ 0 & -\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(g(\widehat{x}(\eta,\xi,t))+\xi+t\overline{\lambda}\big) & \widehat{\lambda}-\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(g\big(\widehat{x}(\eta,\xi,t)\big)+\xi+t\overline{\lambda}\big)\overline{\lambda}\end{bmatrix}.\end{aligned} \tag{4.17}$$

Denoting

$$\begin{aligned}A(\eta,\xi,t)&=\mathcal{J}_{(x,\widetilde{\mu},\widetilde{\lambda})}F_o\big(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t),t\big),\\ Z_o(\eta,\xi,t)&=g\big(\widehat{x}(\eta,\xi,t)\big)+\xi+t\overline{\lambda},\end{aligned}$$

we have from (4.17) that

$$\begin{bmatrix}\mathcal{J}_{\eta,\xi}\widehat{\mu}(\eta,\xi,t)\\ \mathcal{J}_{\eta,\xi}\widehat{\lambda}(\eta,\xi,t)\end{bmatrix}=\begin{pmatrix}0 & I_q & 0\\ 0 & 0 & I_p\end{pmatrix}A(\eta,\xi,t)^{-1}\begin{bmatrix}0 & 0\\ -I_q & 0\\ 0 & -\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_o(\eta,\xi,t)\big)\end{bmatrix}. \tag{4.18}$$

We can easily obtain the following expression of $A(\eta,\xi,t)$:

$$A(\eta,\xi,t)=\begin{bmatrix}\nabla^2_{xx}\mathcal{L}\big(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t)\big) & \mathcal{J}h\big(\widehat{x}(\eta,\xi,t)\big)^\top & \mathcal{J}g\big(\widehat{x}(\eta,\xi,t)\big)^\top\\ \mathcal{J}h\big(\widehat{x}(\eta,\xi,t)\big) & -tI_q & 0\\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_o(\eta,\xi,t)\big)\mathcal{J}g\big(\widehat{x}(\eta,\xi,t)\big) & 0 & -tI_p\end{bmatrix}.$$

From the equality

$$\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_o(\eta,\xi,t)\big)=\mathcal{J}\Pi_{\mathbb{R}^p_+}\big(t^{-1}g\big(x(\eta,u,t)\big)+\lambda\big),$$

that

$$A(\eta,\xi,t)^{-1}=\begin{bmatrix}\nabla^2_{xx}L_{t^{-1}}\big(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t)\big)^{-1} & \Theta_{12}(\eta,\xi,t)\\ \Theta_{21}(\eta,\xi,t) & \Theta_{22}(\eta,\xi,t)\end{bmatrix} \tag{4.19}$$

with

$$\begin{aligned}\Theta_{12}(\eta,\xi,t)&=-t^{-1}\nabla^2_{xx}L_{t^{-1}}\big(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t)\big)^{-1}\\ &\quad\times\Big[\mathcal{J}h\big(\widehat{x}(\eta,\xi,t)\big)^\top\quad \mathcal{J}g\big(\widehat{x}(\eta,\xi,t)\big)^\top\Big],\\ \Theta_{21}(\eta,\xi,t)&=-t^{-1}\begin{bmatrix}\mathcal{J}h\big(x(\eta,\xi,t)\big)\\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_o(\eta,\xi,t)\big)\mathcal{J}g\big(x(\eta,\xi,t)\big)\end{bmatrix}\\ &\quad\times\nabla^2_{xx}L_{t^{-1}}\big(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t)\big)^{-1},\end{aligned}$$

$$\Theta_{22}(\eta,\xi,t) = -t^{-1}I + t^{-2}\begin{bmatrix} \mathcal{J}h\big(x(\eta,\xi,t)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_o(\eta,\xi,t)\big)\mathcal{J}g\big(x(\eta,\xi,t)\big)\end{bmatrix}$$
$$\times \nabla^2_{xx}L_{t^{-1}}\big(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t)\big)^{-1}\begin{bmatrix}\mathcal{J}h\big(x(\eta,\xi,t)\big) \\ \mathcal{J}g\big(x(\eta,\xi,t)\big)\end{bmatrix}^\top.$$

