Hindawi Publishing Corporation
Abstract and Applied Analysis
Volume 2014, Article ID 210284, 7 pages
http://dx.doi.org/10.1155/2014/210284

Research Article

Second-Order Multiplier Iteration Based on a Class of Nonlinear Lagrangians

Yong-Hong Ren[1,2]

[1] *School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China*
[2] *School of Mathematics, Liaoning Normal University, Dalian 116029, China*

Correspondence should be addressed to Yong-Hong Ren; ryhong@163.com

Received 24 December 2013; Accepted 11 March 2014; Published 9 April 2014

Academic Editor: Gaohang Yu

Nonlinear Lagrangian algorithm plays an important role in solving constrained optimization problems. It is known that, under appropriate conditions, the sequence generated by the first-order multiplier iteration converges superlinearly. This paper aims at analyzing the second-order multiplier iteration based on a class of nonlinear Lagrangians for solving nonlinear programming problems with inequality constraints. It is suggested that the sequence generated by the second-order multiplier iteration converges superlinearly with order at least two if in addition the Hessians of functions involved in problem are Lipschitz continuous.

1. Introduction

Lagrangians play an important role for solving constrained optimization problems. Hestenes [1] and Powell [2] introduced the proximal augmented Lagrangian for problems with equality constraints and Rockafellar [3] developed the proximal augmented Lagrangian for problems with both equality and inequality constraints.

Based on the above Lagrangians, Bertsekas [4, 5] discussed the convergence of sequence generated by the second-order multiplier iteration. The same author further improved the convergence and convergent rate of the second-order multiplier iteration using Newton's method in 1982. Besides, Brusch [6] and Fletcher [7] first independently proposed the second-order multiplier iteration using quasi-Newton's method, respectively. Bertsekas [8] developed new framework of quasi-Newton's method in 1982.

Consider the following inequality constrained optimization problem:

$$\begin{aligned} &\text{minimize} \quad f_0(x) \\ &\text{subject to} \quad f_i(x) \geq 0, \quad i = 1, \ldots, m, \end{aligned} \tag{INP}$$

where $f_i : \mathbb{R}^2 \to \mathbb{R}$, $i = 0, \ldots, m$ are continuous differentiable functions.

As nonlinear Lagrangians can be used to develop dual algorithms for nonlinear programming, requiring no restrictions on primal feasibility, important contributions on this topic have been done by many authors.

Polyak and Teboulle [9] discussed a class of Lagrange functions of the form

$$H(x, u, c) = f_0(x) - c\sum_{i=1}^{m} u_i \psi\left(c^{-1} f_i(x)\right) \tag{1}$$

for solving (INP), where $c > 0$ is penalty parameter and ψ is twice continuous differentiable function. Furthermore, Polyak and Griva [10] proposed a general primal-dual nonlinear rescaling (PDNR) method for convex optimization with inequality constraints, and Griva and Polyak [11] developed a general primal-dual nonlinear rescaling method with dynamic scaling parameter update. Besides the works by Polyak and his coauthors, Auslender et al. [12] and Ben-Tal and Zibulevsky [13] studied other nonlinear Lagrangians and obtained interesting convergence results for convex programming problems, too. Under appropriate conditions, the sequence generated by the first-order multiplier iteration converges superlinearly.

Ren and Zhang [14] analysed the following nonlinear Lagrangians:

$$H(x,u,c) = f_0(x) - k^{-1}\sum_{i=1}^{m} u_i \psi(k f_i(x)) \tag{2}$$

and constructed the dual algorithm based on minimizing $H(x,u,k)$ as follows.

D-Algorithm

Step 1. Given $k > 0$ large enough, $\varepsilon \geq 0$ small enough, $u^0 \in \mathbb{R}^m_{++}$, and $x^0 \in \mathbb{R}^n$, set $s = 0$.

Step 2. Solve (approximately)

$$\text{minimize} \quad H(x, u^s, k) \tag{3}$$

and obtain its (approximate) solution x^s.

Step 3. If $u_i^s f_i(x^s) \leq \varepsilon$, $i = 1,\ldots,m$, stop; otherwise go to Step 4.

Step 4. Update Lagrange multiplier u

$$u^{s+1} = \operatorname*{diag}_{1\leq i\leq m}\left(\psi'(k f_i(x^s))\right) u^s. \tag{4}$$

Step 5. Set $s = s+1$ and return to Step 2.

It was shown that, under a set of conditions, dual algorithm based on this class of Lagrange is locally convergent when the penalty parameter is larger than a threshold.

