A Smoothing Newton-Type Algorithm of Stronger Convergence for the Quadratically Constrained Convex Quadratic Programming

ZHENG-HAI HUANG*   zheng-haihuang@yahoo.com.cn
Department of Mathematics, School of Science, Tianjin University, Tianjin 300072, P.R. China

DEFENG SUN   matsundf@nus.edu.sg
GONGYUN ZHAO   matzgy@nus.edu.sg
Department of Mathematics, National University of Singapore, 2 Science Drive 2, Singapore 117543, Republic of Singapore

Abstract.   In this paper we propose a smoothing Newton-type algorithm for the problem of minimizing a convex quadratic function subject to finitely many convex quadratic inequality constraints. The algorithm is shown to converge globally and possess stronger local superlinear convergence. Preliminary numerical results are also reported.

Keywords: smoothing Newton method, global convergence, superlinear convergence

Mathematics Subject Classification (1991):   90C33, 65K10

1.   Introduction

The quadratically constrained quadratic programming (QCQP) is the following

$$
\begin{aligned}
\min \quad & f_0(\hat{x}) \\
\text{s.t.} \quad & f(\hat{x}) \leq 0,
\end{aligned}
\tag{1.1}
$$

where $f(\hat{x}) = (f_1(\hat{x}), \ldots, f_m(\hat{x}))^T$, $f_j(\hat{x}) = \frac{1}{2}\hat{x}^T P^j \hat{x} + (a^j)^T \hat{x} + b_j$, $\hat{x} \in \Re^n$, $a^j \in \Re^n$, $P^j \in \Re^{n \times n}$, $b_j \in \Re$ for all $j \in \mathcal{J}_0 := \{0, 1, \ldots, m\}$ (we reserve x for later use). In this paper, we are interested in (1.1) when all f_j, $j \in \mathcal{J}_0$ are convex functions, i.e., all P^j, $j \in \mathcal{J}_0$ are symmetric positive semidefinite matrices.

Interior point methods (IPMs) have been successfully applied to solve convex optimization problems and monotone complementarity problems [21, 36]. Besides the polynomial complexity, the superlinear convergence has always been an important topic in IPMs. Early superlinear convergence analysis of IPMs requires many conditions [39, 40]. Under the strictly complementary condition, Ye and Anstreicher [38] proved the quadratic convergence of a predictor-corrector IPM for solving the monotone linear complementarity problem (LCP). The strictly complementary condition has been successfully eliminated in proving the superlinear convergence of some predictor-corrector IPMs for LCPs [20, 31, 32, 41]. The local superlinear convergence analysis of the existing IPMs for solving nonlinear optimization problems depends on the strictly complementary condition [27, 28, 34].

Smoothing Newton-type methods are originally designed for solving nonsmooth equations arising from the mathematical programming field. So far, a number of globally and locally superlinearly convergent smoothing Newton-type methods have been proposed. For a comprehensive treatment on this topic, see [8, Chap. 11]. A key condition for the superlinear convergence of smoothing Newton-type methods is the nonsingularity of the generalized Jacobian of the function involved in the nonsmooth equations. For the P_0 function nonlinear complementarity problem (NCP), this condition implies that the solution set consists of a single element. Several authors have investigated ways to relax such a relatively restrictive condition in smoothing Newton-type methods for linear programming [4, 5] and LCPs [14, 35], and in the Levenberg-Marquardt method for the nonlinear equation [37].

In this paper, we will propose a smoothing Newton-type method for solving the convex QCQP (1.1). It should be noted that such a QCQP can be cast as a second-order-cone problem (SOCP) [17, 21]. Since the SOCP can be solved efficiently by using IPMs, one often reformulates the convex QCQP as the SOCP and then solves the corresponding SOCP. Such an approach may not be practical for those convex QCQPs with a huge number of constraints. For example, the subproblems of second-order methods for solving semi-infinite programming are convex QCQPs of a large number of constraints (see, [22, Chap. 3]). Here, we will reformulate the QCQP as a system of nonsmooth equations instead of an SOCP. The proposed algorithm is shown to possess the local convergence features of both smoothing Newton-type methods and interior point methods. Specifically, the superlinear convergence of the algorithm is obtained under either the nonsingularity condition or the strictly complementary condition. To some extent, this paper can be regarded as an extension of [14] for solving the LCP to the convex QCQP. However, due to the nonlinear structure of the QCQP, substantial differences exist and new techniques are needed.

The paper is organized as follows. In the next section, (1.1) is reformulated as a system of parameterized smooth equations. In Section 3, we propose the smoothing Newton-type algorithm and discuss its global convergence. In Sections 4 and 5, we investigate the local superlinear convergence of the proposed algorithm under the nonsingularity and the strictly complementary condition, respectively. Numerical results are reported in Section 6. Proofs of some technical lemmas are given in Appendix.

2.   Preliminaries

Throughout this paper, we use the following assumption:

Assumption 2.1.

 (i) P^j are symmetric positive semidefinite matrices for all $j \in \mathcal{J}_0$.
 (ii) (Slater Constraint Qualification) There is a point x^0 such that $f(x^0) < 0$.

   It is well known that, under Assumption 2.1, solving (1.1) is equivalent to solving the following system

$$\nabla f_0(\hat{x}) + f'(\hat{x})^T \hat{\lambda} = 0, \ \hat{\lambda} \geq 0, \ -f(\hat{x}) \geq 0, \ \hat{\lambda}^T f(\hat{x}) = 0, \tag{2.1}$$

where $\hat{\lambda} \in \Re^m$, $f'(\hat{x}) = (\nabla f_1(\hat{x}), \ldots, \nabla f_m(\hat{x}))^T$, and $\nabla f_j(\hat{x})$ is the gradient of $f_j(\hat{x})$ at $\hat{x} \in \Re^n$, $j \in \mathcal{J}_0$. Every solution of (2.1) is called a KKT point of problem (1.1). Let

$$\hat{w} := \begin{pmatrix} \hat{x} \\ \hat{\lambda} \end{pmatrix}, \ F(\hat{w}) := \begin{pmatrix} \nabla f_0(\hat{x}) + f'(\hat{x})^T \hat{\lambda} \\ -f(\hat{x}) \end{pmatrix},$$
$$K := \{(\hat{x}, \hat{\lambda}) \in \Re^{n+m} : \hat{x} \in \Re^n, \hat{\lambda} \in \Re_+^m\}.$$

Then, solving (2.1) is equivalent to finding a vector $\hat{w}^* \in K$ such that

$$(\hat{w} - \hat{w}^*)^T F(\hat{w}^*) \geq 0 \quad \text{for all } \hat{w} \in K. \tag{2.2}$$

Let w denote the vector $(x^T, \lambda^T)^T$ and $\Pi_K(w)$ denote the Euclidean projection of w onto K. It is well known that problem (2.2) is equivalent to the following normal equation

$$F(\Pi_K(w)) + w - \Pi_K(w) = 0 \tag{2.3}$$

in the sense that if $w^* \in \Re^{n+m}$ is a solution of (2.3), then $\hat{w}^* := \Pi_K(w^*)$ is a solution of (2.2), and conversely if $\hat{w}^*$ is a solution of (2.2), then $w^* := \hat{w}^* - F(\hat{w}^*)$ is a solution of (2.3) [16, 30]. Let $H_0(x, \lambda) := F(\Pi_K(w)) + w - \Pi_K(w)$. Then (2.3) becomes

$$H_0(x, \lambda) = \begin{pmatrix} \nabla f_0(x) + f'(x)^T \lambda_+ \\ -f(x) + \lambda - \lambda_+ \end{pmatrix} = 0, \tag{2.4}$$

where λ_+ is a vector whose i-th component is $\max\{0, \lambda_i\}$, $i \in \mathcal{J}$.
   Since λ_+ is not differentiable everywhere, the function H_0 is not differentiable. We now introduce the following smoothing function:

$$\Phi(\mu, \lambda) := \text{vec}\{\phi(\mu, \lambda_i) : i \in \mathcal{J}\}, \tag{2.5}$$

where for $u \in \Re^m$, vec$\{u_i : i \in \mathcal{J}\}$ denotes an m-vector whose i-th component is u_i; and $\phi : \Re^2 \to \Re$ is defined by

$$\phi(a, b) := \frac{1}{2}(b + \sqrt{b^2 + 4a^2}), \ (a, b) \in \Re \times \Re. \tag{2.6}$$

The function ϕ is continuously differentiable around any $(a, b) \in \Re \times \Re$ such that $(a, b) \neq 0$,

$$\phi'(a, b) = \left(\frac{2a}{\sqrt{b^2 + 4a^2}}, \frac{1}{2}\left(1 + \frac{b}{\sqrt{b^2 + 4a^2}}\right) \right). \tag{2.7}$$

Define $\bar{H} : \Re \times \Re^n \times \Re^m \to \Re \times \Re^n \times \Re^m$ by

$$\bar{H}(\mu, x, \lambda) := \begin{pmatrix} \mu \\ \nabla f_0(x) + f'(x)^T \Phi(\mu, \lambda) \\ -f(x) + \lambda - \Phi(\mu, \lambda) \end{pmatrix}, \ (\mu, x, \lambda) \in \Re \times \Re^n \times \Re^m.$$

Then, $(x^*, \lambda^*) \in \Re^n \times \Re^m$ solves (2.4) if and only if $(0, x^*, \lambda^*)$ solves $\bar{H}(\mu, x, \lambda) = 0$. Based on the function $\bar{H}(\cdot)$, we define the following smoothing function:

$$H(\mu, x, \lambda) := \begin{pmatrix} \mu \\ \nabla f_0(x) + f'(x)^T \Phi(\mu, \lambda) + g_1(\mu)x \\ -f(x) + \lambda - \Phi(\mu, \lambda) + g_2(\mu)\lambda + g_3(\mu)\pi(\mu, x, \lambda) \end{pmatrix}, \tag{2.8}$$

where $\pi(\mu, x, \lambda) := \text{vec}\{\pi_i(\mu, x, \lambda) : i \in \mathcal{J}\}$ with

$$\pi_i(\mu, x, \lambda) := \phi(\mu, \lambda_i)\phi(\mu, -f_i(x)), \quad i \in \mathcal{J}, \tag{2.9}$$

and for each $i \in \{1, 2, 3\}$, $g_i : \Re \to \Re_+$ is a twice continuously differentiable function satisfying

$$g_i(\mu) \geq 0 \quad \forall \mu, \ g_i(0) = 0, \ g_i(\mu) = O(\mu^2), \quad \text{and} \quad |g_i'(\mu)| = O(\mu). \tag{2.10}$$

When $\mu = 0$, $\pi_i(\mu, x, \lambda) = \pi_i(0, x, \lambda) = (\lambda_i)_+(-f_i(x))_+, i \in \mathcal{J}$. If $g_1(\mu) = g_2(\mu) = g_3(\mu) = 0$, then $H(\mu, x, \lambda) = \bar{H}(\mu, x, \lambda)$. The terms $g_1(\mu)x$ and $g_2(\mu)\lambda$ represent the regularized part for H and $g_3(\mu)\pi(\mu, x, \lambda)$ the smoothed penalized part. From both theoretical and practical points of view, we require

$$g_i(\mu) > 0, \quad \forall \mu \neq 0, \quad i = 1, 2, 3. \tag{2.11}$$

It is evident that $(x^*, \lambda^*) \in \Re^n \times \Re^m$ solves (2.4) if and only if $(0, x^*, \lambda^*)$ solves $H(\mu, x, \lambda) = 0$.

For any positive integer p, E^p denotes the $p \times p$ identity matrix. For any $u \in \Re^m$, let diag$\{u_i : i \in \mathcal{J}\}$ denote an $m \times m$ diagonal matrix whose i-th diagonal element is u_i.

Then, for any $z = (\mu, x, \lambda) \in \Re^{1+n+m}$ with $\mu \neq 0$, it follows from (2.8) that

$$H'(z) = \begin{bmatrix} 1 & 0 & 0 \\ f'(x)^T(\Phi(y))'_\mu + g'_1(\mu)x & M(z) & f'(x)^T(\Phi(y))'_\lambda \\ P(z) & R(z)[-f'(x)] & N(z) \end{bmatrix}, \quad (2.12)$$

where $(\Phi(y))'_\mu$ is the partial derivative of $\Phi(y) = \Phi(\mu, \lambda)$ with respect to μ, $(\Phi(y))'_\lambda$ is the partial derivative of $\Phi(y) = \Phi(\mu, \lambda)$ with respect to λ, and

$$M(z) := P^0 + \sum_{i \in \mathcal{J}} P^i \Phi_i(y) + g_1(\mu)E^n, \quad (2.13)$$

$$P(z) := -(\Phi(y))'_\mu + g'_2(\mu)\lambda + g'_3(\mu)\pi(z) + g_3(\mu)(\pi(z))'_\mu, \quad (2.14)$$

$$R(z) := E^m + Q(z), \quad \text{where} \quad Q(z) := \text{diag}\{q_i(z) : i \in \mathcal{J}\} \quad \text{with}$$

$$q_i(z) := \frac{1}{2}\left(1 - \frac{f_i(x)}{\sqrt{(-f_i(x))^2 + 4\mu^2}}\right)\Phi_i(y)g_3(\mu), \quad (2.15)$$

$$N(z) := (E^m - (\Phi(y))'_\lambda) + g_2(\mu)E^m$$
$$+ g_3(\mu)(\Phi(y))'_\lambda \text{diag}\{\phi(\mu, -f_i(x)) : i \in \mathcal{J}\}. \quad (2.16)$$

Proposition 2.1. *Under Assumption 2.1, for any $z = (\mu, x, \lambda) \in \Re^{1+n+m}$ with $\mu \neq 0$, the matrix $H'(z)$ is nonsingular.*

Proof. Let $dz := (d\mu, dx, d\lambda) \in \Re \times \Re^n \times \Re^m$ satisfy $H'(z)dz = 0$. Then from (2.12), it follows that $d\mu = 0$ and

$$M(z)dx + f'(x)^T(\Phi(y))'_\lambda d\lambda = 0, \quad (2.17)$$
$$R(z)[-f'(x)]dx + N(z)d\lambda = 0. \quad (2.18)$$

From (2.15) we know that $R(z)$ is a positive definite diagonal matrix. By multiplying Eq. (2.17) on the left side by $(dx)^T$, we have

$$(dx)^T M(z)dx + (f'(x)dx)^T(\Phi(y))'_\lambda d\lambda = 0,$$

which together with (2.18) implies

$$(dx)^T M(z)dx + (d\lambda)^T N(z)R(z)^{-1}(\Phi(y))'_\lambda d\lambda = 0. \quad (2.19)$$