Thus, we have from (4.18) that

$$\begin{bmatrix}\mathcal{J}_{\eta,\xi}\widehat{\mu}(\eta,\xi,t) \\ \mathcal{J}_{\eta,\xi}\widehat{\lambda}(\eta,\xi,t)\end{bmatrix} = \Theta_{22}(\eta,\xi,t)\begin{bmatrix} I_q & 0 \\ 0 & \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_o(\eta,\xi,t)\big)\end{bmatrix}, \tag{4.20}$$

which implies

$$\begin{bmatrix}\mathcal{J}_{\eta,\xi}\widehat{\mu}(\eta,\xi,t) \\ \mathcal{J}_{\eta,\xi}\widehat{\lambda}(\eta,\xi,t)\end{bmatrix} = t^{-1}\begin{bmatrix} I_q & 0 \\ 0 & \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_o(\eta,\xi,t)\big)\end{bmatrix}$$
$$- t^{-2}\begin{bmatrix} \mathcal{J}h\big(x(\eta,\xi,t)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_o(\eta,\xi,t)\big)\mathcal{J}g\big(x(\eta,\xi,t)\big)\end{bmatrix}$$
$$\times \nabla^2_{xx}L_{t^{-1}}\big(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t)\big)^{-1}$$
$$\times \begin{bmatrix} \mathcal{J}h\big(x(\eta,\xi,t)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_o(\eta,\xi,t)\big)\mathcal{J}g\big(x(\eta,\xi,t)\big)\end{bmatrix}^\top.$$

Then we get

$$\begin{pmatrix}\widehat{\mu}(\eta,\xi,t)-\overline{\mu} \\ \widehat{\lambda}(\eta,\xi,t)-\overline{\lambda}\end{pmatrix} - \begin{pmatrix}\widehat{\mu}(\eta,\xi,t)-\mu(0,0,t) \\ \widehat{\lambda}(\eta,\xi,t)-\lambda(0,0,t)\end{pmatrix}$$
$$= \widehat{\mu}(t,\gamma)-\widehat{\mu}(0,\gamma) = \int_0^1 \begin{bmatrix}\mathcal{J}_{\eta,\xi}\widehat{\mu}(s\eta,sU,t) \\ \mathcal{J}_{\eta,\xi}\widehat{\lambda}(s\eta,sU,t)\end{bmatrix}\begin{pmatrix}\eta \\ \xi\end{pmatrix}\mathrm{d}s$$
$$= \int_0^1 \Bigg\{ t^{-1}\begin{bmatrix} I_q & 0 \\ 0 & \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_o(\eta,\xi,t)\big)\end{bmatrix} - t^{-2}\begin{bmatrix} \mathcal{J}h\big(x(\eta,\xi,t)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_o(\eta,\xi,t)\big)\mathcal{J}g\big(x(\eta,\xi,t)\big)\end{bmatrix}$$
$$\times \nabla^2_{xx}L_{t^{-1}}\big(\widehat{x}(\eta,\xi,t),\widehat{\mu}(\eta,\xi,t),\widehat{\lambda}(\eta,\xi,t)\big)^{-1}$$
$$\times \begin{bmatrix} \mathcal{J}h\big(x(\eta,\xi,t)\big) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(Z_o(\eta,\xi,t)\big)\mathcal{J}g\big(x(\eta,\xi,t)\big)\end{bmatrix}^\top \Bigg\}\begin{pmatrix}\eta \\ \xi\end{pmatrix}\mathrm{d}s.$$

Substituting $t = 1/c, \eta = (\mu-\overline{\mu})/c, \eta = (\lambda-\overline{\lambda})/c, x(\mu,\lambda,c) = \widehat{x}(\eta,\xi,t)$, and

$$\widetilde{\mu}(\mu,\lambda,c) = \widehat{\mu}(\eta,\xi,t) = \mu - ch\big(\widehat{x}(\eta,\xi,t)\big),$$
$$\widetilde{\lambda}(\mu,\lambda,c) = \widehat{\lambda}(\eta,\xi,t) = \Pi_{\mathbb{R}^p_+}\big(\lambda + cg\big(\widehat{x}(\eta,\xi,t)\big)\big).$$

We obtain the desired result. □

Theorem 4.2. *Assume that $(\overline{x},\overline{\mu},\overline{\lambda}) \in \mathbb{R}^n\times\mathbb{R}^q\times\mathbb{R}^p$ satisfies the Jacobian uniqueness conditions, $c^*>0$, δ and ε are given by Theorem* 3.1. *Assume that*

$$\overline{c}^* > \max\left\{c^*,\ 2\max_{1\le i\le q+p}\left|\lambda_i\big(\nabla^2 p(0)\big)\right|\right\}, \tag{4.21}$$

where $\lambda_i(\nabla^2 p(0))$ denotes the i-th eigenvalue of $\nabla^2 p(0)$. Then there must exist a scalar $\overline{\delta}_1 \in (0, \delta]$ such that if $\{c_k\}$ and (μ^0, λ^0) satisfy