In view of interpretation of the multiplier iteration as the steepest ascent method, it is natural to consider Newton's method for maximizing the dual functional. Using known results for Newton's method, we expect that the second-order iteration will yield a vector u^{s+1} which is closer to u^* than u^s. This paper aims at discussing the second-order multiplier iteration based on nonlinear Lagrangians of the form (2). It is suggested that the sequence generated by the second-order multiplier iteration converges superlinearly with order at least two if $\nabla^2 f_i(x)$ $(i = 0,\ldots,m)$ are Lipschitz continuous.

We introduce the following notation to end this section:

$$\begin{aligned}
&\nabla f(x) = \left(\nabla f_1(x),\ldots,\nabla f_m(x)\right),\\
&u^* = \left(u_1^*,\ldots,u_m^*\right) \in \mathbb{R}^m,\\
&u^*_{(m-r)} = \left(u_{r+1}^*,\ldots,u_m^*\right) \in \mathbb{R}^{m-r},\\
&S(y,\varepsilon) = \left\{x \in \mathbb{R}^n : \|x-y\| \leq \varepsilon\right\}.\\
&\nabla f_{(r)}(x) = \left(\nabla f_1(x),\ldots,\nabla f_r(x)\right),\\
&u^*_{(r)} = \left(u_1^*,\ldots,u_r^*\right) \in \mathbb{R}^r,\\
&\|x\| = \|x\|_\infty = \max_{1\leq i\leq n}|x_i|.
\end{aligned} \tag{5}$$

2. Preliminaries

Consider the inequality constrained optimization problem (INP). Let

$$L(x,u) = f_0(x) - \sum_{i=1}^{m} u_i f_i(x) \tag{6}$$

denote the Lagrange function for problem (INP) and $I(x) = \{i \mid f_i(x) = 0,\ i = 1,\ldots,m\}$.

For the convenience of description in the sequel, we list the following assumptions, some of which will be used somewhere.

(a) Functions $f_i(x)$ $(i = 0,\ldots,m)$ are twice continuously differentiable.

(b) For convenience of statement, we assume $I(x^*) = \{i \mid f_i(x^*) = 0, i = 1,\ldots,m\} = \{1,\ldots,r\}$.

(c) Let $(x^*,u^*) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfy the Kuhn-Tucker conditions

$$\begin{aligned}
&\nabla_x L(x^*,u^*) = 0, \quad u^* \geq 0,\\
&u_i^* f_i(x^*) = 0, \quad i = 1,\ldots,m.
\end{aligned} \tag{7}$$

(d) Strict complementary condition holds; that is,

$$u_i^* > 0 \quad \text{for } i \in I(x^*). \tag{8}$$

(e) The set of vectors $\{\nabla f_i(x^*) \mid i \in I(x^*)\}$ are linearly independent.

(f) For all $y \neq 0$ satisfying $\nabla f_i(x^*)^T y = 0$, $i \in I(x^*)$, the following inequality holds:

$$y^T \nabla_x^2 L(x^*,u^*)\, y > 0. \tag{9}$$

Let function ψ in $H(x,u,k)$ defined in (2) and its derivatives satisfy the following conditions:

(H1) $\psi(0) = 0$;

(H2) $\psi'(t) > 0$, for all $t \in (b,+\infty)$, with $-\infty \leq b < 0$, and $\psi'(0) = 1$;

(H3) $\psi''(t) < 0$, for all $t \in (b,+\infty)$, with $-\infty \leq b < 0$;

(H4) $k\psi'(kt)$ is bounded, where $t \in (b,+\infty)$, with $-\infty \leq b < 0$, and for $k > 0$ large enough.

The following proposition concerns properties of $H(x,u,k)$ at a Kuhn-Tucker point (x^*,u^*).

Proposition 1 (see [14]). *Assume that (a)–(f) and (H1)–(H3) hold. For any $k > 0$ and any Kuhn-Tucker point (x^*,u^*) the following properties are valid:*

(i) $H(x^*,u^*,k) = L(x^*,u^*) = f(x^*)$;

(ii) $\nabla_x H(x^*,u^*,k) = \nabla_x L(x^*,u^*) = \nabla f(x^*) - \sum_{i=1}^m u_i^* \nabla f_i(x^*) = 0$;

(iii) $\nabla_x^2 H(x^*,u^*,k) = \nabla_x^2 L(x^*,u^*) - k\psi''(0)\nabla f_{(r)}(x^*)^T U^* \nabla f_{(r)}(x^*)$, *where* $U^* = \operatorname{diag}_{1\leq i\leq r}(u_i^*)$;

(iv) *there exist* $c_0 > 0$ *and* $\mu > 0$ *such that, for any* $c > c_0$,

$$\left\langle \nabla_x^2 H(x^*, u^*, k)\, y, y \right\rangle \geq \mu \langle y, y \rangle, \quad \forall y \in \mathbb{R}^n \quad \text{satisfying } \nabla f_{(r)}(x^*)^T y = 0. \tag{10}$$