From (2.7) we know that $(\Phi(y))'_\lambda$ is a positive definite diagonal matrix. In addition, since P^j are positive semidefinite for all $j \in \mathcal{J}_0$, $\Phi_i(y) > 0$ for all $i \in \mathcal{J}$, and $g_1(\mu) > 0$, it follows that $P^0 + \sum_{i \in \mathcal{J}} P^i \Phi_i(y) + g_1(\mu)E^n$ is positive definite. Similarly, we can obtain that $N(z)R(z)^{-1}(\Phi(y))'_\lambda$ is also positive definite. Thus, (2.19) implies that $dx = 0$ and $d\lambda = 0$. Therefore, the Jacobian matrix $H'(z)$ is nonsingular. $\square$

The following notation will be used in this paper. For any vectors $u, v \in \Re^n$, we write $(u^T, v^T)^T$ as (u, v) for simplicity. For any $\mathcal{K}, \mathcal{L} \subseteq \{1, \ldots, n\}$, we denote by $u_\mathcal{K}$

the vector obtained after removing from u those u_i with $i \notin \mathcal{K}$; and for any $A \in \Re^{n \times n}$ we denote by $A_{\mathcal{KL}}$ the submatrix of A obtained by removing all rows of A with indices outside of $\mathcal{K}$ and removing all columns of A with indices outside of $\mathcal{L}$, and denote by $A_{\mathcal{K}.}$ the submatrix of A obtained by removing all rows of A with indices outside of $\mathcal{K}$. For any $\alpha \in \Re_+$ and $\beta \in \Re_{++}$, we write $\alpha = O(\beta)$ (respectively, $\alpha = o(\beta)$) to mean α/β is uniformly bounded from above (respectively, tends to zero) as $\beta \to 0$, write $\alpha = O(1)$ to mean that there is a constant $\mathcal{C} > 0$ such that $\alpha \le \mathcal{C}$. Let $k \ge 0$ denote the iteration index, and write $\alpha = \Theta(\beta)$ to mean that there are two constants $\mathcal{C}_2 \ge \mathcal{C}_1 > 0$ such that $\mathcal{C}_1 \beta \le \alpha \le \mathcal{C}_2 \beta$. For any $(\mu, x, \lambda), (\mu_k, x^k, \lambda^k) \in \Re_+ \times \Re^n \times \Re^m$, we denote

$$w := (x, \lambda), \ w^k := (x^k, \lambda^k), \ y := (\mu, \lambda), \ y^k := (\mu_k, \lambda^k), \ z := (\mu, x, \lambda),$$
$$z^k := (\mu_k, x^k, \lambda^k).$$

Let p and q be two positive integers. The kernel or null space of a matrix $A \in \Re^{p \times q}$ is $\mathrm{Ker}A := \{d \in \Re^q : Ad = 0\}$, while the range space is denoted by $\mathrm{Ran}A := \{Ad : d \in \Re^q\}$.

3. Algorithm and its global convergence

Let $\bar{\mu}, \kappa \in (0, +\infty)$, $t_1 \in (0, 1]$, and $t_2, \tau, \gamma \in (0, 1)$. Define $\theta, \psi, \beta : \Re^{1+n+m} \to \Re_+$ by

$$\theta(z) := \|H(z)\|, \quad \psi(z) := \theta^{1+t_1}(z), \quad \text{and} \quad \beta(z) := \gamma \min\{1, \psi(z)\}, \tag{3.1}$$

respectively. For $z = (\mu, x, \lambda) \in \Re^{1+n+m}$ and $y = (\mu, \lambda)$ with $\mu \ne 0$, let

$$u_1(z) := f'(x)^T \hat{u}_1(z) + \tilde{u}_1(z) \quad \text{and} \quad u_2(z) := \hat{u}_2(z) + \tilde{u}_2(z), \tag{3.2}$$

where

$$\hat{u}_1(z) := (\Phi(y))'_\mu \mu_0 \beta(z) - \frac{1}{2}\mu(\Phi(y))'_\mu, \tag{3.3}$$

$$\hat{u}_2(z) := -(\Phi(y))'_\mu \mu_0 \beta(z) + \frac{1}{2}\mu(\Phi(y))'_\mu, \tag{3.4}$$

$$\tilde{u}_1(z) := g_1(\mu)x + g'_1(\mu)(-\mu + \bar{\mu}\beta(z))x, \tag{3.5}$$

$$\tilde{u}_2(z) := g_2(\mu)\lambda + g_3(\mu)\pi(z) + [g'_2(\mu)\lambda$$
$$+ g'_3(\mu)\pi(z) + g_3(\mu)(\pi(z))'_\mu](-\mu + \bar{\mu}\beta(z)). \tag{3.6}$$

Define $u : \Re^{1+n+m} \to \Re^{n+m}$ by

$$u(z) := \begin{cases} \begin{pmatrix} u_1(z) \\ u_2(z) \end{pmatrix} & \text{if } \mu \ne 0 \\ \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

and $v : \Re^{1+n+m} \to \Re^{n+m}$ by

$$v(z) := \begin{pmatrix} v_1(z) \\ v_2(z) \end{pmatrix} := \begin{cases} \tau \mu e & \text{if } \tau \mu \sqrt{n+m} \le \|u(z)\| \\ u(z) & \text{otherwise,} \end{cases} \tag{3.8}$$

where e denotes the $(n+m)$-vector of all ones. Define $\xi : \Re^m \to \Re_+$ by

$$\xi(\lambda) := \min_{i \in \mathcal{I}} |\lambda_i| \tag{3.9}$$

and $\Upsilon : \Re^{1+n+m} \to \Re^{1+n+m}$ by

$$\Upsilon(z) := \begin{pmatrix} \bar{\mu}\beta(z) \\ v_1(z) \\ v_2(z) \end{pmatrix} \quad \text{if} \quad \xi(\lambda) > \kappa\mu^{t_2}; \qquad \Upsilon(z) := \begin{pmatrix} \bar{\mu}\beta(z) \\ 0 \\ 0 \end{pmatrix} \quad \text{if} \quad \xi(\lambda) \le \kappa\mu^{t_2}. \tag{3.10}$$

Our smoothing Newton-type algorithm is now formally stated as follows.

Algorithm 3.1. *(A Smoothing Newton-Type Algorithm)*

Step 0. Choose $t_1 \in (0, 1]$, $t_2, \delta, \sigma \in (0, 1)$, and $\kappa, \bar{\mu} \in (0, \infty)$. Let $(x^0, \lambda^0) \in \Re^{n+m}$ be an arbitrary vector. Set $\mu_0 := \bar{\mu}$ and $z^0 := (\mu_0, x^0, \lambda^0)$. Choose $\gamma \in (0, 1)$ and $\tau \in (0, 1)$ such that $\gamma\mu_0 + \tau\sqrt{n+m} < 1$. Set $\eta := \gamma\mu_0 + \tau\sqrt{n+m}$ and $k := 0$.
Step 1. If $\theta(z^k) = 0$, stop.
Step 2. Compute $\Delta z^k := (\Delta\mu_k, \Delta x^k, \Delta\lambda^k) \in \Re^{1+n+m}$ by

$$H(z^k) + H'(z^k)\Delta z^k = \Upsilon(z^k). \tag{3.11}$$

Step 3. Let χ_k be the maximum of the values $1, \delta, \delta^2, \cdots$ such that

$$\theta(z^k + \chi_k\Delta z^k) \le [1 - \sigma(1 - \eta)\chi_k]\theta(z^k). \tag{3.12}$$

Step 4. Set $z^{k+1} := z^k + \chi_k\Delta z^k$ and $k := k + 1$. Go to Step 1.

The above algorithm is based on smoothing Newton methods in [26] for the NCP and box constrained variational inequality problem and in [14] for the P_0 and monotone LCP. The function $\Upsilon(\cdot)$ defined by (3.10) is quite different from those used in [14, 26]. This difference is vital in our local convergence analysis, which will be seen later.

Denote $\Omega := \{z = (\mu, x, \lambda) \in \Re \times \Re^n \times \Re^m : \mu \ge \mu_0\beta(z)\}$.

Lemma 3.1. *Let Assumption 2.1 be satisfied. Then*
 (i) *Algorithm 3.1 is well-defined.*
 (ii) *Algorithm 3.1 generates an infinite sequence $\{z^k\}$ with $\mu_k > 0$.*
 (iii) *$z^k \in \Omega$ for all $k \ge 0$.*

Proof: For the result (i), we need to show that Eq. (3.11) is solvable and line search (3.12) terminates finitely. The former holds from Proposition 2.1, and the latter can be proved similarly as in Lemma 2 of [14]. In addition, parts (ii) and (iii) can be obtained similarly as in Lemma 5 and Proposition 6 of [26]. We omit the details here. □

Lemma 3.2. *Let Assumption 2.1 be satisfied. Then Algorithm 3.1 generates an infinite iteration sequence $\{z^k\}$ with $\lim_{k\to\infty} \theta(z^k) = 0$. In particular, any accumulation point of $\{z^k\}$ is a solution of $H(z) = 0$.*

Proof: By using Lemma 3.1, we can prove this lemma similarly as in Theorem 4.1 of [33]. □

Lemma 3.2 shows that if $\{z^k\}$ has an accumulation point z^*, then z^* is a solution of $H(z) = 0$. This does not necessarily mean that there exists an accumulation point. In order to assure that $\{z^k\}$ has an accumulation point, we need the following sufficient condition.

Assumption 3.1. The solution set of (2.4) is nonempty and bounded.

Remark 3.1. (i) It should be noted that the Tikhonov-type regularization method for the monotone variational inequality problem can converge to a solution even if the solution set of the problem concerned is unbounded [1]. For problem (1.1), if we regularize objective function f_0 itself, we may show the global convergence of some regularization method without the boundedness of the solution set. In this paper, however, our main purpose is to improve the local convergence of the smoothing algorithms for the QCQP by using the norm map. It is difficult for us to unify such a better global convergence to the improved local convergence. (ii) In fact, Assumption 3.1 has been used extensively in regularized methods [7, 13, 24, 33]. It is known that Assumption 3.1 is weaker than those required by most existing smoothing (non-interior continuation) Newton-type methods [12]. For the monotone NCP, Assumption 3.1 is equivalent to say that the NCP has a strictly feasible solution [15, 18]. The latter has been used extensively in IPMs for the LP and the LCP.

Theorem 3.1. *Let Assumptions 2.1 and 3.1 be satisfied. Then the infinite sequence $\{z^k\}$ generated by Algorithm 3.1 is bounded and any accumulation point of $\{z^k\}$ is a solution of $H(z) = 0$.*

Proof: It is not difficult to show that the functions $H_0 : \Re^{n+m} \to \Re^{n+m}$ and $H : \Re^{1+n+m} \to \Re^{1+n+m}$ defined by (2.4) and (2.8), respectively are weakly univalent functions defined in [11]. Since Assumption 3.1 implies that the inverse image $H_0^{-1}(0)$ is nonempty and bounded, by using Theorem 2.5 in [29] we obtain that the sequence $\{z^k\}$ is bounded. Hence, by Lemma 3.2 , any accumulation point of $\{z^k\}$ is a solution of $H(z) = 0$. □

Let $z^* := (\mu_*, x^*, \lambda^*) \in \Re_+ \times \Re^n \times \Re^m$ be an accumulation point of the iteration sequence generated by Algorithm 3.1. Theorem 3.1 implies that $\mu_* = 0$ and $w^* :=$

(x^*, λ^*) is a solution of (2.4). Next, we consider the local convergence of Algorithm 3.1. The convergence analysis is divided into two parts and is discussed in the following two sections.

4. Superlinear convergence under nonsingularity

In this section, we consider the case that w^* satisfies a nonsingularity condition but may not satisfy the strictly complementary condition. In order to discuss the local superlinear convergence of the algorithm, we need the concept of semismoothness, which was originally introduced by Mifflin [19] for functionals and was extended the definition of semismoothness to vector valued functions by Qi and Sun [25].

Definition 4.1.   A locally Lipschitz function $F : \Re^{m_1} \to \Re^{m_2}$, which has the generalized Jacobian $\partial F(x)$ in the sense of Clarke [2], is said to be semismooth at $x \in \Re^{m_1}$, if

$$\lim_{\substack{V \in \partial F(x+th') \\ h' \to h, t \downarrow 0}} \{V h'\}$$

exists for any $h \in \Re^{m_1}$. F is said to be strongly semismooth at x if F is semismooth at x and for any $V \in \partial F(x+h)$, $h \to 0$, it follows that $F(x+h) - F(x) - Vh = O(\|h\|^2)$.

Remark 4.1.   Since $\phi(\cdot)$ is strongly semismooth at any $(a, b) \in \Re^2$, from [9] we know that the function $H(\cdot)$ defined by (2.8) is strongly semismooth everywhere.

The following lemma is key to show the main convergence result in this section.