$$\frac{1}{c_0}\left\|\begin{pmatrix}\mu^0-\overline{\mu}\\ \lambda^0-\overline{\lambda}\end{pmatrix}\right\| < \overline{\delta}_1, \quad \overline{c}^* \le c_k \le c_{k+1}, \quad \forall\, k = 0, 1, 2, \ldots, \tag{4.22}$$

then the sequence $\{(\mu^k, \lambda^k)\}$ generated by

$$\begin{aligned} \mu^{k+1} &= \mu^k + c_k h\big(x\big(\mu^k, \lambda^k, c_k\big), \lambda\big(\mu^k, \lambda^k, c_k\big)\big), \\ \lambda^{k+1} &= \Pi_{\mathbb{R}^p_+}\big(\lambda^k + c_k g\big(x\big(\mu^k, \lambda^k, c_k\big), \lambda\big(\mu^k, \lambda^k, c_k\big)\big)\big) \end{aligned} \tag{4.23}$$

is well-defined, and $(\mu^k, \lambda^k) \to (\overline{\mu}, \overline{\lambda})$ and $(x(\mu^k, c_k), \lambda(\mu^k, c_k)) \to (\overline{x}, \overline{\lambda})$. Furthermore if

$$\limsup_{k\to\infty} c_k = c_* < \infty$$

and $(\mu^k, \lambda^k) \neq (\overline{\mu}, \overline{\lambda})$ for all k, then

$$\limsup_{k\to\infty} \frac{\|(\mu^{k+1}, \lambda^{k+1}) - (\overline{\mu}, \overline{\lambda})\|}{\|(\mu^k, \lambda^k) - (\overline{\mu}, \overline{\lambda})\|} \le \max_{1\le i\le q+p}\left|\frac{\lambda_i\big(\nabla^2 p(0)\big)}{c_* + \lambda_i\big(\nabla^2 p(0)\big)}\right|, \tag{4.24}$$

while if $c_k \to \infty$ and $(\mu^k, \lambda^k) \neq (\overline{\mu}, \overline{\lambda})$ for all k, then

$$\lim_{k\to\infty} \frac{\|(\mu^{k+1}, \lambda^{k+1}) - (\overline{\mu}, \overline{\lambda})\|}{\|(\mu^k, \lambda^k) - (\overline{\mu}, \overline{\lambda})\|} = 0. \tag{4.25}$$

Proof. In view of Υ_c of (4.15), we have that

$$\begin{aligned} \Upsilon_c(\overline{\mu}, \overline{\lambda}) = &\begin{bmatrix} I_q & 0 \\ 0 & \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(cg(\overline{x}) + \overline{\lambda}\big)\end{bmatrix} \\ &- c\begin{bmatrix}\mathcal{J}h(\overline{x}) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(cg(\overline{x}) + \overline{\lambda}\big)\mathcal{J}g(\overline{x})\end{bmatrix} \nabla^2_{xx} L_c\left(\overline{x}, \overline{\mu}, \overline{\lambda}\right)^{-1} \begin{bmatrix}\mathcal{J}h(\overline{x}) \\ \mathcal{J}\Pi_{\mathbb{R}^p_+}\big(cg(\overline{x}) + \overline{\lambda}\big)\mathcal{J}g(\overline{x})\end{bmatrix}^{\top}. \end{aligned}$$

Using (4.13), we obtain

$$\Upsilon_c(\overline{\mu}, \overline{\lambda}) = I - c\big(\nabla^2 p(0) + cI\big)^{-1}$$

and thus, for the i-th eigenvalue $\lambda_i(\Upsilon_c)$), one has

$$\lambda_i\big(\Upsilon_c(\overline{\mu}, \overline{\lambda})\big) = \frac{\lambda_i\big(\nabla^2 p(0)\big)}{c + \lambda_i\big(\nabla^2 p(0)\big)},$$

where $\lambda_i(\nabla^2 p(0))$ denotes the i-th eigenvalue of $\nabla^2 p(0)$. It is easy to see that for any $\varepsilon_1 > 0$, there exists $\overline{\delta}_1 \in (0, \delta]$ such that, for all (μ, λ, c) with $\|(\mu - \overline{\mu}, \lambda - \overline{\lambda})\|/c < \overline{\delta}_1, c \ge c^*$, we have