Let $\delta > 0$ be small enough, $0 < \varepsilon < \min\{u_i^* \mid i = 1, \ldots, r\}$, and k_0 large enough satisfying (iv) of Proposition 1. For any fixed $k > k_0$, define

$$U_k^i(\varepsilon, \delta) = \begin{cases} \{u_i \mid \max\{\varepsilon, u_i^* - \delta k\} \\ \quad \leq u_i \leq u_i^* + \delta k\}, & i = 1, \ldots, r, \\ \{u_i \mid 0 \leq u_i \leq \delta k\}, & i = r+1, \ldots, m, \end{cases} \tag{11}$$

$$U_k(\varepsilon, \delta) = U_k^1(\varepsilon, \delta) \times \ldots \times U_k^r(\varepsilon, \delta) \times \ldots \times U_k^m(\varepsilon, \delta).$$

For any $k_1 > k_0$, we denote

$$D(\varepsilon, \delta) = \{(u, k) \mid u \in U_k(\varepsilon, \delta), k \in [k_0, k_1]\}. \tag{12}$$

Let $\sigma = \min\{f_i(x^*) \mid r+1 \leq i \leq m\} > 0$, I_r is the $r \times r$ identity matrix, and 0_r is the $r \times r$ zero matrix.

Theorem 2 (see [14]). *Assume that (a)–(f) and (H1)–(H4) hold. Then there exists* $k_0 > 0$ *large enough such that, for any* $k_1 > k_0$, *there exist* $\varepsilon_1 > 0, \delta > 0$, *satisfying that for any* $(u, k) \in D(\varepsilon, \delta)$, *the following statements hold.*

(i) *There exists a vector*

$$\hat{x} = \hat{x}(u, k) \in \arg\min\{H(x, u, k)\, x \in S(x^*, \varepsilon_1)\}. \tag{13}$$

(ii) *For* $\hat{x}$ *in (i) and* $\hat{u} = \hat{u}(u, k) = \operatorname{diag}_{1 \leq i \leq m}(\psi'(kf_i(\hat{x})))u$, *the following estimate is valid:*

$$\max\{\|\hat{x} - x^*\|, \|\hat{u} - u^*\|\} \leq ck^{-1}\|u - u^*\|, \tag{14}$$

where $c > 0$ *is a scalar independent of* k_0 *and* k_1.

(iii) *Function* $H(x, u, k)$ *is strongly convex in a neighborhood of* $\hat{x}$.

3. The Second-Order Multiplier Iteration

Based on the nonlinear Lagrange function $H(x, u, k)$, we consider the dual function defined on $S(x^*, \varepsilon_1) \times \mathbb{R}_+^m$ as follows:

$$d_k(u) = \inf\{H(x, u, k) \mid x \in S(x^*, \varepsilon_1)\} - \delta(u \mid U_k(\varepsilon, \delta)), \tag{15}$$

where $\delta(u \mid U_k(\varepsilon, \delta)) = \begin{cases} 0 & \text{if } u \in U_k(\varepsilon, \delta) \\ +\infty & \text{if } u \notin U_k(\varepsilon, \delta) \end{cases}$ is the indicator function of $U_k(\varepsilon, \delta)$.

Lemma 3. *Assume that conditions (a)–(f) and (H1)–(H4) hold; then for any fixed* $k \geq k_0$ *function* $d_k(u)$ *is twice continuously differentiable and concave on* $U_k(\varepsilon, \delta)$.

Proof. Obviously, for $k > 0$, function $d_k(u)$ is concave. In view of Theorem 2, for any $(u, k) \in D(\varepsilon, \delta)$, function $H(x, u, k)$ is strong convex in the neighborhood of $\hat{x} = \hat{x}(u, k)$. So $\hat{x}(u, k)$ is unique minimizer of function $H(x, u, k)$ with respect to x in the neighborhood of point $\hat{x}$, and $d_k(u) = H(\hat{x}(u, k), u, k)$ is smooth in $U_k(\varepsilon, \delta)$; that is, the Jacobian of $d_k(u)$ exists, and

$$\begin{aligned} \nabla_u d_k(u) &= \nabla_u \hat{x}(u, k)\, \nabla_x H(\hat{x}(u, k), u, k) \\ &\quad + \nabla_u H(\hat{x}(u, k), u, k) \\ &= \left(\nabla_{u_1} d_k(u), \ldots, \nabla_{u_m} d_k(u)\right)^T. \end{aligned} \tag{16}$$