Lemma 4.1.   *Let Assumptions 2.1 and 3.1 hold. Let $t_1 \in (0, 1]$ and $t_2 \in (0, 1)$ be given as in Algorithm 3.1, and the infinite sequence $\{z^k\}$ be generated by Algorithm 3.1. Then $\|\Upsilon(z^k)\| = O(\theta^\zeta(z^k))$ holds for all z^k sufficiently close to z^*, where $\zeta := \min\{1 + t_1, 2 - t_2\}$.*

Proof: From Lemma 3.2 we know that $\lim_{k \to \infty} \theta(z^k) = 0$. This, together with the definition of $\beta(z^k)$ (see (3.1)), implies that $\beta(z^k) = \gamma \psi(z^k)$ holds for all z^k sufficiently close to z^*. For each k, we have either $\xi(\lambda^k) \leq \kappa(\mu_k)^{t_2}$ or $\xi(\lambda^k) > \kappa(\mu_k)^{t_2}$, where the function $\xi(\cdot)$ is defined by (3.9). For the former case, we have that for all z^k sufficiently close to z^*,

$$\|\Upsilon(z^k)\| = \mu_0 \beta(z^k) = \mu_0 \gamma \psi(z^k) = O(\theta^\zeta(z^k)).$$

Hence, we only need to consider the latter case. From the definition of $\hat{u}_1(\cdot)$ (see (3.3)) it follows that for all z^k sufficiently close to z^*,

$$\|\hat{u}_1(z^k)\| \leq \|(\Phi(y^k))'_\mu\| \mu_0 \beta(z^k) + \frac{1}{2}\mu_k \|(\Phi(y^k))'_\mu\|$$

$$= \left\|\operatorname{vec}\left\{2\mu_k \Big/ \sqrt{(\lambda_i^k)^2 + 4(\mu_k)^2} : i \in \mathcal{J}\right\}\right\| \left(\mu_0 \beta(z^k) + \frac{1}{2}\mu_k\right)$$

$$\leq \left\| \mathrm{vec}\left\{2\mu_k \Big/ \sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2} : i \in \mathcal{J}\right\}\right\| \times \left(\mu_k + \frac{1}{2}\mu_k\right)$$

$$\leq 2\sqrt{n}(\mu_k)^{1-t_2} \Big/ \sqrt{\kappa^2 + (\mu_k)^{2-2t_2}} \times 3\mu_k/2$$

$$= O((\mu_k)^{2-t_2}),$$

where the first equality is due to (2.7), the second inequality is due to $z^k \in \Omega$ (by the result (iii) of Lemma 3.1), and the third inequality is due to the condition $\xi(\lambda^k) > \kappa(\mu_k)^{t_2}$; and from the definition of $\tilde{u}_1(\cdot)$ (see (3.5)) it follows that for all z^k sufficiently close to z^*,

$$\begin{aligned}
\|\tilde{u}_1(z^k)\| &\leq g_1(\mu)\|x^k\| + |g_1'(\mu_k)|(\mu_k + \bar\mu\beta(z^k))\|x^k\| \\
&\leq g_1(\mu)\|x^k\| + 2\mu_k|g_1'(\mu_k)|\|x^k\| \\
&= O((\mu_k)^2),
\end{aligned}$$

where the second inequality is due to $z^k \in \Omega$ and the last equality is due to (2.10), (2.11), and Theorem 3.1. Thus, by (3.2) we obtain that for all z^k sufficiently close to z^*,

$$\|u_1(z^k)\| \leq \|f'(x^k)\|\|\hat{u}_1(z^k)\| + \|\tilde{u}_1(z^k)\| = O((\mu_k)^{2-t_2}). \tag{4.1}$$

Similarly, by (3.2), (3.4), and (3.6) we obtain that for all z^k sufficiently close to z^*,

$$\|u_2(z^k)\| \leq \|\hat{u}_2(z^k)\| + \|\tilde{u}_2(z^k)\| = O((\mu_k)^{2-t_2}). \tag{4.2}$$

Furthermore, by combining (3.7) with (4.1) and (4.2), we have that for all z^k sufficiently close to z^*,

$$\|u(z^k)\| \leq \|u_1(z^k)\| + \|u_2(z^k)\| = O((\mu_k)^{2-t_2}). \tag{4.3}$$

Now, by the definition of the function $v(\cdot)$ (see (3.8)), it is easy to see that (4.3) implies that $v(z^k) = u(z^k)$ for all z^k sufficiently close to z^*. Hence, for all z^k sufficiently close to z^*,

$$\|\Upsilon(z^k)\| = \sqrt{(\bar\mu\beta(z^k))^2 + \|u(z^k)\|^2} = \sqrt{[\bar\mu\gamma\psi(z^k)]^2 + [O((\mu_k)^{2-t_2})]^2} = O(\theta^\zeta(z^k)).$$

The proof is completed.   □

Theorem 4.1. *Let Assumptions 2.1 and 3.1 be satisfied. Suppose that z^* is an accumulation point of the sequence $\{z^k\}$ generated by Algorithm 3.1. Let $\zeta := \min\{1+t_1, 2-t_2\}$. If all $V \in \partial H(z^*)$ are nonsingular, then the whole iteration sequence $\{z^k\}$ converges to z^*,*

$$\|z^{k+1} - z^*\| = O(\|z^k - z^*\|^\zeta), \quad \text{and} \quad \mu_{k+1} = O((\mu_k)^\zeta).$$

Proof: By using Lemma 4.1, we can prove this theorem in a similar way as Theorem 8 in [26]. We omit the details here.   □

In Theorem 4.1, we provide a superlinear convergence result for Algorithm 3.1 under the nonsingularity of $\partial H(z^*)$. The latter condition implies that the problem has a unique solution. In the discussion of next section, we will not assume the nonsingularity of $\partial H(z^*)$, but will assume a strictly complementary condition instead.

5. Superlinear convergence under strict complementarity

Let $\mathcal{S}$ denote the solution set of (2.4), i.e.,

$$\mathcal{S} = \{w := (x, \lambda) \in \Re^{n+m} : H_0(w) = 0\}.$$

5.1. Assumptions

Recall that $z^* := (\mu_*, x^*, \lambda^*) \in \Re_+ \times \Re^n \times \Re^m$ is an accumulation point of the iteration sequence generated by Algorithm 3.1, and $\mu_* = 0$ and $w^* := (x^*, \lambda^*)$ is a solution of (2.4).

Assumption 5.1.   Suppose that $\lambda_i^* \neq 0$ holds for all $i \in \mathcal{J}$.

It is not difficult to see from (2.2) and (2.4) that $\hat{w}^* = \Pi_K(w^*)$ is a strictly complementary solution to (2.2) is equivalent to $\lambda_i^* \neq 0$ for all $i \in \mathcal{J}$, where $w^* := (x^*, \lambda^*)$. In the sequel, let $\mathcal{B} := \{i \in \mathcal{J} : \lambda_i^* > 0\}$, $\mathcal{N} := \{i \in \mathcal{J} : \lambda_i^* < 0\}$, and $f_{\mathcal{B}}'(x) := (f'(x))_{\mathcal{B}}$. for all $x \in \Re^n$.

Let $\mathcal{S}^{x^*} := \{\lambda \in \Re^m : (x^*, \lambda) \in \mathcal{S}\}$. The two-side projection of a square matrix $Q \in \Re^{n \times n}$ onto the kernel of another matrix $\bar{Q} \in \Re^{p \times n}$ is any matrix of the form $X^T Q X$, where the column of X form a basis of $\mathrm{Ker}\bar{Q}$.

Assumption 5.2.   For each $\lambda \in \mathcal{S}^{x^*}$, the two-side projection of the matrix $P^0 + \sum_{i \in \mathcal{B}} \lambda_i P^i$ onto $\mathrm{Ker} f_{\mathcal{B}}'(x^*)$ is invertible.

Assumption 5.2 is an invertibility condition on the projection of the matrix $P^0 + \sum_{i \in \mathcal{B}} \lambda_i P^i$ onto the kernel of the active constraint Jacobian, which is essentially a second-order sufficient condition for optimality (see, for examples, [27, 28]).

Let $M(\cdot)$ and $R(\cdot)$ be defined by (2.13) and (2.15), respectively. Denote $M^0(z) := M(z) - g_1(\mu)E^n$. Let

$$A(z) := \begin{bmatrix} -M(z) & -f'(x)^T \\ R(z)_{\mathcal{B}\mathcal{B}} f_{\mathcal{B}}'(x) & 0 \\ 0 & -E_{\mathcal{N}\cdot}^m \end{bmatrix}, \quad B(z) := \begin{bmatrix} -M^0(z) & -f'(x)^T \\ f_{\mathcal{B}}'(x) & 0 \\ 0 & -E_{\mathcal{N}\cdot}^m \end{bmatrix},$$

$$C(z) := \begin{bmatrix} M(z)^T & -R(z)f'(x)^T \\ f_{\mathcal{B}}'(x) & 0 \\ 0 & -E_{\mathcal{N}\cdot}^m \end{bmatrix}. \tag{5.1}$$

For $\varepsilon > 0$, we define

$$\check{\mathcal{F}}^{\varepsilon,z^*} := \{z := (\mu, x, \lambda) \in \Re_+ \times \Re^n \times \Re^m : \|z - z^*\| \le \varepsilon\}.$$

Assumption 5.3. There exists a scalar $\varepsilon > 0$ such that for any $\check{u}, \check{v} \in \check{\mathcal{F}}^{\varepsilon,z^*}$, it holds that

$$\mathrm{Ran}A(\check{u}) = \mathrm{Ran}B(\check{u}) = \mathrm{Ran}C(\check{u}) = \mathrm{Ran}A(\check{v}) = \mathrm{Ran}B(\check{v}) = \mathrm{Ran}C(\check{v}).$$

It is noted that matrices $A(z)$, $B(z)$, and $C(z)$ have quite similar structures. A similar condition to Assumption 5.3 has been used in [34].

5.2. *An error bound result*

Lemma 5.1. *Suppose that Assumptions* 2.1, 5.1*, and* 5.2 *are satisfied. For $\varepsilon > 0$, let*

$$\mathcal{X}^{\varepsilon,*} := \{x \in \Re^n : (x, \lambda) \in \mathcal{S} \quad \text{and} \quad \|(x, \lambda) - (x^*, \lambda^*)\| \le \varepsilon\}.$$

Then (i) there exists $\varepsilon_0 > 0$ such that $\mathcal{X}^{\varepsilon,*} = \{x^*\}$ holds for all $\varepsilon \in (0, \varepsilon_0]$; and (ii) the solution set $\mathcal{S}$ of (2.4) *is convex and compact.*

Proof: By using the relations between the solutions of (2.1) and (2.4), we can obtain the results of this lemma directly from Lemmas 4.1 and 4.2 in [27]. $\square$

Let the function H_0 be defined by (2.4). Denote

$$\mathcal{S}_0 := \{w = (x, \lambda) \in \mathcal{S} : \lambda_i > 0 \,\forall i \in \mathcal{B} \quad \text{and} \quad \lambda_i < 0 \,\forall i \in \mathcal{N}\}.$$

Lemma 5.2. *Suppose that Assumptions* 2.1, 5.1*, and* 5.2 *are satisfied. Then there are a constant $\mathcal{C}_0 > 0$ and a sufficiently small constant $\varepsilon > 0$ such that* $\mathrm{dist}(w, \mathcal{S}) \le \mathcal{C}_0\|H_0(w)\|$ *holds for all $w = (x, \lambda) \in \Re^{n+m}$ sufficiently close to $w^* = (x^*, \lambda^*)$, where* $\mathrm{dist}(w, \mathcal{S})$ *is the Euclidean distance of $w \in \Re^{n+m}$ to the set $\mathcal{S}$.*

Proof: The proof can be found in Appendix I. $\square$

For the KKT system of nonlinear programming, similar error bound results have already been established in [6, 27] (Some more general error bound results for generalized equations can be found in [10]). Here, because a normal map is used, a proof is necessary.

Theorem 5.1. *Suppose that Assumptions* 2.1, 3.1, 5.1*, and* 5.2 *are satisfied. Then there is a constant $\mathcal{C} > 0$ such that* $\mathrm{dist}(w^k, \mathcal{S}) \le \mathcal{C}\theta(z^k)$ *holds for all z^k sufficiently close to* z^*.

Proof: Since $w^k \to w^*$ as $k \to \infty$, by Lemma 5.2 we have

$$\text{dist}(w^k, \mathcal{S}) \leq \mathcal{C}_0 \|H_0(w^k)\| \tag{5.2}$$

for all z^k sufficiently close to z^*. Let $G : \Re^{1+n+m} \to \Re^{n+m}$ be defined by

$$H(z) := \begin{pmatrix} \mu \\ G(x) \end{pmatrix}, \ z = (\mu, x, \lambda) \in \Re^{1+n+m}.$$

Then, from the definitions of $H_0(\cdot)$ and $G(\cdot)$, we have

$$
\begin{aligned}
\|H_0(w^k)\| &\leq \|G(z^k)\| + \|H_0(w^k) - G(z^k)\| \\
&= \|G(z^k)\| + \left\| \begin{pmatrix} f'(x^k)^T(\Phi(y^k) - (\lambda^k)_+) + g_1(\mu_k)x^k \\ \lambda_+^k - \Phi(y^k) + g_2(\mu_k)\lambda^k + g_3(\mu_k)\pi(z^k) \end{pmatrix} \right\| \\
&\leq \theta(z^k) + \|f'(x^k)\|\|\Phi(y^k) - (\lambda^k)_+\| + g_1(\mu_k)\|x^k\| \\
&\quad + \|(\lambda^k)_+ - \Phi(y^k)\| + g_2(\mu_k)\|\lambda^k\| + g_3(\mu_k)\|\pi(z^k)\|,
\end{aligned}
\tag{5.3}
$$

where $y^k = (\mu_k, \lambda^k)$. For any $i \in \mathcal{B}$ and z^k sufficiently close to z^*, we have

$$
\begin{aligned}
|\Phi_i(y^k) - (\lambda_i^k)_+| &= \left| \left(\lambda_i^k + \sqrt{(\lambda_i^k)^2 + 4(\mu_k)^2} \right) \Big/ 2 - \lambda_i^k \right| \\
&= \left| \lambda_i^k - \sqrt{(\lambda_i^k)^2 + 4(\mu_k)^2} \right| \Big/ 2 \\
&= 2(\mu_k)^2 \Big/ \left(\lambda_i^k + \sqrt{(\lambda_i^k)^2 + 4(\mu_k)^2} \right) = O((\mu_k)^2).
\end{aligned}
\tag{5.4}
$$

Similarly, for any $i \in \mathcal{N}$ and z^k sufficiently close to z^*, we have

$$\left|\Phi_i(y^k) - (\lambda_i^k)_+\right| = \left|\phi_i\left(\mu_k, \lambda_i^k\right)\right| = (\mu_k)^2 \Big/ \left(\sqrt{(\lambda_i^k)^2 + 4(\mu_k)^2} - \lambda_i^k \right) = O((\mu_k)^2). \tag{5.5}$$

Moreover, by (2.10), (2.11), and Theorem 3.1, we have

$$
\begin{aligned}
&g_1(\mu_k)\|x^k\| = O((\mu_k)^2), \ g_2(\mu_k)\|\lambda^k\| = O((\mu_k)^2), \text{ and} \\
&g_3(\mu_k)\|\pi(z^k)\| = O((\mu_k)^2).
\end{aligned}
\tag{5.6}
$$

By using (5.4)–(5.6), we can obtain from (5.3) that for all z^k sufficiently close to z^*,

$$\|H_0(w^k)\| = \theta(z^k) + O\left(\mu_k^2\right). \tag{5.7}$$

Since $\mu_k \to 0$ as $k \to \infty$, we have $(\mu_k)^2 \leq \mu_k \leq \theta(z^k)$ for all z^k sufficiently close to z^*. This, together with (5.2) and (5.7), implies that there is a constant $\mathcal{C} > 0$ such that $\text{dist}(w^k, \mathcal{S}) \leq \mathcal{C}\theta(z^k)$ holds for all z^k sufficiently close to z^*. This completes the proof. $\quad\square$

5.3.   Upper bound of $\|\Delta z^k\|$

Lemma 5.3. *Suppose that Assumptions 2.1, 3.1, 5.1, and 5.2 are satisfied. Then $v(z^k) = u(z^k)$ for all z^k sufficiently close to z^*, where $v(\cdot)$ and $u(\cdot)$ are defined by (3.8) and (3.7), respectively.*