$$\begin{aligned} \|\Upsilon_c(\mu, \lambda)\| \le \|\Upsilon_c(\overline{\mu}, \overline{\lambda})\| + \varepsilon_1 &= \max_{1\le i\le q+p}\big|\lambda_i\big(\Upsilon_c(\overline{\mu}, \overline{\lambda})\big)\big| + \varepsilon_1 \\ &= \max_{1\le i\le q+p}\left|\frac{\lambda_i\big(\nabla^2 p(0)\big)}{c + \lambda_i\big(\nabla^2 p(0)\big)}\right| + \varepsilon_1, \end{aligned}$$

where $\|\cdot\|$ denotes the spectral norm of the operator. Using (4.14), we obtain for all (μ, λ, c) chosen as the above,

$$\left\|\begin{pmatrix}\widetilde{\mu}(\mu, \lambda, c) - \overline{\mu} \\ \widetilde{\lambda}(\mu, \lambda, c) - \overline{\lambda}\end{pmatrix}\right\| \le \left(\max_{1\le i\le q+p}\left|\frac{\lambda_i(\nabla^2 p(0))}{c + \lambda_i\big(\nabla^2 p(0)\big)}\right| + \varepsilon_1\right)\left\|\begin{pmatrix}\mu - \overline{\mu} \\ \lambda - \overline{\lambda}\end{pmatrix}\right\|. \tag{4.26}$$

From (4.21) and (4.22), we have

$$\max_{1\le i\le q+p}\left|\frac{\lambda_i\big(\nabla^2 p(0)\big)}{c+\lambda_i\big(\nabla^2 p(0)\big)}\right|<1.$$

Therefore by choosing ε_1 small enough, we have that there exists $\varrho_1\in(0,1)$ such that

$$\left\|\begin{pmatrix}\widetilde{\mu}(\mu,\lambda,c)-\overline{\mu}\\ \widetilde{\lambda}(\mu,\lambda,c)-\overline{\lambda}\end{pmatrix}\right\|\le\varrho_1\left\|\begin{pmatrix}\mu-\overline{\mu}\\ \lambda-\overline{\lambda}\end{pmatrix}\right\|$$

for (μ,λ,c) with $\|(\mu,\lambda)-(\overline{\mu},\overline{\lambda})\|/c<\delta_1, c\ge\overline{c}^*$. From this, (3.2) and (4.22), we obtain that $\mu^k\to\overline{\mu}$ and $(x(\mu^k,c_k),\lambda(\mu^k,c_k))\to(\overline{x},\overline{\lambda})$. The estimates (4.24) and (4.25) for the rate of convergence can be easily obtained from (4.26). □

It follows from (4.24) and (4.26) that the sequence of the multipliers has at least Q-linear convergence if $\{c_k\}$ is bounded and the convergence is superlinear if c_k is increasing to ∞.

Acknowledgements. The work was supported by the National Key R&D Program of China (Project No. 2022YFA1004000) and by the National Natural Science Foundation of China (Grant Nos. 12201097 and 12071055).

References

[1] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, 1982.

[2] A.R. Conn, N.I.M. Gould and P.L. Toint, A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds, *SIAM J. Numer. Anal.*, **28** (1991), 545–572.

[3] L. Contesse-Becker, Extended convergence results for the method of multipliers for non-strictly binding inequality constraints, *J. Optim. Theory Appl.*, **79** (1993), 273–310.

[4] G. Debreu, Definite and semidefinite quadratic forms, *Econometrica*, **20** (1952), 295–300.

[5] M.R. Hestenes, Multiplier and gradient methods, *J. Optim. Theory Appl.*, **4** (1969), 303–320.

[6] K. Ito and K. Kunisch, The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces, *Math. Program.*, **46** (1990), 341–360.

[7] M.J.D. Powell, A method for nonlinear constraints in minimization problems, in: *Optimization*, Academic Press, (1969), 283–298.

[8] S. Puntanen and G.P.H. Styan, Historical introduction: Issai Schur and the early development of the Schur complement, in: *The Schur Complement and Its Application. Numerical Methods and Algorithms*, Vol. 4, Springer, (2005), 1–16.

[9] R.T. Rockafellar, A dual approach to solving nonlinear programming problems by unconstrained optimization, *Math. Program.*, **5** (1973), 354–373.

[10] R.T. Rockafellar, The multiplier method of Hestenes and Powell applied to convex programming, *J. Optim. Theory Appl.*, **12** (1973), 555–562.

[11] R.T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Math. Oper. Res.*, **1** (1976), 97–116.

[12] D.F. Sun, J. Sun, and L.W. Zhang, The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming, *Math. Program.*, **114** (2008), 349–391.