For $(u, k) \in D(\varepsilon, \delta)$, matrix $\nabla_x^2 H(x, u, k)$ is positive definite, and system $\nabla_x H(x, u, k) = 0_n$ generates unique vector-valued function $\hat{x}(u, k)$ satisfying $\hat{x}(u^*, k) = x^*$ and

$$\nabla_u \hat{x}(u, k) = -\nabla_{xu}^2 H(\hat{x}(u, k), u, k)\left(\nabla_x^2 H(\hat{x}(u, k), u, k)\right)^{-1}, \quad \forall (u, k) \in D(\varepsilon, \delta). \tag{17}$$

In view of $\nabla_x H(\hat{x}(u, k), u, k) = 0_n$, we have

$$\begin{aligned} \nabla_u d_k(u) &= \nabla_u H(\hat{x}(u, k), u, k) \\ &= -k^{-1}\left(\psi(kf_1(\hat{x}(u, k))), \ldots, \psi(kf_m(\hat{x}(u, k)))\right)^T. \end{aligned} \tag{18}$$

It follows from (18) that

$$\begin{aligned} &\nabla_{ux}^2 H(\hat{x}(u, k), u, k) \\ &\quad = -\left(\psi'(kf_1(\hat{x}(u, k)))\nabla f_1(\hat{x}(u, k)), \ldots, \right. \\ &\qquad \left. \psi'(kf_m(\hat{x}(u, k)))\nabla f_m(\hat{x}(u, k))\right) \\ &\quad = -\nabla f(\hat{x}(u, k))\, \psi'(kf(\hat{x}(u, k))), \end{aligned} \tag{19}$$

which means

$$\begin{aligned} \nabla_{xu}^2 H(\hat{x}(u, k), u, k) &= \left(\nabla_{ux}^2 H(\hat{x}(u, k), u, k)\right)^T \\ &= -\psi'(kf(\hat{x}(u, k)))\, \nabla f(\hat{x}(u, k))^T. \end{aligned} \tag{20}$$

Thus,

$$\begin{aligned} \nabla_u^2 d_k(u) &= \nabla_u x(u, k)\, \nabla_{ux}^2 H(\hat{x}(u, k), u, k) \\ &= -\nabla_{xu}^2 H(\hat{x}(u, k), u, k)\left(\nabla_x^2 H(\hat{x}(u, k), u, k)\right)^{-1} \\ &\quad \times \nabla_{ux}^2 H(\hat{x}(u, k), u, k) \\ &= -\psi'(kf(\hat{x}(u, k)))\left(\nabla f(\hat{x}(u, k))\right)^T \\ &\quad \times \left(\nabla_x^2 H(\hat{x}(u, k), u, k)\right)^{-1} \\ &\quad \times \left(\nabla f(\hat{x}(u, k))\right)\psi'(kf(\hat{x}(u, k))). \end{aligned} \tag{21}$$

So,

$$\nabla_u^2 d_k(u^*) = -\psi'(kf(x^*))(\nabla f(x^*))^T\left(\nabla_x^2 H(x^*,u^*,k)\right)^{-1} \times (\nabla f(x^*))\psi'(kf(x^*)). \tag{22}$$

□

Let $\hat{x}(u,k)$ be the minimizer of $H(x,u,k)$ in a neighborhood of x^*; then we obtain that

$$\begin{aligned}\nabla_u d_k(u) &= \nabla_u H(\hat{x}(u,k),u,k)\\ &= -k^{-1}(\psi(kf_1(\hat{x}(u,k))),\ldots,\psi(kf_m(\hat{x}(u,k))))^T,\\ \nabla_u^2 d_k(u) &= -\psi'(kf(\hat{x}(u,k)))\nabla f(\hat{x}(u,k))^T\\ &\quad\times\left(\nabla_x^2 H(\hat{x}(u,k),u,k)\right)^{-1}\\ &\quad\times\nabla f(\hat{x}(u,k))\psi'(kf(\hat{x}(u,k))).\end{aligned} \tag{23}$$

In view of the interpretation of the multiplier iteration as the steepest ascent method, it is natural to consider Newton's method for maximizing the dual functional d_k which is given by

$$u^{s+1} = u^s - \left[\nabla^2 d_k(u^s)\right]^{-1}\nabla d_k(u^s). \tag{24}$$

In view of (23), this iteration can be written as

$$u^{s+1} = u^s - B_k^{-1}k^{-1}\psi(kf(x(u^s,k))), \tag{25}$$

where

$$\begin{aligned}B_k &= \psi'(kf(x(u^s,k)))\nabla f(x(u^s,k))^T\\ &\quad\times\left[\nabla_x^2 H(x(u^s,k),u^s,k)\right]^{-1}\\ &\quad\times\nabla f(x(u^s,k))\psi'(kf(x(u^s,k))).\end{aligned} \tag{26}$$

We will provide a convergence and rate of convergence result for iteration (25) and (26).