Proof: The proof can be found in Appendix II.   □

Suppose that Assumptions 2.1, 3.1, 5.1, and 5.2 are satisfied. Then, from Algorithm 3.1 and Lemma 5.3, we have that for all z^k sufficiently close to z^*,

$$\Delta\mu_k = -\mu_k + \mu_0\beta(z^k), \tag{5.8}$$

$$(f'(x^k)^T(\Phi(y^k))'_\mu + g'_1(\mu_k)x^k)\Delta\mu_k + M(z^k)\Delta x^k + f'(x^k)^T(\Phi(y^k))'_\lambda\Delta\lambda^k$$
$$= -(\nabla f_0(x^k) + f(x^k)^T\Phi(y^k) + g_1(\mu_k)x^k) + u_1(z^k), \tag{5.9}$$

$$P(z^k)\Delta\mu_k + R(z^k)(-f'(x^k))\Delta x^k + N(z^k)\Delta\lambda^k$$
$$= -(-f(x^k) + \lambda^k - \Phi(y^k) + g_2(\mu_k)\lambda^k + g_3(\mu_k)\pi(z^k)) + u_2(z^k), \tag{5.10}$$

where $y^k = (\mu_k, \lambda^k)$, matrices $M(\cdot)$, $P(\cdot)$, $Q(\cdot)$, and $N(\cdot)$ are given by (2.13), (2.14), (2.15), and (2.16), respectively; and vectors $u_1(\cdot)$ and $u_2(\cdot)$ are defined by (3.2). From (5.8) we have

$$|\Delta\mu_k| = O(\theta(z^k)). \tag{5.11}$$

Thus, we need only to derive the upper bounds of $\|\Delta x^k\|$ and $\|\Delta\lambda^k\|$. Let

$$t^k := \Phi(y^k), \quad r^k := -(\lambda^k - \Phi(y^k)), \tag{5.12}$$

and

$$\Delta t^k := (\Phi(y^k))'_\lambda\Delta\lambda^k + (\Phi(y^k))'_\mu\Delta\mu_k - \hat{u}_1(z^k), \tag{5.13}$$
$$\Delta r^k := -N(z^k)\Delta\lambda^k + (\Phi(y^k))'_\mu\Delta\mu_k + \hat{u}_2(z^k). \tag{5.14}$$

Then, by (5.9), (5.10), and (3.2)–(3.6),

$$M(z^k)\Delta x^k + f'(x^k)^T\Delta t^k = -(\nabla f_0(x^k) + f'(x^k)^T t^k),$$
$$R(z^k)(-f'(x^k))\Delta x^k - \Delta r^k = -(-f(x^k) - r^k),$$

and, by (5.13) and (5.14),

$$N(z^k)\Delta t^k + (\Phi(y^k))'_\lambda\Delta r^k$$
$$= N(z^k)[(\Phi(y^k))'_\mu\Delta\mu_k - \hat{u}_1(z^k)] + (\Phi(y^k))'_\lambda[(\Phi(y^k))'_\mu\Delta\mu_k + \hat{u}_2(z^k)],$$

where $N(\cdot)$ is defined by (2.16).

For each k, let $w^{k*} := (x^{k*}, \lambda^{k*}) \in \mathcal{S}$ be such that

$$\|w^k - w^{k*}\| = \mathrm{dist}(w^k, \mathcal{S}). \tag{5.15}$$

Denote

$$y^{k*} := (0, \lambda^{k*}), \quad t^{k*} := \Phi(y^{k*}), \quad \text{and} \quad r^{k*} := -(\lambda^{k*} - \Phi(y^{k*})). \tag{5.16}$$

Let

$$\overline{\Delta x}^k := x^k - x^{k*} + \Delta x^k, \ \overline{\Delta t}^k := t^k - t^{k*} + \Delta t^k, \ \text{and} \ \overline{\Delta r}^k := r^k - r^{k*} + \Delta r^k. \tag{5.17}$$

Then we can further obtain

$$M(z^k)\overline{\Delta x}^k + f'(x^k)^T \overline{\Delta t}^k = h_0(z^k), \qquad R(z^k)(-f'(x^k))\overline{\Delta x}^k - \overline{\Delta r}^k = h_1(x^k),$$

and

$$\begin{aligned}
N(z^k)&\overline{\Delta t}^k + (\Phi(y^k))'_\lambda \overline{\Delta r}^k \\
&= N(z^k)(t^k - t^{k*}) + (\Phi(y^k))'_\lambda (r^k - r^{k*}) \\
&\quad + N(z^k)[(\Phi(y^k))'_\mu \Delta\mu_k - \hat{u}_1(z^k)] + (\Phi(y^k))'_\lambda [(\Phi(y^k))'_\mu \Delta\mu_k + \hat{u}_2(y^k)] \\
&= N(z^k)(\Phi(y^k) - \Phi(y^{k*})) + (\Phi(y^k))'_\lambda (\lambda^{k*} - \Phi(y^{k*}) - (\lambda^k - \Phi(y^k))) \\
&\quad + N(z^k)[(\Phi(y^k))'_\mu \Delta\mu_k - \hat{u}_1(z^k)] + (\Phi(y^k))'_\lambda [(\Phi(y^k))'_\mu \Delta\mu_k + \hat{u}_2(z^k)] \\
&= N(z^k)h_2(z^k) + (\Phi(y^k))'_\lambda h_3(z^k),
\end{aligned}$$

where

$$h_0(z^k) = M(z^k)(x^k - x^{k*}) + f'(x^k)^T(t^k - t^{k*}) - (\nabla f_0(x^k) + f'(x^k)^T t^k), \tag{5.18}$$
$$h_1(y^k) = f(x^k) - f(x^{k*}) - R(z^k)f'(x^k)(x^k - x^{k*}), \tag{5.19}$$
$$h_2(y^k) = \Phi(y^k) - \Phi(y^{k*}) + (\Phi(y^k))'_\mu \Delta\mu_k - \hat{u}_1(z^k), \tag{5.20}$$
$$h_3(y^k) = \lambda^{k*} - \Phi(y^{k*}) - (\lambda^k - \Phi(y^k)) + (\Phi(y^k))'_\mu \Delta\mu_k + \hat{u}_2(z^k). \tag{5.21}$$

Thus, (5.9) and (5.10) become

$$\begin{bmatrix} M(z^k) & f'(x^k)^T & 0 \\ R(z^k)(-f'(x^k)) & 0 & -E^m \\ 0 & N(z^k) & (\Phi(y^k))'_\lambda \end{bmatrix} \begin{bmatrix} \overline{\Delta x}^k \\ \overline{\Delta t}^k \\ \overline{\Delta r}^k \end{bmatrix} = \begin{bmatrix} h_0(z^k) \\ h_1(x^k) \\ h_4(z^k) \end{bmatrix}, \tag{5.22}$$

where $M(\cdot)$ and $N(\cdot)$ are defined by (2.13) and (2.16), respectively, and $h_4(\cdot)$ is defined by $h_4(z^k) := N(z^k)h_2(z^k) + (\Phi(y^k))'_\lambda h_3(z^k)$. Furthermore, (5.22) can be split into the

following two systems:

$$
\begin{bmatrix}
M(z^k) & f'(x^k)^T & 0 \\
R(z^k)(-f'(x^k)) & 0 & -E^m \\
0 & N(z^k) & (\Phi(y^k))'_\lambda
\end{bmatrix}
\begin{bmatrix}
\widehat{\Delta x}^k \\
\widehat{\Delta t}^k \\
\widehat{\Delta r}^k
\end{bmatrix}
=
\begin{bmatrix}
0 \\
0 \\
h_4(z^k)
\end{bmatrix}
\tag{5.23}
$$

and

$$
\begin{bmatrix}
M(z^k) & f'(x^k)^T & 0 \\
R(z^k)(-f'(x^k)) & 0 & -E^m \\
0 & N(z^k) & (\Phi(y^k))'_\lambda
\end{bmatrix}
\begin{bmatrix}
\widetilde{\Delta x}^k \\
\widetilde{\Delta t}^k \\
\widetilde{\Delta r}^k
\end{bmatrix}
=
\begin{bmatrix}
h_0(z^k) \\
h_1(x^k) \\
0
\end{bmatrix},
\tag{5.24}
$$

where

$$
\overline{\Delta x}^k = \widehat{\Delta x}^k + \widetilde{\Delta x}^k, \ \overline{\Delta t}^k = \widehat{\Delta t}^k + \widetilde{\Delta t}^k, \ \text{and} \ \overline{\Delta r}^k = \widehat{\Delta r}^k + \widetilde{\Delta r}^k.
\tag{5.25}
$$

In the subsequent analysis, we first obtain upper bounds of $\|\widehat{\Delta x}^k\|$, $\|\widehat{\Delta t}^k\|$, $\|\widehat{\Delta r}^k\|$ and $\|\widetilde{\Delta x}^k\|$, $\|\widetilde{\Delta t}^k\|$, $\|\widetilde{\Delta r}^k\|$ by using (5.23) and (5.24), respectively. From these bounds and (5.25) we obtain the upper bounds of $\|\overline{\Delta x}^k\|$, $\|\overline{\Delta t}^k\|$, and $\|\overline{\Delta r}^k\|$, which, together with (5.17), yield the upper bounds of $\|\Delta x^k\|$, $\|\Delta t^k\|$ and $\|\Delta r^k\|$. Finally, by (5.13) and (5.14), we derive the upper bound of $\|\Delta \lambda^k\|$.

The upper bounds of $\|\widehat{\Delta t}^k\|$, $\|\widehat{\Delta r}^k\|$ and $\|\widehat{\Delta x}^k\|$ can be obtained from the following two lemmas. Their proofs can be found in Appendix III.

Lemma 5.4. *Suppose that Assumptions* 2.1, 3.1, 5.1, *and* 5.2 *are satisfied. Let* $\widehat{\Delta x}^k$, $\widehat{\Delta t}^k$, *and* $\widehat{\Delta r}^k$ *be generated by* (5.23). *Then* $\max\{\|\widehat{\Delta t}^k\|, \|\widehat{\Delta r}^k\|\} = O(\theta(z^k))$ *holds for all* z^k *sufficiently close to* z^*.

Lemma 5.5. *Suppose that Assumptions* 2.1, 3.1, 5.1, *and* 5.2 *are satisfied. Let* $\widehat{\Delta x}^k$, $\widehat{\Delta t}^k$, *and* $\widehat{\Delta r}^k$ *be generated by* (5.23). *Then* $\|\widehat{\Delta x}^k\| = O(\theta(z^k))$ *holds for all* z^k *sufficiently close to* z^*.

The proof of the following proposition can be found in Appendix IV.

Proposition 5.1. *Suppose that Assumptions* 2.1, 3.1, 5.1, *and* 5.2 *are satisfied. Let* $\widetilde{\Delta x}^k$, $\widetilde{\Delta t}^k$, *and* $\widetilde{\Delta r}^k$ *be generated by* (5.24). *Then* $\|\widetilde{\Delta x}^k\| = O(\theta^2(z^k)) + O(\theta(z^k))\|\widetilde{\Delta t}_B^k\|$ *holds for all* z^k *sufficiently close to* z^*.

In the following, let $D^k := [(\Phi(y^k))'_\lambda]^{-1/2}[N(z^k)]^{1/2}$. By Proposition 5.1, it is sufficient to estimate the upper bounds of $\|\widetilde{\Delta t}^k\|$ and $\|\widetilde{\Delta r}^k\|$ in order to obtain the upper bounds of $\|\widetilde{\Delta x}^k\|$, $\|\widetilde{\Delta t}^k\|$, and $\|\widetilde{\Delta r}^k\|$.

The upper bounds of $\|\widetilde{\Delta t}_N^k\|$ and $\|\widetilde{\Delta r}_B^k\|$ can be obtained from the following lemma whose proof can be found in Appendix V.

Lemma 5.6. *Suppose that Assumptions 2.1, 3.1, 5.1, and 5.2 are satisfied. Let $C(\cdot)$ be defined by (5.1), and $\widetilde{\Delta x}^k$, $\widetilde{\Delta t}^k$, and $\widetilde{\Delta r}^k$ be generated by (5.24). Then $\max\{\|\widetilde{\Delta t}^k_\mathcal{N}\|,$ $\|\widetilde{\Delta r}^k_\mathcal{B}\|\} = O(\theta(z^k))$ holds for all z^k sufficiently close to z^*.*

To obtain the upper bound of $\|\widetilde{\Delta t}^k_\mathcal{B}\|$ and $\|\widetilde{\Delta r}^k_\mathcal{N}\|$, we need to establish the following proposition whose proof can be found in Appendix VI.

Proposition 5.2. *Suppose that Assumptions 2.1 and 5.1–5.3 are satisfied. Let $C(\cdot)$ be defined by (5.1). Then for all z^k sufficiently close to z^*, $(\widetilde{\Delta x}^k, \widetilde{\Delta t}^k_\mathcal{B}, \widetilde{\Delta r}^k_\mathcal{N})$ is the solution of the following (weighted) least squares problem*

$$\min \quad \frac{1}{2}\left\|(D^k)_{\mathcal{B}\mathcal{B}}\widetilde{\Delta t}^k_\mathcal{B}\right\|^2 + \frac{1}{2}\left\|(D^k)^{-1}_{\mathcal{N}\mathcal{N}}\widetilde{\Delta r}^k_\mathcal{N}\right\|^2$$

$$\text{s.t.} \quad C(z^k)^T \begin{bmatrix} \widetilde{\Delta x}^k \\ \widetilde{\Delta t}^k_\mathcal{B} \\ \widetilde{\Delta r}^k_\mathcal{N} \end{bmatrix} = \begin{bmatrix} h_0(z^k) - f'_\mathcal{N}(x^k)^T\widetilde{\Delta t}^k_\mathcal{N} \\ h_1(x^k) + \widetilde{\Delta r}^k_\mathcal{B} \end{bmatrix}. \tag{5.26}$$

By using Proposition 5.2, we can obtain the upper bounds of $\|\widetilde{\Delta t}^k_\mathcal{B}\|$ and $\|\widetilde{\Delta r}^k_\mathcal{N}\|$ in the following lemma whose proof can be found in Appendix VII.

Lemma 5.7. *Suppose that Assumptions 2.1, 3.1, and 5.1–5.3 are satisfied. Let $\widetilde{\Delta x}^k$, $\widetilde{\Delta t}^k$, and $\widetilde{\Delta r}^k$ be generated by (5.24). Then $\max\{\|\widetilde{\Delta t}^k_\mathcal{B}\|, \|\widetilde{\Delta r}^k_\mathcal{N}\|\} = O(\theta(z^k))$ holds for all z^k sufficiently close to z^*.*

By Proposition 5.1 and Lemma 5.7, we further obtain

Lemma 5.8. *Suppose that Assumptions 2.1, 3.1, and 5.1–5.3 are satisfied. Then $\|\widetilde{\Delta x}^k\| = O(\theta(z^k))$ holds for all z^k sufficiently close to z^*.*

The next theorem is about the upper bound of $\|\Delta z^k\|$.

Theorem 5.2. *Suppose that Assumptions 2.1, 3.1, and 5.1–5.3 are satisfied. Let z^k and Δz^k be generated by Algorithm 3.1. Then there exists a constant $\mathcal{C}_1 > 0$ such that $\|\Delta z^k\| \le \mathcal{C}_1\theta(z^k)$ holds for all z^k sufficiently close to z^*.*

Proof: For all z^k sufficiently close to z^*, from Lemmas 5.6, 5.7, and 5.8 we obtain

$$\max\{\|\widetilde{\Delta x}^k\|, \|\widetilde{\Delta t}^k\|, \|\widetilde{\Delta r}^k\|\} = O(\theta(z^k)),$$

and from Lemmas 5.4 and 5.5 we have

$$\max\{\|\widehat{\Delta x}^k\|, \|\widehat{\Delta t}^k\|, \|\widehat{\Delta r}^k\|\} = O(\theta(z^k)).$$

Thus, by (5.25), we obtain that for all z^k sufficiently close to z^*,

$$\|\overline{\Delta x}^k\| \le \|\widehat{\Delta x}^k\| + \|\widetilde{\Delta x}^k\| = O(\theta(z^k)),$$
$$\|\overline{\Delta t}^k\| \le \|\widehat{\Delta t}^k\| + \|\widetilde{\Delta t}^k\| = O(\theta(z^k)),$$
$$\|\overline{\Delta r}^k\| \le \|\widehat{\Delta r}^k\| + \|\widetilde{\Delta r}^k\| = O(\theta(z^k)).$$

Furthermore, by (5.17), we obtain that for all z^k sufficiently close to z^*,

$$\|\Delta x^k\| \le \|\overline{\Delta x}^k\| + \|x^k - x^{k*}\| = O(\theta(z^k)),$$
$$\|\Delta t^k\| \le \|\overline{\Delta t}^k\| + \|t^k - t^{k*}\| = O(\theta(z^k)), \tag{5.27}$$
$$\|\Delta r^k\| \le \|\overline{\Delta r}^k\| + \|r^k - r^{k*}\| = O(\theta(z^k)).$$

From (5.13) and (5.14) it follows that

$$\Delta t_{\mathcal{B}}^k = ((\Phi(y^k))'_\lambda)_{\mathcal{B}\mathcal{B}} \Delta\lambda_{\mathcal{B}}^k + ((\Phi(y^k))'_\mu)_{\mathcal{B}} \Delta\mu_k - (\hat{u}_1(y^k))_{\mathcal{B}},$$
$$\Delta r_{\mathcal{N}}^k = -N(z^k)_{\mathcal{N}\mathcal{N}} \Delta\lambda_{\mathcal{N}}^k + ((\Phi(y^k))'_\mu)_{\mathcal{N}} \Delta\mu_k + (\hat{u}_2(y^k))_{\mathcal{N}}.$$

Thus, for all z^k sufficiently close to z^*,

$$\left\|\Delta\lambda_{\mathcal{B}}^k\right\| \le \|[((\Phi(y^k))'_\lambda)_{\mathcal{B}\mathcal{B}}]^{-1}\|\left[\left\|\Delta t_{\mathcal{B}}^k\right\| + \left\|((\Phi(y^k))'_\mu)_{\mathcal{B}}\Delta\mu_k\right\| + \left\|(\hat{u}_1(y^k))_{\mathcal{B}}\right\|\right]$$
$$= O(\theta(z^k)), \tag{5.28}$$
$$\left\|\Delta\lambda_{\mathcal{N}}^k\right\| \le \|[N(z^k)_{\mathcal{N}\mathcal{N}}]^{-1}\|\left[\left\|\Delta r_{\mathcal{N}}^k\right\| + \left\|((\Phi(y^k))'_\mu)_{\mathcal{N}}\Delta\mu_k\right\| + \left\|(\hat{u}_2(y^k))_{\mathcal{N}}\right\|\right]$$
$$= O(\theta(z^k)). \tag{5.29}$$

By using (5.11), (5.27), (5.28), and (5.29), we obtain that $\|\Delta z^k\| = O(\theta(z^k))$ for all z^k sufficiently close to z^*, which completes the proof. $\qquad\square$

5.4.   Superlinear convergence

In this subsection, we always assume that Assumptions 2.1, 3.1, and 5.1–5.3 are satisfied.

Lemma 5.9.   *Let z^k and Δz^k be generated by Algorithm 3.1. Then there exists a constant $C_2 > 0$ such that*

$$z^{k+1} = z^k + \Delta z^k \quad \text{and} \quad \theta(z^{k+1}) \le C_2 \theta^{1+t_1}(z^k)$$

holds for all z^k sufficiently close to z^*.

Proof: By using Theorem 5.2, we can prove this lemma in a similar way as Lemma 8 in [14]. We omit the details here. $\qquad\square$

By Lemma 5.9, there exists a constant $\mathcal{C}_3 > 0$ such that for all z^k sufficiently close to z^*,

$$\theta(z^{k+1}) \leq (\mathcal{C}_3\theta(z^k))^{1+t_1}. \tag{5.30}$$

For given $\varepsilon > 0$, define $\mathcal{N}(z^*, \varepsilon) := \{z \in \mathfrak{R}_+ \times \mathfrak{R}^n \times \mathfrak{R}^m : \|z - z^*\| \leq \varepsilon\}$. Since H is locally Lipschitz continuous around z^*, there exists a constant $\mathcal{L} > 0$ such that $\|H(z^1) - H(z^2)\| \leq \mathcal{L}\|z^1 - z^2\|$ holds for any $z^1, z^2 \in N(z^*, \varepsilon)$. Let

$$\bar{\varepsilon} := \min\{\varepsilon/(2 + 2\mathcal{C}_1\mathcal{L} + 4\mathcal{C}_1\mathcal{C}_3\mathcal{L}), \ 1/(2\mathcal{C}_3\mathcal{L})\}, \tag{5.31}$$

where $\mathcal{C}_1$ and $\mathcal{C}_3$ are given by Theorem 5.2 and (5.30), respectively. Then the following lemma can be proved in a similar way as in [14, Lemma 5.5]. We omit the proof here.

Lemma 5.10. *Let $\bar{\varepsilon}$ be defined by (5.31). If for some k the iterate $z^k \in N(z^*, \bar{\varepsilon})$ and ε is sufficiently small, then $z^{k+q} \in N(z^*, \varepsilon/2)$ for all $q = 0, 1, 2, \ldots$ and $\{z^{k+q}\}_{q=1}^{\infty}$ is a convergent sequence.*

Theorem 5.3. *Let z^* be an accumulation point of the iteration sequence $\{z^k\}$ generated by Algorithm 3.1. Suppose that Assumptions 2.1 and 5.1–5.3 are satisfied. Then*
  (i) *the whole sequence $\{z^k\}$ converges to z^*,*
  (ii) *$\theta(z^{k+1}) = O(\theta^{1+t_1}(z^k))$, $\mu_{k+1} = O((\mu_k)^{1+t_1})$,   and*
  (iii) *$\text{dist}(w^{k+1}, \mathcal{S}) = O((\text{dist}(w^k, \mathcal{S}))^{1+t_1})$.*

Proof:
  (i) This is obtained by Lemma 5.10.
  (ii) By Lemma 5.9, we know that for all z^k sufficiently close to z^*, $z^{k+1} = z^k + \Delta z^k$ and $\theta(z^{k+1}) = O(\theta^{1+t_1}(z^k))$. In addition, since $z^{k+1} = z^k + \Delta z^k$ for all z^k sufficiently close to z^*, it follows that $\mu_{k+1} = \mu_k + \Delta\mu_k = \gamma\mu_0\theta^{1+t_1}(z^k)$ for all z^k sufficiently close to z^*. This, together with $\theta(z^{k+1}) = O(\theta^{1+t_1}(z^k))$, implies that $\mu_{k+1} = O((\mu_k)^{1+t_1})$ holds for all z^k sufficiently close to z^*. Thus, by the convergence of $\{z^k\}$, (ii) holds.
  (iii) By Theorem 5.1 and (ii), we have

$$\text{dist}(w^{k+1}, \mathcal{S}) = O(\theta(z^{k+1})) = O(\theta^{1+t_1}(z^k)). \tag{5.32}$$

On the other hand, by the Lipschitz continuity of H and the boundedness of $\mathcal{S}$,

$$\theta^{1+t_1}(z^k) = \|H(z^k) - H(\tilde{z})\|^{1+t_1} = O(\|z^k - \tilde{z}\|^{1+t_1}), \ \forall\tilde{z} := (0, \tilde{w}) \text{ with } \tilde{w} \in \mathcal{S}.$$

In particular, we take $\tilde{z} := z^{k*}$ where w^{k*} is the projection of w^k onto $\mathcal{S}$, then

$$\theta^{1+t_1}(z^k) = O(\|z^k - z^{k*}\|^{1+t_1}) = O((\text{dist}(w^k, \mathcal{S}))^{1+t_1}). \tag{5.33}$$

By combining (5.32) with (5.33), we obtain that (iii) holds. The proof is complete. □

Table 1.   The numerical results of Examples 6.1–6.7

Prob	IT	NF	Val
Example 1	5	7	3.39×10^{-9}
Example 2	8	12	1.41×10^{-8}
Example 3	12	36	7.99×10^{-8}
Example 4	10	13	4.82×10^{-9}
Example 5	4	6	3.40×10^{-7}
Example 6	5	6	1.33×10^{-7}
Example 7	5	6	1.78×10^{-7}

Table 2.   The numerical results of Example 6.8

n	m	AIT	ANF
500	100	6.9	7.9
500	500	10.3	13.8
500	1000	20.7	14.7

6.   Numerical results

In this section, we report some numerical experiments for Algorithm 3.1 running in Matlab. Throughout the computational experiments, we chose starting points as

$$x^0 = (0, \ldots, 0)^T \in \Re^n, \ \lambda^0 = (0, \ldots, 0)^T \in \Re^m, \ \mu_0 = 1.0;$$

and the parameters used in the algorithm were chosen as

$$\sigma = 10^{-5}, \ \delta = 0.5, \ t_1 = 0.2, \ t_2 = 0.5, \ \kappa = 0.1, \ \tau = 1/(10\sqrt{n+m}), \ \gamma = 0.02.$$

We used $\|H_0(w^k)\| \leq 10^{-6}$ as the stopping criterion, where the function H_0 is defined by (2.4). We tested the three groups of problems: Examples 6.1–6.7 (group 1), Example 6.8 (group 2), and Examples 6.9–6.10 (group 3). Numerical results for problems in groups 1 and 3 are reported in Tables 1 and 3, where **Prob** denotes the problem to be tested; **IT** denotes the number of iterations; **NF** denotes the number of function evaluations for the function H defined by (2.8); and **Val** denotes the value of $\|H_0(w^k)\|$ when the algorithm stops. Numerical results for problems in group 2 are reported in Table 2, where, for each given pair (n, m), the problem is run ten times, **AIT** denotes the average number of iterations among the ten runs; and **ANF** denotes the average number of function evaluations for H among the ten runs.

The numerical results in Tables 1–3 show that we only need a small number of iterations for each example tested.

First, we test the following seven problems with small sizes. The tested results are listed in Table 1.

Table 3.   The numerical results of Examples 6.9–6.10

Prob	m	IT	NF	Val
Example 6.9	1000	17	44	3.68×10^{-8}
	1500	18	47	1.89×10^{-7}
	2000	20	52	6.48×10^{-8}
Example 6.10	1000	17	46	3.68×10^{-7}
	1500	19	50	9.42×10^{-8}
	2000	19	48	1.46×10^{-7}

Example 6.1.   (Problem 1 in [3])

$$\min 0.5(x_1 - 5)^2 + 0.5x_2^2 \quad \text{s.t.} \quad \begin{cases} 0.5x_2^2 + x_1 - 4 \leq 0 \\ 0.5x_1^2 + x_1 - 20 \leq 0. \end{cases}$$

Example 6.2.   (Problem 2 in [3])

$$\min 0.5(x_1 - 5)^2 + 0.5x_2^2 \quad \text{s.t.} \quad \begin{cases} 0.5x_2^2 + x_1 - 4 \leq 0 \\ 0.5x_1^2 + x_2 - 10 \leq 0. \end{cases}$$

Example 6.3.   (Problem 3 in [3])

$$\min \quad 0.5\left(5x_1^2 + 14x_1x_2 + 13x_2^2\right) - 18x_1 - 32x_2$$
$$\text{s.t.} \quad \begin{cases} 2.5x_1^2 - x_1x_2 + 5x_2^2 + 2x_1 + 3x_2 - 11.5 \leq 0 \\ 2x_1^2 - 2x_1x_2 + 0.5x_2^2 - 2x_1 + x_2 - 1 \leq 0. \end{cases}$$

Example 6.4.   (Problem 4 in [3])

$$\min \quad 0.5\left(10x_1^2 + 38x_1x_2 + 41x_2^2\right) - 47.5x_1 - 63x_2$$
$$\text{s.t.} \quad \begin{cases} 5x_1^2 + x_1x_2 + 2.5x_2^2 + x_1 + x_2 - 3.125 \leq 0 \\ 2.5x_1^2 + 7x_1x_2 + 6.5x_2^2 - x_1 + 2x_2 - 5 \leq 0 \\ 2.5x_1^2 - x_1x_2 + 5x_2^2 + 3x_1 + x_2 - 3.625 \leq 0 \\ 2x_1^2 - 2x_1x_2 + 0.5x_2^2 + 2x_1 + 3x_2 - 5.5 \leq 0 \\ 4.5x_1^2 + 6x_1x_2 + 2x_2^2 - 2x_1 + x_2 - 2.625 \leq 0. \end{cases}$$

Example 6.5.   (Example 1 in [27])

$$\min x_1 + x_2 \quad \text{s.t.} \ (x_1 - 1)^2 + (x_2 - 1)^2 \leq 2.$$

Example 6.6.   (Example 2 in [27])

$$\min x_1 \qquad \text{s.t.} \quad \begin{bmatrix} (x_1 - 2)^2 + x_2^2 \\ (x_1 - 4)^2 + x_2^2 \end{bmatrix} \leq \begin{bmatrix} 4 \\ 16 \end{bmatrix}.$$

Example 6.7.   (Example 3 in [27])

$$
\begin{aligned}
\min \quad & x_1^2 + x_1 x_2 + 2x_2^2 + x_1 + x_2 \\
\text{s.t.} \quad & \frac{1}{2}(x_1 - 2)^2 + \frac{1}{2}(x_2 - 1)^2 \leq \frac{5}{2}, \\
& x_1 \geq 0, \, x_2 \geq 0.
\end{aligned}
$$

Next, we consider the following min-max problem:

$$\min_{x \in \Re^{n-1}} \left\{ \max_{i \in \mathcal{J}} f_i(x) \right\} + f_0(x), \tag{6.1}$$

where $\mathcal{J} = \{1, 2, \ldots, m\}$ and

$$
\begin{aligned}
f_i(x) &:= \frac{1}{2} x^T (a^i (a^i)^T) x + (b^i)^T x + c^i, \quad \forall i \in \mathcal{J}, \\
f_0(x) &:= \frac{1}{2} x^T (A A^T + E^{n-1}) x + (b^0)^T x,
\end{aligned}
$$

with $x, a^i, b^i, b^0 \in \Re^{n-1}$, $c^i \in \Re$, and $A \in \Re^{(n-1) \times m}$.