For $k>0$ and $(x,u)\in\mathbb{R}^{n+m}$, we define

$$\begin{aligned}A^+(x,u) &= \{i \mid u_i\psi(kf_i(x)) > 0,\ i=1,\ldots,m\}\\ A^-(x,u) &= \{i \mid i\notin A^+(x,u),\ i=1,\ldots,m\}.\end{aligned} \tag{27}$$

For a given (x,u), assume (by reordering indices if necessary) that $A^+(x,u)$ contains the first r indices where r is an integer with $0\le r\le m$. Define

$$\begin{aligned}\psi_+(kf(x)) &= \begin{pmatrix}\psi(kf_1(x))\\ \vdots\\ \psi(kf_r(x))\end{pmatrix}\\ \psi_-(kf(x)) &= \begin{pmatrix}\psi(kf_{r+1}(x))\\ \vdots\\ \psi(kf_m(x))\end{pmatrix}\\ u_+ &= (u_1,\ldots,u_r)^T,\qquad u_- = (u_{r+1},\ldots,u_m)^T,\\ H_+(x,u,k) &= f_0(x) - k^{-1}u_+^T\psi_+(kf(x)).\end{aligned} \tag{28}$$

We note that r,ψ_+,ψ_-,u_+,u_- and H_+ depend on (x,u), but to simplify notation we do not show explicitly this dependence. Now, we consider Newton's method for solving the system of necessary conditions

$$\begin{aligned}&\nabla_x H_+(x,u,k) = \nabla f_0(x) - \sum_{i=1}^r u_i\psi'(kf_i(x))\nabla f_i(x) = 0,\\ &k^{-1}\psi(kf_i(x)) = 0,\quad i=1,\ldots,r.\end{aligned} \tag{29}$$

Considering the extension of Newton's method, given (x,u), we denote the next iterate by $(\hat{x},\hat{u})$ where $\hat{u} = (\hat{u}_1,\ldots,\hat{u}_m)^T$. We also write

$$\hat{u}_+ = (\hat{u}_1,\ldots,\hat{u}_r)^T,\qquad \hat{u}_- = (\hat{u}_{r+1},\ldots,\hat{u}_m)^T. \tag{30}$$

The iteration, roughly speaking, consists of setting the multipliers of the inactive constraints ($j\in A^-(x,u)$) to zero and treating the remaining constraints as equalities. More precisely, we set $\hat{u}_- = 0_{m-r}$ and obtain $\hat{x},\hat{u}_+$ by solving the system

$$\begin{pmatrix}\nabla_x^2 H_+(x,u,k) & -\nabla f_{(r)}(x)\psi_+'(kf(x))\\ \psi_+'(kf(x))\nabla f_{(r)}(x)^T & 0\end{pmatrix}\begin{pmatrix}\hat{x}-x\\ \hat{u}_+-u_+\end{pmatrix} = \begin{pmatrix}-\nabla_x H_+(x,u,k)\\ -k^{-1}\psi_+(kf(x))\end{pmatrix}, \tag{31}$$

where $\psi_+'(kf(x)) = [\operatorname{diag}\psi'(kf_i(x))]_{i=1}^r$.

If $\nabla_x^2 H_+(x,u,k)$ is invertible and $\nabla f_{(r)}(x)$ has rank r, we can solve system (31) explicitly. It follows from (31) that

$$\nabla_x^2 H_+(x,u,k)(\hat{x}-x) - \nabla f_{(r)}(x)\psi_+'(kf(x))(\hat{u}_+-u_+) = -\nabla_x H_+(x,u,k), \tag{32}$$

$$\psi_+'(kf(x))\nabla f_{(r)}(x)^T(\hat{x}-x) = -k^{-1}\psi_+(kf(x)). \tag{33}$$

Premultiplying (32) with $\psi_+'(kf(x))\nabla f_{(r)}(x)^T[\nabla_x^2 H_+(x,u,k)]^{-1}$ and using (33), we obtain

$$\hat{x}-x = \left[\nabla_x^2 H_+(x,u,k)\right]^{-1}\left\{\nabla f_{(r)}(x)\psi_+'(kf(x))(\hat{u}_+-u_+) - \nabla_x H_+(x,u,c)\right\}, \tag{34}$$

from which, we have

$$\hat{u}_+ = u_+ - \left\{\psi_+'(kf(x))\nabla f_{(r)}(x)^T\left[\nabla_x^2 H_+(x,u,k)\right]^{-1} \times\nabla f_{(r)}(x)\psi_+'(kf(x))\right\}^{-1}k^{-1}\psi_+(kf(x)). \tag{35}$$

Substitution in (32) yields

$$\hat{x} = x - \nabla_x^2 H_+(x,u,k)\nabla_x H_+(x,\hat{u},k). \tag{36}$$

Return to (25) and (26), and using the fact that $\nabla_x H_+(x(u,k),u,k) = 0$, we see that iteration (25) and (26) is of the form (35).

For a triple (x,u,k) for which the matrix on the left-hand side of (31) is invertible, we denote by $\widehat{x}(x,u,k)$, $\widehat{u}_+(x,u,k)$ the unique solution of (31) and say that $\widehat{x}(x,u,k)$, $\widehat{u}_+(x,u,c)$ are well defined.

Define

$$u_+^{s+1} = \widehat{u}_+\left(x\left(u^s,k\right),u^s,k\right), \qquad u_-^{s+1} = 0. \tag{37}$$

Proposition 4. *Let k be a scalar. For every triple (x,u,k), if ψ' satisfies*

$$\psi_+'^2 - 2\psi_+' + I = 0, \tag{38}$$

then the vectors $\widehat{x}(x,u,k)$, $\widehat{u}_+(x,u,k)$ are well defined if and only if the vectors $\widehat{x}(x,\psi'(kf(x))u,0)$, $\widehat{u}_+(x,\psi'(kf(x))u,0)$ are well defined.

Furthermore,

$$\widehat{x}(x,u,k) = \widehat{x}\left(x,\psi'\left(kf(x)\right)u,0\right), \tag{39}$$

$$\widehat{u}_+(x,u,k) = \widehat{u}_+\left(x,\psi'\left(kf(x)\right)u,0\right). \tag{40}$$

Proof. By calculating, we have

$$\begin{aligned}\nabla_x H_+(x,u,k) &= \nabla_x L\left(x,\psi'\left(kf(x)\right)u\right),\\ \nabla_x^2 H_+(x,u,k) &= \nabla_x^2 L\left(x,\psi'\left(kf(x)\right)u\right)\\ &\quad - k\sum_{i=1}^{r} u_i\psi''\left(kf_i(x)\right)\nabla f_i(x)\nabla f_i(x)^T.\end{aligned} \tag{41}$$

As a result, the system (31) can be written as

$$\begin{aligned}&\begin{pmatrix}\nabla_x^2 L\left(x,\psi'\left(kf(x)\right)u\right) - k\sum_{i=1}^{r} u_i\psi''\left(kf_i(x)\right)\nabla f_i(x)\nabla f_i(x)^T & -\nabla f_{(r)}(x)\psi_+'\left(kf(x)\right)\\ \psi_+'\left(kf(x)\right)\nabla f_{(r)}(x)^T & 0\end{pmatrix}\times\begin{pmatrix}\widehat{x}-x\\ \widehat{u}_+-u_+\end{pmatrix}\\ &\quad = \begin{pmatrix}-\nabla_x L\left(x,\psi'\left(kf(x)\right)u\right)\\ -k^{-1}\psi_+\left(kf(x)\right)\end{pmatrix}.\end{aligned} \tag{42}$$

The second equation yields

$$\psi_+'\left(kf(x)\right)\nabla f_{(r)}(x)^T\left(\widehat{x}-x\right) = -k^{-1}\psi_+\left(kf(x)\right). \tag{43}$$

If we form the second-order Taylor series expansion of ψ around t_k,

$$\psi(t) = \psi\left(t_k\right) + \psi'\left(t_k\right)\left(t-t_k\right) + \frac{1}{2}\left(t-t_k\right)^T\psi''\left(t_k\right)\left(t-t_k\right), \tag{44}$$

we obtain

$$\psi'(t) = \psi'\left(t_k\right) + \psi''\left(t_k\right)\left(t-t_k\right). \tag{45}$$

Take $t = kf_i(\widehat{x})$, $t_k = kf_i(x)$, $i = 1,\ldots,r$, and it follows that

$$\psi'\left(kf_i(x)\right) = 1 - k(\widehat{x}-x)^T\nabla f_i(x)\psi''\left(kf_i(x)\right), \quad i = 1,\ldots,r. \tag{46}$$

Substituting (46) into (43), we have

$$\begin{aligned}&\operatorname*{diag}_{1\le i\le r}\left(1 - k(\widehat{x}-x)^T\nabla f_i(x)\psi''\left(kf_i(x)\right)\right)\\ &\quad\times\nabla f_{(r)}(x)^T\left(\widehat{x}-x\right)\\ &\quad = -k^{-1}\psi_+\left(kf(x)\right)\end{aligned} \tag{47}$$

which, when substituted into the first equation in (42), yields

$$\begin{aligned}&\nabla_x^2 L\left(x,\psi'\left(kf(x)\right)u\right)\left(\widehat{x}-x\right) - \nabla f_{(r)}(x)\psi_+'\left(kf(x)\right)\widehat{u}_+\\ &\quad + 2\nabla f_{(r)}(x)\psi_+'\left(kf(x)\right)u_+ - \nabla f_{(r)}(x)u_+\\ &\quad = -L\left(x,\psi'\left(kf(x)\right)u\right).\end{aligned} \tag{48}$$