Problem (6.1) is equivalent to the following QCQP

$$
\begin{aligned}
\min \quad & t + f_0(x) \\
\text{s.t.} \quad & f_i(x) - t \leq 0, \quad i \in \mathcal{J}.
\end{aligned}
\tag{6.2}
$$

Thus, we can solve problem (6.1) by making use of Algorithm 3.1 to solve (6.2).

For any two positive integers n_1, n_2, one positive real number r_1, and one real number $r_2 \in [0, 1]$, we used the following notation: $\lfloor r_1 \rfloor$ denotes the maximal integer which is not large than r_1; $rand(n_1, n_2)$ denotes an $n_1 \times n_2$ matrix whose entries are randomly chosen in $(0, 1)$; and $sprand(n_1, n_2, r_2)$ denotes a random, $m \times n$, sparse matrix with approximately $r_2 * n_1 * n_2$ uniformly distributed nonzero entries. In our testing, we consider the following example:

Example 6.8.   Consider the problem (6.1) with $b^0 := sprand(n - 1, 1, 0.1)$, $A := rand(n - 1, m)$, and for each $i \in \mathcal{J}$,

$$a^i := \left| 1 - \frac{i}{\lfloor n/2 \rfloor + 1} \right| sprand(n - 1, 1, 0.1), \quad b^i := sprand(n - 1, 1, 0.1), \text{ and}$$

$$c^i := rand(1).$$

The tested results are listed in Table 2.

Finally, we consider the semi-infinite min-max problem (see [23]):

$$\min_{x \in \Re^n} F(\psi(x)), \tag{6.3}$$

where $F : \Re^{n_1} \to \Re$ is a smooth function; and $\psi : \Re^n \to \Re^{n_1}$ is a nonsmooth function. The components of $\psi(\cdot)$ are the form:

$$\psi^j(x) = \max_{y_j \in Y_j} \psi^j(x, y_j) \text{ with } Y_j \in R^{m_j} \text{ and } j \in \{1, 2, \ldots, n_1\}.$$

To solve (6.3), the following subproblem needs to be solved at each iteration [23]:

$$\min_{(p,h)} \quad \nabla F(\psi(x))^T p + \frac{1}{2} p^T \nabla^2 F(\psi(x)) p$$

$$\text{s.t.} \quad p^j - \psi^j(x) - \nabla_x \psi^j(x, y_j)^T h - h^T (\nabla^2 \phi^j(x, y_j) h + \psi^j(x)) \geq 0, \tag{6.4}$$
$$j \in \{1, 2, \ldots, n_1\}, \quad y_j \in Y_j.$$

In this paper, we tested the following problem:

$$\min_{x \in \Re^2} F(\psi^1(x), \psi^2(x)),$$

where $F : \Re^2 \to \Re$ and

$$\psi^1(x) = \max_{t \in Y_1} \{t^2 - (tx_1 + e^t x_2) + (x_1 + x_2)^2 + x_1^2 + x_2^2 + e^{(x_1+x_2)}\},$$

$$\psi^2(x) = \max_{t \in Y_2} \{-(t-1)^2 + 0.5(x_1 + x_2)^2 - 2t(x_1 + x_2) + 0.5[x_1^2 + x_2^2]\}$$

with $x \in \Re^2$, $Y_1 = [0, 1]$, and $Y_2 = [-1, 0]$. The function F is chosen as follows.

Example 6.9. $F(z) = z_1 + z_2, z \in \Re^2$.

Example 6.10. $F(z) = 0.5(z_1 + \sqrt{z_1^2 + 4}) + \ln(1 + e^{z_2}) + 0.5(z_1^2 + z_2^2), z \in \Re^2$.

In our testing about Examples 6.9 and 6.10, we considered the corresponding sub-problem 6.4 with discretized Y_1 and Y_2. For Example 6.9 we take

$$t_i = \frac{i - 1}{m/2 - 1} \quad \forall i \in \{1, \ldots, m/2\};$$

and for Example 6.10 we take

$$t_i = -\frac{i - m/2 - 1}{m/2 - 1} \quad \forall i \in \{m/2 + 1, \ldots, m\},$$

where m is an even number. The test results are listed in Table 3.

Appendix

Appendix I

Proof of Lemma 5.2: Assume, on the contrary, that the result is not true. Then, by Lemma 5.1 we may choose a subsequence $\{w^l\} := \{(x^l, \lambda^l)\} \subseteq \Re^{n+m}$ of infinite number such that for all w^l sufficiently close to w^*,

$$\|w^l - w^{l*}\| \geq l\|H_0(w^l)\|, \tag{7.1}$$

where $w^{l*} := (x^*, \lambda^{l*})$ is the projection of w^l onto $\mathcal{S}$. Since $\|w^* - w^{l*}\| \leq \|w^* - w^l\|$, by taking a subsequence if necessary, we can assume that both $\{w^l\}$ and $\{w^{l*}\}$ converge to $w^* := (x^*, \lambda^*)$. Let $\alpha_l := \|w^l - w^{l*}\|$. Then we can further assume that

$$(w^l - w^{l*})/\alpha_l \to (dx, d\lambda) \quad \text{as } l \to \infty. \tag{7.2}$$

In what follows, we show that $(dx, d\lambda) = 0$, which contradicts $\|(dx, d\lambda)\| = 1$.

First, we show that $(dx, d\lambda)$ is in the normal cone to $\mathcal{S}$ at $w^* = (x^*, \lambda^*)$, i.e.,

$$\begin{pmatrix} dx \\ d\lambda \end{pmatrix}^T \begin{pmatrix} x^* - x^* \\ \lambda - \lambda^* \end{pmatrix} \leq 0 \quad \forall (x^*, \lambda) \in \mathcal{S}. \tag{7.3}$$

Since w^{l*} is the projection of w^l onto $\mathcal{S}$ and $\mathcal{S}$ is convex, it follows that

$$\begin{pmatrix} x^l - x^* \\ \lambda^l - \lambda^{l*} \end{pmatrix}^T \begin{pmatrix} x^* - x^* \\ \lambda - \lambda^* \end{pmatrix} \leq 0 \quad \forall (x^*, \lambda) \in \mathcal{S}. \tag{7.4}$$

Thus, by (7.4), we have

$$\begin{pmatrix} (x^l - x^*)/\alpha_l \\ (\lambda^l - \lambda^{l*})/\alpha_l \end{pmatrix}^T \begin{pmatrix} x^* - x^* \\ \lambda - \lambda^* \end{pmatrix} \leq 0 \quad \forall (x^*, \lambda) \in \mathcal{S},$$

which, together with (7.2), implies (7.3).

Next, we show that $(dx, d\lambda)$ is in the tangent cone to $\mathcal{S}$ at w^*, i.e.,

$$w^* + \varepsilon_0(dx, d\lambda) \in \mathcal{S} \quad \text{for all sufficiently small } \varepsilon_0 > 0. \tag{7.5}$$

Since both $\{w^l\}$ and $\{w^{l*}\}$ converge to $w^* \in \mathcal{S}_0$, it follows that, for all sufficiently large l,

$$\lambda_i^l > 0, \lambda_i^{l*} > 0 \;\; \forall i \in \mathcal{B} \quad \text{and} \quad \lambda_i^l < 0, \lambda_i^{l*} < 0 \;\; \forall i \in \mathcal{N}. \tag{7.6}$$

Using the fact that $\nabla f_0(x^*) + f'(x^*)^T(\lambda^{l*})_+ = 0$ ($w^{l*} \in \mathcal{S}$), we have

$$\frac{\nabla f_0(x^l) + f'(x^l)^T(\lambda^l)_+}{\alpha_l}$$

$$
= \frac{(\nabla f_0(x^l) + f'(x^l)^T (\lambda^l)_+) - (\nabla f_0(x^*) + f'(x^*)^T (\lambda^{l*})_+)}{\alpha_l}
$$

$$
= \frac{\nabla f_0(x^l) - \nabla f_0(x^*)}{\alpha_l} + \frac{(f'(x^l)^T - f'(x^*)^T)(\lambda^{l*})_+}{\alpha_l} + \frac{f'(x^l)^T ((\lambda^l)_+ - (\lambda^{l*})_+)}{\alpha_l}
$$

$$
= P^0 \frac{x^l - x^*}{\alpha_l} + \sum_{i \in \mathcal{B}} P^i \frac{x^l - x^*}{\alpha_l} \lambda_i^{l*} + \sum_{i \in \mathcal{B}} f_i'(x^l) \frac{\lambda_i^l - \lambda_i^{l*}}{\alpha_l}
$$

$$
\rightarrow \left(P^0 + \sum_{i \in \mathcal{B}} \lambda_i^{l*} P^i \right) dx + \sum_{i \in \mathcal{B}} f_i'(x^*)(d\lambda)_i, \quad \text{as } l \rightarrow \infty, \tag{7.7}
$$

where the third equality is due to (7.6), and the last relation is due to (7.2). Using the fact that $-f(x^*) + \lambda^{l*} - (\lambda^{l*})_+ = 0$ $((x^*, \lambda^{l*}) \in \mathcal{S})$, we have

$$
\frac{-f(x^l) + \lambda^l - (\lambda^l)_+}{\alpha_l} = \frac{(-f(x^l) + \lambda^l - (\lambda^l)_+) - (-f(x^*) + \lambda^{l*} - (\lambda^{l*})_+)}{\alpha_l}
$$

$$
= \frac{-(f(x^l) - f(x^*)) + (\lambda^l - (\lambda^l)_+) - (\lambda^{l*} - (\lambda^{l*})_+)}{\alpha_l}. \tag{7.8}
$$

By (7.6) and (7.2), it follows from (7.8) that for $i \in \mathcal{B}$,

$$
\frac{[-f(x^l) + \lambda^l - (\lambda^l)_+]_i}{\alpha_l} = \frac{-(f_i(x^l) - f_i(x^*))}{\alpha_l} \rightarrow -f_i'(x^*)^T dx, \tag{7.9}
$$

and for $i \in \mathcal{N}$,

$$
\frac{[-f(x^l) + \lambda^l - (\lambda^l)_+]_i}{\alpha_l} = \frac{-(f_i(x^l) - f_i(x^*)) + \lambda_i^l - \lambda_i^{l*}}{\alpha_l} \rightarrow -f_i'(x^*)^T dx + (d\lambda)_i. \tag{7.10}
$$

Since (7.1) implies

$$
\lim_{l \to \infty} \frac{\|H_0(w^l)\|}{\alpha_l} \leq \lim_{l \to \infty} \frac{1}{l} \times \frac{w^l - w^{l*}}{\alpha_l} = 0,
$$

by (7.7), (7.9), and (7.10) we have

$$
\left(P^0 + \sum_{i \in \mathcal{B}} \lambda_i^{l*} P^i \right) dx + \sum_{i \in \mathcal{B}} f_i'(x^*)(d\lambda)_i = 0,
$$

$$
f_{\mathcal{B}}'(x^*)^T dx = 0, \tag{7.11}
$$

$$
-f_{\mathcal{N}}'(x^*)^T dx + (d\lambda)_{\mathcal{N}} = 0.
$$

From the first two equations in (7.11) we have

$$(dx)^T \left(P^0 + \sum_{i \in \mathcal{B}} \lambda_i^{l*} P^i \right) dx = - \sum_{i \in \mathcal{B}} (dx)^T f_i'(x^*)(d\lambda)_i = 0.$$

Thus, by Assumption 5.2, we obtain $dx = 0$. This and the third equation in (7.11) yield $(d\lambda)_{\mathcal{N}} = 0$. Thus, for any sufficiently small $\varepsilon_0 > 0$, we have $(\lambda^* + \varepsilon_0 d\lambda)_+ = (\lambda^l)_+$, and hence

$$\nabla f_0(x^* + \varepsilon_0 dx) + f'(x^* + \varepsilon_0 dx)^T (\lambda^* + \varepsilon_0 d\lambda)_+ = \nabla f_0(x^*) + f'(x^*)^T (\lambda^*)_+ = 0$$

and

$$-f(x^* + \varepsilon_0 dx) + \lambda^* + \varepsilon_0 d\lambda - (\lambda^* + \varepsilon_0 d\lambda)_+ = -f(x^*) + \lambda^* - (\lambda^*)_+ = 0,$$

i.e., $H_0(x^* + \varepsilon_0 dx, \lambda^* + \varepsilon_0 d\lambda) = 0$. This shows that (7.5) holds.

Therefore, we obtain $(dx, d\lambda) = 0$. This complemetes the proof. $\qquad\square$

Appendix II

Proof of Lemma 5.3: Since the strictly complementary condition holds, i.e., $|\lambda_i^*| > 0$ for all $i \in \mathcal{J}$, it follows that there is a constant $\varepsilon > 0$ such that $|\lambda_i^k| \geq \varepsilon$ for all $i \in \mathcal{J}$ and all λ^k sufficiently close to λ^*. Thus, by (2.7) we have that, for all z^k sufficiently close to z^*,

$$|(\Phi_i(y^k))_\mu'| = 2\mu_k / \sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2} \leq 2\mu_k / \sqrt{\varepsilon^2} = O(\mu_k). \tag{7.12}$$

Furthermore, by combining (2.10), (2.11), (3.2), (3.3), (3.5), (7.12), and the fact that $z^k \in \Omega$, we have for all z^k sufficiently close to z^* that

$$\begin{aligned}
\|u_1(z^k)\| &\leq \|f'(x^k)\| \|\hat{u}_1(z^k)\| + \|\tilde{u}_1(z^k)\| \\
&\leq \|f'(x^k)\| \left(\|(\Phi(y^k))_\mu'\| \mu_0 \beta(z^k) + \frac{1}{2}\mu_k \|(\Phi(y^k))_\mu'\| \right) \\
&\quad + g_1(\mu_k)\|x^k\| + |g_1'(\mu_k)|(\mu_k + \mu_0\beta(z^k))\|x^k\| \\
&\leq \|f'(x^k)\| \left(\|(\Phi(y^k))_\mu'\| \mu_k + \frac{1}{2}\mu_k \|(\Phi(y^k))_\mu'\| \right) \\
&\quad + (g_1(\mu_k) + 2\mu_k|g_1'(\mu_k)|)\|x^k\| \\
&= \frac{3}{2}\mu_k \|f'(x^k)\| \left\| \text{vec}\left\{ 2\mu_k \Big/ \sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2} : i \in \mathcal{J} \right\} \right\| + O((\mu_k)^2) \\
&= O((\mu_k)^2). \tag{7.13}
\end{aligned}$$

In addition, by combining (2.10), (2.11), (3.2), (3.4), and (3.6), we have

$$\|u_2(z^k)\| \leq \|\hat{u}_2(z^k)\| + \|\tilde{u}_2(z^k)\| = O((\mu_k)^2) \tag{7.14}$$

for all z^k sufficiently close to z^*.