Thus, in view of condition $\psi_+'^2 - 2\psi_+' + I = 0$, system (42) is equivalent to

$$\begin{aligned}&\begin{pmatrix}\nabla_x^2 L\left(x,\psi'\left(kf(x)\right)u\right) & -\nabla f_{(r)}(x)\psi_+'\left(kf(x)\right)\\ \psi_+'\left(kf(x)\right)\nabla f_{(r)}(x)^T & 0\end{pmatrix}\\ &\quad\times\begin{pmatrix}\widehat{x}-x\\ \widehat{u}_+ - \psi_+'\left(kf(x)\right)u_+\end{pmatrix}\\ &\quad = \begin{pmatrix}-\nabla_x L\left(x,\psi'\left(kf(x)\right)u\right)\\ -k^{-1}\psi_+\left(kf(x)\right)\end{pmatrix}.\end{aligned} \tag{49}$$

This shows (39) and (40). □

In view of (40), we can write (37) as

$$\begin{aligned}u_+^{s+1} &= \widehat{u}_+\left(x\left(u^s,k\right),\widetilde{u}\left(u^s,k\right),0\right),\\ u_-^{s+1} &= 0_{m-r},\end{aligned} \tag{50}$$

where

$$\widetilde{u}\left(u^s,k\right) = \psi'\left(kf\left(x\left(u^s,k\right)\right)\right)u^s. \tag{51}$$

This means that one can carry out the second-order multiplier iteration (25), (26) in two stages. First execute the first-order iteration (51) and then the second-order iteration (50), which is part of Newton's iteration at $(x(u^s,k)), \widetilde{u}(u^s,k)$ for solving the system of necessary conditions (29).

Now, we know that $x(u^s,k), \widetilde{u}(u^s,k)$ is close to (x^*,u^*) for (u^s,k) in an appropriate region of $\mathbb{R}^{m+1}$. Therefore, using known results for Newton's method, we expect that (50) will yield a vector u^{s+1} which is closer to u^* than u^s. This argument is the basis for the proof of the following proposition.

Proposition 5. *Assume (a)–(f) hold, and let $k_0 > 0$, $\delta > 0$ be as in Theorem 2. Then, given any scalar $\gamma > 0$, there exists a scalar δ_1 with $0 < \delta_1 \leq \delta$ such that for all $(u,k) \in D_1 = \{(u,k) : u \in U_k(\varepsilon,\delta_1), k \geq k_0\}$ there holds*

$$\left\|(\widehat{x}(u,k),\widehat{u}(u,k)) - (x^*,u^*)\right\| \leq \gamma k^{-1}\left\|u-u^*\right\|, \tag{52}$$

where

$$u^{s+1} = u^s - B_k^{-1}k^{-1}\psi\left(kf\left(x\left(u^s,k\right)\right)\right), \tag{53}$$

where

$$\begin{aligned} B_k = {} & \psi'\left(kf\left(x\left(u^s,k\right)\right)\right)\nabla f\left(x\left(u^s,k\right)\right)^T \\ & \times\left[\nabla_x^2 H\left(x\left(u^s,k\right),u^s,k\right)\right]^{-1}\nabla f\left(x\left(u^s,k\right)\right)\psi' \\ & \times\left(kf\left(x\left(u^s,k\right)\right)\right). \end{aligned} \tag{54}$$

If, in addition, $\nabla^2 f_i(x)$, $i = 0,\ldots,m$ are Lipschitz continuous in a neighborhood of x^*, there exists a scalar $\gamma_1 > 0$ such that, for all $(u,k) \in D_1$, there holds

$$\left\|(\widehat{x}(u,k),\widehat{u}(u,k)) - (x^*,u^*)\right\| \leq \gamma_1 k^{-2}\left\|u-u^*\right\|^2. \tag{55}$$

Proof. In view of Theorem 2, given any $\gamma > 0$, there exist $\varepsilon_1 > 0, \varepsilon_2 > 0$ and $M > 0$ such that if $x(u,k) \in S(x^*,\varepsilon_1)$ and $\widetilde{u}(u,k) \in S(u^*,\varepsilon_2)$, there holds

$$\begin{aligned} & \left\|(\widehat{x}(x(u,k),\widetilde{u}(u,k),0),\widehat{u}(x(u,k),\widetilde{u}(u,k),0)) - (x^*,u^*)\right\| \\ & \leq \frac{\gamma}{M}\left\|(x(u,k),\widetilde{u}(u,k)) - (x^*,u^*)\right\| \end{aligned} \tag{56}$$

(compare with Proposition 1.17, Bertsekas [8]). Take δ_1 sufficiently small so that, for all $(u,k) \in D_1$, we have $x(u,k) \in S(x^*,\varepsilon_1)$, $\widetilde{u}(u,k) \in S(u^*,\varepsilon_2)$, and

$$\left\|(x(u,k),\widetilde{u}(u,k)) - (x^*,u^*)\right\| \leq Mk^{-1}\left\|u-u^*\right\|. \tag{57}$$

From (50), we have

$$\begin{aligned} & \left\|(\widehat{x}(x(u,k),\widetilde{u}(u,k),0),\widehat{u}(x(u,k),\widetilde{u}(u,k),0)) - (x^*,u^*)\right\| \\ & \leq \frac{\gamma}{M}\cdot Mk^{-1}\left\|u-u^*\right\| = \gamma k^{-1}\left\|u-u^*\right\|. \end{aligned} \tag{58}$$

If $\nabla^2 f_i(x)$ $(i = 0,\ldots,m)$ are Lipschitz continuous, then there exists a $\gamma_1 > 0$ such that for $x(u,k) \in S(x^*,\varepsilon)$ and $\widetilde{u}(u,k) \in S(u^*,\varepsilon)$, we have

$$\begin{aligned} & \left\|(\widehat{x}(x(u,k),\widetilde{u}(u,k),0),\widehat{u}(x(u,k),\widetilde{u}(u,k),0)) - (x^*,u^*)\right\| \\ & \leq \frac{\gamma_1}{2M^2}\left\|(x(u,k),\widetilde{u}(u,k)) - (x^*,u^*)\right\|^2 \\ & \leq \frac{\gamma_1}{2M^2}\left(\left(Mk^{-1}\right)^2\left\|u-u^*\right\|^2 + \left(Mk^{-1}\right)^2\left\|u-u^*\right\|^2\right) \\ & = \gamma_1 k^{-2}\left\|u-u^*\right\|^2. \end{aligned} \tag{59}$$

□

From the above analysis, we know that the sequence generated by the second-order multiplier iteration converges superlinearly with order at least two if the Hessians of functions involved in problem are Lipschitz continuous.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This project is supported by the National Natural Science Foundation of China (Grant no. 11171138).

References

[1] M. R. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications*, vol. 4, no. 5, pp. 303–320, 1969.

[2] M. J. D. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*, pp. 283–298, Academic Press, New York, NY, USA, 1969.

[3] R. T. Rockafellar, "Augmented Lagrange multiplier functions and duality in nonconvex programming," *SIAM Journal on Control*, vol. 12, pp. 268–285, 1974.

[4] D. P. Bertsekas, "Multiplier methods: a survey," *Automatica*, vol. 12, no. 2, pp. 133–145, 1976.

[5] D. P. Bertsekas, "On the convergence properties of second-order multiplier methods," *Journal of Optimization Theory and Applications*, vol. 25, no. 3, pp. 443–449, 1978.

[6] R. B. Brusch, "A rapidly convergent methods for equality constrained function minimization," in *Proceedings of the IEEE Conference on Decision and Control*, pp. 80–81, San Diego, Calif, USA, 1973.

[7] R. Fletcher, "An ideal penalty function for constrained optimization," in *Nonlinear Programming 2*$''$, pp. 121–163, Academic Press, New York, NY, USA, 1975.

[8] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, NY, USA, 1982.

[9] R. A. Polyak and M. Teboulle, "Nonlinear rescaling and proximal-like methods in convex optimization," *Mathematical Programming, Series B*, vol. 76, no. 2, pp. 265–284, 1997.

[10] R. A. Polyak and I. Griva, "Primal-dual nonlinear rescaling method for convex optimization," *Journal of Optimization Theory and Applications*, vol. 122, no. 1, pp. 111–156, 2004.

[11] I. Griva and R. A. Polyak, "Primal-dual nonlinear rescaling method with dynamic scaling parameter update," *Mathematical Programming*, vol. 106, no. 2, pp. 237–259, 2006.

[12] A. Auslender, R. Cominetti, and M. Haddou, "Asymptotic analysis for penalty and barrier methods in convex and linear programming," *Mathematics of Operations Research*, vol. 22, no. 1, pp. 43–62, 1997.

[13] A. Ben-Tal and M. Zibulevsky, "Penalty/barrier multiplier methods for convex programming problems," *SIAM Journal on Optimization*, vol. 7, no. 2, pp. 347–366, 1997.

[14] Y.-H. Ren and L.-W. Zhang, "The dual algorithm based on a class of nonlinear Lagrangians for nonlinear programming," in *Proceedings of the 6th World Congress on Intelligent Control and Automation (WCICA '06)*, pp. 934–938, 2006.