Since $\mu_k \to 0$ as $k \to \infty$, (7.13) and (7.14) imply that $\|u(z^k)\| \le \tau \sqrt{n+m} \mu_k$ for all z^k sufficiently close to z^*. Therefore, by noting the definition of $v(\cdot)$ (see (3.8)), we have $v(z^k) = u(z^k)$ for all z^k sufficiently close to z^*. This completes the proof. $\square$

Appendix III

Proof of Lemma 5.4: By (2.6), we have $\Phi_i(y^k) > 0$ for all $i \in \mathcal{J}$, and hence by (5.12), we have $t_i^k > 0$ for all $i \in \mathcal{J}$. This, together with Assumption 2.1, implies that $P^0 + \sum_{i \in \mathcal{J}} t_i^k P^i$ is positive semidefinite, and so is $M(z^k)$ by (2.13), (2.10), and (2.11). Thus, from the first two equations of (5.23) it follows that

$$(R(z^k)^{-1}\widehat{\Delta t}^k)^T \widehat{\Delta r}^k = (\widehat{\Delta t}^k)^T R(z^k)^{-1} \widehat{\Delta r}^k = (\widehat{\Delta t}^k)^T (-f'(x^k)\widehat{\Delta x}^k)$$
$$= -(f'(x^k)^T \widehat{\Delta t}^k)^T \widehat{\Delta x}^k = (\widehat{\Delta x}^k)^T M(z^k)^T \widehat{\Delta x}^k \ge 0.$$

From the third equation of (5.23) we have

$$\min_{i \in \mathcal{J}}[(\Phi(y^k))'_\lambda N(z^k)]_{ii} \| R(z^k)^{-1/2}[(\Phi(y^k))'_\lambda]^{-1}\widehat{\Delta t}^k\|^2$$
$$\le [(R(z^k)^{-1}\widehat{\Delta t}^k)^T[(\Phi(y^k))'_\lambda]^{-1}][N(z^k)(\Phi(y^k))'_\lambda][[(\Phi(y^k))'_\lambda]^{-1}\widehat{\Delta t}^k]$$
$$= (R(z^k)^{-1}\widehat{\Delta t}^k)^T[(\Phi(y^k))'_\lambda]^{-1}N(z^k)\widehat{\Delta t}^k$$
$$\le (R(z^k)^{-1}\widehat{\Delta t}^k)^T[(\Phi(y^k))'_\lambda]^{-1}N(z^k)\widehat{\Delta t}^k + (R(z^k)^{-1}\widehat{\Delta t}^k)^T\widehat{\Delta r}^k$$
$$= (R(z^k)^{-1}\widehat{\Delta t}^k)^T[(\Phi(y^k))'_\lambda]^{-1}[N(z^k)\widehat{\Delta t}^k + (\Phi(y^k))'_\lambda\widehat{\Delta r}^k]$$
$$= (R(z^k)^{-1}\widehat{\Delta t}^k)^T[(\Phi(y^k))'_\lambda]^{-1}[N(z^k)h_2(z^k) + (\Phi(y^k))'_\lambda h_3(z^k)]$$
$$\le \| R(z^k)^{-1/2}[(\Phi(y^k))'_\lambda]^{-1}\widehat{\Delta t}^k\|\|\check{h}(z^k)\|,$$

where $\check{h}(z^k) := R(z^k)^{-1/2}[N(z^k)h_2(z^k) + (\Phi(y^k))'_\lambda h_3(z^k)]$. Since

$$\|\check{h}(z^k)\| \le \|[R(z^k)^{-1/2}(N(z^k)h_2(z^k) + (\Phi(y^k))'_\lambda h_3(z^k))]_\mathcal{B}\|$$
$$+ \|[R(z^k)^{-1/2}(N(z^k)h_2(z^k) + (\Phi(y^k))'_\lambda h_3(z^k))]_\mathcal{N}\|$$
$$\le \| R(z^k)^{-1/2}\|(\|[N(z^k)]_{\mathcal{B}\mathcal{B}}\|\|[h_2(z^k)]_\mathcal{B}\| + \|[(\Phi(y^k))'_\lambda]_{\mathcal{B}\mathcal{B}}\|\|[h_3(z^k)]_\mathcal{B}\|)$$
$$+ \| R(z^k)^{-1/2}\|(\|[N(z^k)]_{\mathcal{N}\mathcal{N}}\|\|[h_2(z^k)]_\mathcal{N}\|$$
$$+ \|[(\Phi(y^k))'_\lambda]_{\mathcal{N}\mathcal{N}}\|\|[h_3(z^k)]_\mathcal{N}\|),$$

we further obtain

$$\| R(z^k)^{-1/2}[(\Phi(y^k))'_\lambda]^{-1}\widehat{\Delta t}^k\|$$
$$\le \| R(z^k)^{-1/2}\|\frac{\|[N(z^k)]_{\mathcal{B}\mathcal{B}}\|\|[h_2(z^k)]_\mathcal{B}\| + \|[(\Phi(y^k))'_\lambda]_{\mathcal{B}\mathcal{B}}\|\|[h_3(z^k)]_\mathcal{B}\|}{\min_{i \in \mathcal{J}}[(\Phi(y^k))'_\lambda N(z^k)]_{ii}}$$
$$+ \| R(z^k)^{-1/2}\|\frac{\|[N(z^k)]_{\mathcal{N}\mathcal{N}}\|\|[h_2(z^k)]_\mathcal{N}\| + \|[(\Phi(y^k))'_\lambda]_{\mathcal{N}\mathcal{N}}\|\|[h_3(z^k)]_\mathcal{N}\|}{\min_{i \in \mathcal{J}}[(\Phi(y^k))'_\lambda N(z^k)]_{ii}}.$$

$$(7.15)$$

Similarly,

$$\|R(z^k)^{-1/2}[N(z^k)]^{-1}\widehat{\Delta r}^k\|$$
$$\leq \|R(z^k)^{-1/2}\|\frac{\|[N(z^k)]_{\mathcal{BB}}\|\|[h_2(z^k)]_{\mathcal{B}}\| + \|[(\Phi(y^k))'_\lambda]_{\mathcal{BB}}\|\|[h_3(z^k)]_{\mathcal{B}}\|}{\min_{i\in\mathcal{J}}[(\Phi(y^k))'_\lambda N(z^k)]_{ii}}$$
$$+\|R(z^k)^{-1/2}\|\frac{\|[N(z^k)]_{\mathcal{NN}}\|\|[h_2(z^k)]_{\mathcal{N}}\| + \|[(\Phi(y^k))'_\lambda]_{\mathcal{NN}}\|\|[h_3(z^k)]_{\mathcal{N}}\|}{\min_{i\in\mathcal{J}}[(\Phi(y^k))'_\lambda N(z^k)]_{ii}}.$$
$$(7.16)$$

Using (2.7), we have

$$\|[(\Phi(y^k))'_\lambda]_{\mathcal{BB}}\| = O(1), \quad \|[(\Phi(y^k))'_\lambda]_{\mathcal{NN}}\| = O((\mu_k)^2), \tag{7.17}$$

and

$$\|[E^m - (\Phi(y^k))'_\lambda]_{\mathcal{BB}}\| = O((\mu_k)^2), \quad \|[E^m - (\Phi(y^k))'_\lambda]_{\mathcal{NN}}\| = O(1).$$

The latter, together with (2.16), (2.10), and (2.11), implies that

$$\|[N(z^k)]_{\mathcal{BB}}\| = O((\mu_k)^2) \quad \text{and} \quad \|[N(z^k)]_{\mathcal{NN}}\| = O(1). \tag{7.18}$$

In addition, using (2.16), (2.5), (2.6), (2.10), and (2.11), we have

$$\min_{i\in\mathcal{J}}[(\Phi(y^k))'_\lambda N(z^k)]_{ii}$$
$$= \min_{i\in\mathcal{J}}[(\Phi(y^k))'_\lambda(E^m - (\Phi(y^k))'_\lambda + g_2(\mu_k)E^m + g_3(\mu_k)(\Phi(y^k))'_\lambda\Phi(\mu_k, -f(x^k)))]_{ii}$$
$$= \min_{i\in\mathcal{J}}\left\{(\Phi_i(y^k))'_\lambda(\lambda_i^k - \Phi_i(y^k))'_\lambda + (\Phi_i(y^k))'_\lambda g_2(\mu_k)\right.$$
$$\left.+g_3(\mu_k)[(\Phi_i(y^k))'_\lambda]^2\Phi_i(\mu_k, -f(x^k))\right\}$$
$$\geq \min_{i\in\mathcal{J}}\left\{(\Phi_i(y^k))'_\lambda(\lambda_i^k - \Phi_i(y^k))'_\lambda\right\}$$
$$= \Theta((\mu_k)^2). \tag{7.19}$$

Thus, in order to give the upper bounds of the right-hand side of (7.15) and (7.16), we need to discuss the upper bounds of $\|[h_2(z^k)]_{\mathcal{B}}\|$, $\|[h_2(z^k)]_{\mathcal{N}}\|$, $\|[h_3(z^k)]_{\mathcal{B}}\|$, and $\|[h_3(z^k)]_{\mathcal{N}}\|$. The discussion can be divided into the following two cases:

Case 1. Consider the upper bounds of $\|[h_2(z^k)]_{\mathcal{B}}\|$ and $\|[h_2(z^k)]_{\mathcal{N}}\|$. Notice that

$$h_2(z^k) = \Phi(y^k) - \Phi(y^{k*}) + (\Phi(y^k))'_\mu\Delta\mu_k - \hat{u}_1(z^k)$$
$$= \Phi(y^k) - \Phi(y^{k*}) - (\Phi(y^k))'_\mu\mu_k + (\Phi(y^k))'_\mu\mu_0\beta(z^k) - \hat{u}_1(z^k)$$
$$= \Phi(y^k) - \Phi(y^{k*}) - \frac{1}{2}(\Phi(y^k))'_\mu\mu_k, \tag{7.20}$$

where the first equality follows from (5.20), the second equality from (5.8), and the last equality from (3.3). Thus, it follows from (7.20) that for any $i \in \mathcal{B}$ and for all z^k sufficiently close to z^*,

$$
\begin{aligned}
|[h_2(z^k)]_i| &\leq |\Phi_i(y^k) - \Phi_i(y^{k*})| + \left| \frac{1}{2}(\Phi_i(y^k))'_\mu \mu_k \right| \\
&= \left| \frac{\lambda_i^k + \sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}}{2} - \frac{\lambda_i^{k*} + \sqrt{\left(\lambda_i^{k*}\right)^2}}{2} \right| + \frac{(\mu_k)^2}{\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}} \\
&= \left| \frac{\lambda_i^k - \lambda_i^{k*}}{2} + \frac{\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2} - \sqrt{\left(\lambda_i^{k*}\right)^2}}{2} \right| + \frac{(\mu_k)^2}{\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}} \\
&\leq \frac{1}{2}|\lambda_i^k - \lambda_i^{k*}| + \frac{|\lambda_i^k - \lambda_i^{k*}||\lambda_i^k + \lambda_i^{k*}| + 4(\mu_k)^2}{2\left(\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2} + \sqrt{\left(\lambda_i^{k*}\right)^2}\right)} + \frac{(\mu_k)^2}{\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}} \\
&= O\left(|\lambda_i^k - \lambda_i^{k*}|\right) + O((\mu_k)^2),
\end{aligned}
$$

which implies

$$
\|[h_2(z^k)]_\mathcal{B}\| = O(\theta(z^k)); \tag{7.21}
$$

and for $i \in \mathcal{N}$ and for all z^k sufficiently close to z^*,

$$
\begin{aligned}
[h_2(z^k)]_i &= \Phi_i(y^k) - \frac{1}{2}(\Phi_i(y^k))'_\mu \mu_k = \frac{\lambda_i^k + \sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}}{2} - \frac{(\mu_k)^2}{\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}} \\
&= \frac{2(\mu_k)^2}{\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2} - \lambda_i^k} - \frac{(\mu_k)^2}{\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}} \\
&= \frac{2\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2} + \lambda_i^k - \sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}}{\left(\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2} - \lambda_i^k\right)\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}}(\mu_k)^2 \\
&= \frac{4(\mu_k)^2}{\left(\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2} - \lambda_i^k\right)^2 \sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}}(\mu_k)^2,
\end{aligned}
$$

which implies

$$
\|[h_2(z^k)]_\mathcal{N}\| = O((\mu_k)^2 \theta(z^k)). \tag{7.22}
$$

Case 2. Consider the upper bounds of $\|[h_3(z^k)]_\mathcal{B}\|$ and $\|[h_3(z^k)]_\mathcal{N}\|$. Notice that

$$
\begin{aligned}
h_3(z^k) &= \lambda^{k*} - \Phi(y^{k*}) - (\lambda^k - \Phi(y^k)) + (\Phi(y^k))'_\mu \Delta\mu_k + \hat{u}_2(z^k) \\
&= \lambda^{k*} - \Phi(y^{k*}) - (\lambda^k - \Phi(y^k)) - (\Phi(y^k))'_\mu \mu_k + (\Phi(y^k))'_\mu \mu_0 \beta(z^k) + \hat{u}_2(z^k) \\
&= \lambda^{k*} - \Phi(y^{k*}) - (\lambda^k - \Phi(y^k)) - \frac{1}{2}(\Phi(y^k))'_\mu \mu_k, \qquad (7.23)
\end{aligned}
$$

where the first equality follows from (5.21), the second equality from (5.8), and the last equality from (3.4). Thus, it follows from (7.23) that for any $i \in \mathcal{B}$,

$$
\begin{aligned}
[h_3(z^k)]_i &= -\left(\lambda_i^k - \Phi_i(y^k)\right) - \frac{1}{2}(\Phi_i(y^k))'_\mu \mu_k \\
&= \frac{-\lambda_i^k + \sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}}{2} - \frac{(\mu_k)^2}{\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}} \\
&= \frac{2(\mu_k)^2}{\lambda_i^k + \sqrt{\left(\lambda_i^k\right)^2 - 4(\mu_k)^2}} - \frac{(\mu_k)^2}{\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}} \\
&= \frac{2\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2} - \lambda_i^k - \sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}}{\left(\lambda_i^k + \sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}\right)\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}}(\mu_k)^2 \\
&= \frac{(\mu_k)^2}{\left(\lambda_i^k + \sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}\right)^2\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}}(\mu_k)^2,
\end{aligned}
$$

which implies

$$
\|[h_3(z^k)]_\mathcal{B}\| = O((\mu_k)^2\theta(z^k)); \qquad (7.24)
$$

and for $i \in \mathcal{N}$, it follows from (7.23) that

$$
\begin{aligned}
|[h_3(z^k)]_i| &\leq |\lambda^{k*} - \Phi(y^{k*}) - (\lambda^k - \Phi(y^k))| + \left|\frac{1}{2}(\Phi_i(y^k))'_\mu \mu_k\right| \\
&= \left|\frac{\lambda_i^{k*} - \lambda_i^k}{2} + \frac{\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2} - \sqrt{\left(\lambda_i^{k*}\right)^2}}{2}\right| + \frac{(\mu_k)^2}{\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}} \\
&\leq \frac{1}{2}|\lambda_i^{k*} - \lambda_i^k| + \frac{|\lambda_i^k - \lambda_i^{k*}||\lambda_i^k + \lambda_i^{k*}| + 4(\mu_k)^2}{2\left(\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2} + \sqrt{\left(\lambda_i^{k*}\right)^2}\right)} + \frac{(\mu_k)^2}{\sqrt{\left(\lambda_i^k\right)^2 + 4(\mu_k)^2}} \\
&= O\left(|\lambda_i^k - \lambda_i^{k*}|\right) + O((\mu_k)^2),
\end{aligned}
$$

which implies

$$\|[h_3(z^k)]_\mathcal{N}\| = O(\theta(z^k)). \tag{7.25}$$

In addition, it is easy to see that, for each $i \in \mathcal{J}$ and each $k \geq 0$, $\Phi_i(z^k)g_3(\mu_k) \geq 0$ and $\Phi_i(z^k)g_3(\mu_k)$ is uniformly bounded above for all z^k sufficiently close to z^*. Thus, for each $i \in \mathcal{J}$, we can obtain from (2.15) that there exists a scalar $c_i > 0$ such that

$$c_i \leq [R(z^k)^{-1}]_{ii} \leq 1 \tag{7.26}$$

for all z^k sufficiently close to z^*.

Now, by combining (7.17)–(7.19) with (7.21), (7.24), and (7.26), we have

$$\frac{\|[N(z^k)]_{\mathcal{BB}}\|\|[h_2(z^k)]_\mathcal{B}\| + \|[(\Phi(y^k))'_\lambda]_{\mathcal{BB}}\|\|[h_3(z^k)]_\mathcal{B}\|}{\min_{i \in \mathcal{J}}[(\Phi(y^k))'_\lambda N(z^k)]_{ii}} = O(\theta(z^k)); \tag{7.27}$$

and by combining (7.17)–(7.19) with (7.22) and (7.25) we have

$$\frac{\|[N(z^k)]_{\mathcal{NN}}\|\|[h_2(z^k)]_\mathcal{N}\| + \|[(\Phi(y^k))'_\lambda]_{\mathcal{NN}}\|\|[h_3(z^k)]_\mathcal{N}\|}{\min_{i \in \mathcal{J}}[(\Phi(y^k))'_\lambda N(z^k)]_{ii}} = O(\theta(z^k)). \tag{7.28}$$

On the one hand, (7.15), together with (7.27) and (7.28), implies that

$$\|R(z^k)^{-1/2}[(\Phi(y^k))'_\lambda]^{-1}\widehat{\Delta t}^k\| = O(\theta(z^k)),$$

which yields,

$$\begin{aligned}
\|\widehat{\Delta t}^k\| &= \|(\Phi(y^k))'_\lambda R(z^k)^{1/2} R(z^k)^{-1/2}[(\Phi(y^k))'_\lambda]^{-1}\widehat{\Delta t}^k\| \\
&\leq \|(\Phi(y^k))'_\lambda\|\|R(z^k)^{1/2}\|\|R(z^k)^{-1/2}[(\Phi(y^k))'_\lambda]^{-1}\widehat{\Delta t}^k\| \\
&= O(\theta(z^k)).
\end{aligned}$$

On the other hand, (7.16), together with (7.27) and (7.28), implies that

$$\|R(z^k)^{-1/2}[N(z^k)]^{-1}\widehat{\Delta r}^k\| = O(\theta(z^k)),$$

and hence,

$$\begin{aligned}
\|\widehat{\Delta r}^k\| &= \|N(z^k)R(z^k)^{1/2} R(z^k)^{-1/2}[N(z^k)]^{-1}\widehat{\Delta r}^k\| \\
&\leq \|N(z^k)\|\|R(z^k)^{1/2}\|\|R(z^k)^{-1/2}[N(z^k)]^{-1}\widehat{\Delta r}^k\| \\
&= O(\theta(z^k)).
\end{aligned}$$

The proof is completed. $\square$

Proof of Lemma 5.5: By the first two equations of (5.23) we have

$$\begin{bmatrix} M(z^k) & f'_\mathcal{B}(x^k)^T \\ R(z^k)_{\mathcal{BB}}(-f'_\mathcal{B}(x^k)) & 0 \end{bmatrix} \begin{bmatrix} \widehat{\Delta x}^k \\ \widehat{\Delta t}^k_\mathcal{B} \end{bmatrix} = \begin{bmatrix} -f'_\mathcal{N}(x^k)^T \widehat{\Delta t}^k_\mathcal{N} \\ \widehat{\Delta r}^k_\mathcal{B} \end{bmatrix},$$

and hence

$$
\begin{bmatrix} P^0 + \sum_{i\in\mathcal{B}} \lambda_i^{k*} P^i & f_{\mathcal{B}}'(x^{k*})^T \\ -f_{\mathcal{B}}'(x^{k*}) & 0 \end{bmatrix} \begin{bmatrix} \widehat{\Delta x}^k \\ \widehat{\Delta t}_{\mathcal{B}}^k \end{bmatrix}
$$

$$
= \begin{bmatrix} M(z^k) & f_{\mathcal{B}}'(x^k)^T \\ R(z^k)_{\mathcal{B}\mathcal{B}}(-f_{\mathcal{B}}'(x^k)) & 0 \end{bmatrix} \begin{bmatrix} \widehat{\Delta x}^k \\ \widehat{\Delta t}_{\mathcal{B}}^k \end{bmatrix}
$$

$$
+ \begin{bmatrix} P^0 + \sum_{i\in\mathcal{B}} \lambda_i^{k*} P^i - M(z^k) & f_{\mathcal{B}}'(x^{k*})^T - f_{\mathcal{B}}'(x^k)^T \\ -f_{\mathcal{B}}'(x^{k*}) + R(z^k)_{\mathcal{B}\mathcal{B}} f_{\mathcal{B}}'(x^k) & 0 \end{bmatrix} \begin{bmatrix} \widehat{\Delta x}^k \\ \widehat{\Delta t}_{\mathcal{B}}^k \end{bmatrix}
$$

$$
= \begin{bmatrix} -f_{\mathcal{N}}'(x^k)^T \widehat{\Delta t}_{\mathcal{N}}^k \\ \widehat{\Delta r}_{\mathcal{B}}^k \end{bmatrix}
$$

$$
+ \begin{bmatrix} \mathcal{P} & (P^{\mathcal{B}}(x^{k*} - x^k))^T \\ -P^{\mathcal{B}}(x^{k*} - x^k) + Q_{\mathcal{B}\mathcal{B}}(z^k)(-f_{\mathcal{B}}'(x^k)) & 0 \end{bmatrix} \begin{bmatrix} \widehat{\Delta x}^k \\ \widehat{\Delta t}_{\mathcal{B}}^k \end{bmatrix},
\tag{7.29}
$$

where $M(\cdot)$ is defined by (2.13),

$$
\mathcal{P} := P^0 + \sum_{i\in\mathcal{B}} \lambda_i^{k*} P^i - M(z^k) = \sum_{i\in\mathcal{B}} P^i\left(t_i^k - t_i^{k*}\right) - \sum_{i\in\mathcal{N}} P^i t_i^k - g_1(\mu_k) E^n,
$$

and $P^{\mathcal{B}}(x^{k*} - x^k)$ denotes the matrix whose i-th row is $(P^i(x^{k*} - x^k))^T$, $i \in \mathcal{B}$. By using (5.15), (5.16), (5.12), (2.10), and (2.11), for all z^k sufficiently close to z^* we have

$$
\begin{aligned}
&\left\| t_{\mathcal{B}}^{k*} - t_{\mathcal{B}}^k \right\| = O(\theta(z^k)), \quad \left\| t_{\mathcal{N}}^k \right\| = O(\theta(z^k)), \\
&\| x^{k*} - x^k \| = O(\theta(z^k)), \quad g_2(\mu_k) = O((\mu_k)^2) = O(\theta(z^k)), \\
&\| Q_{\mathcal{B}\mathcal{B}}(z^k)(-f_{\mathcal{B}}(x^k)) \| = O(g_3(\mu_k)) = O((\mu_k)^2) = O(\theta(z^k)).
\end{aligned}
\tag{7.30}
$$

Similar to the proof given in [27, Lemma 5.2], by partitioning $\widehat{\Delta x}^k$ into its components in $\mathrm{Ker}\, f_{\mathcal{B}}'(x^*)$ and $\mathrm{Ran}\, f_{\mathcal{B}}'(x^*)$, it follows from Assumption 5.2, (7.29), (7.30), and Lemma 5.4 that $\|\widehat{\Delta x}^k\|$ is bounded by the right-hand side of (7.29), i.e.,

$$
\|\widehat{\Delta x}^k\| = O(\theta(z^k)) + O(\theta(z^k))\|\widehat{\Delta x}^k\|.
$$

Since, by Theorem 3.1, $\theta(z^k) \to 0$ as $k \to \infty$, it follows from the above equality that $\|\widehat{\Delta x}^k\| = O(\theta(z^k))$ for all z^k sufficiently close to z^*. This completes the proof. $\qquad \square$

Appendix IV

Proof of Proposition 5.1: Let $M(\cdot)$ and $N(\cdot)$ be defined by (2.13) and (2.16), respectively. By eliminating $\widetilde{\Delta r}^k$ and $\widetilde{\Delta t}^k_{\mathcal{N}}$ from (5.24), we have

$$
\begin{bmatrix} M(z^k) & f'_{\mathcal{B}}(x^k)^T \\ R(z^k)_{\mathcal{B}\mathcal{B}}(-f'_{\mathcal{B}}(x^k)) & 0 \end{bmatrix} \begin{bmatrix} \widetilde{\Delta x}^k \\ \widetilde{\Delta t}^k_{\mathcal{B}} \end{bmatrix}
$$

$$
= \begin{bmatrix} h_0(z^k) - f'_{\mathcal{N}}(x^k)^T((\Phi(y^k))'_\lambda[N(z^k)]^{-1}R(z^k))_{\mathcal{N}\mathcal{N}}(R(z^k)_{\mathcal{N}\mathcal{N}}f'_{\mathcal{N}}(x^k)\widetilde{\Delta x}^k + (h_1(x^k))_{\mathcal{N}}) \\ (h_1(x^k))_{\mathcal{B}} - ([(\Phi(y^k))'_\lambda]^{-1}N(z^k))_{\mathcal{B}\mathcal{B}}\widetilde{\Delta t}^k_{\mathcal{B}} \end{bmatrix}
$$

$$
= \begin{bmatrix} -f'_{\mathcal{N}}(x^k)^T((\Phi(y^k))'_\lambda[N(z^k)]^{-1})_{\mathcal{N}\mathcal{N}}R(z^k)_{\mathcal{N}\mathcal{N}}f'_{\mathcal{N}}(x^k) \\ -([(\Phi(y^k))'_\lambda]^{-1}N(z^k))_{\mathcal{B}\mathcal{B}} \end{bmatrix} \begin{bmatrix} \widetilde{\Delta x}^k \\ \widetilde{\Delta t}^k_{\mathcal{B}} \end{bmatrix}
$$

$$
+ \begin{bmatrix} h_0(z^k) - f'_{\mathcal{N}}(x^k)^T((\Phi(y^k))'_\lambda[N(z^k)]^{-1})_{\mathcal{N}\mathcal{N}}(h_1(x^k))_{\mathcal{N}} \\ (h_1(x^k))_{\mathcal{B}} \end{bmatrix},
$$

which implies

$$
\begin{bmatrix} P^0 + \sum_{i\in\mathcal{B}}\lambda_i^{k*}P^i & f'_{\mathcal{B}}(x^{k*})^T \\ -f'_{\mathcal{B}}(x^{k*}) & 0 \end{bmatrix} \begin{bmatrix} \widetilde{\Delta x}^k \\ \widetilde{\Delta t}^k_{\mathcal{B}} \end{bmatrix}
$$

$$
= \begin{bmatrix} M(z^k) & f'_{\mathcal{B}}(x^k)^T \\ R(z^k)_{\mathcal{B}\mathcal{B}}(-f'_{\mathcal{B}}(x^k)) & 0 \end{bmatrix} \begin{bmatrix} \widetilde{\Delta x}^k \\ \widetilde{\Delta t}^k_{\mathcal{B}} \end{bmatrix}
$$

$$
+ \begin{bmatrix} P^0 + \sum_{i\in\mathcal{B}}\lambda_i^{k*}P^i - M(z^k) & f'_{\mathcal{B}}(x^{k*})^T - f'_{\mathcal{B}}(x^k)^T \\ -f'_{\mathcal{B}}(x^{k*}) + R(z^k)_{\mathcal{B}\mathcal{B}}f'_{\mathcal{B}}(x^k) & 0 \end{bmatrix} \begin{bmatrix} \widetilde{\Delta x}^k \\ \widetilde{\Delta t}^k_{\mathcal{B}} \end{bmatrix}
$$

$$
= \begin{bmatrix} -f'_{\mathcal{N}}(x^k)^T((\Phi(y^k))'_\lambda[N(z^k)]^{-1})_{\mathcal{N}\mathcal{N}}R(z^k)_{\mathcal{N}\mathcal{N}}f'_{\mathcal{N}}(x^k) \\ -([(\Phi(y^k))'_\lambda]^{-1}N(z^k))_{\mathcal{B}\mathcal{B}} \end{bmatrix} \begin{bmatrix} \widetilde{\Delta x}^k \\ \widetilde{\Delta t}^k_{\mathcal{B}} \end{bmatrix}
$$

$$
+ \begin{bmatrix} h_0(z^k) - f'_{\mathcal{N}}(x^k)^T((\Phi(y^k))'_\lambda[N(z^k)]^{-1})_{\mathcal{N}\mathcal{N}}(h_1(x^k))_{\mathcal{N}} \\ (h_1(x^k))_{\mathcal{B}} \end{bmatrix}
$$

$$
+ \begin{bmatrix} P^0 + \sum_{i\in\mathcal{B}}\lambda_i^{k*}P^i - M(z^k) & f'_{\mathcal{B}}(x^{k*})^T - f'_{\mathcal{B}}(x^k)^T \\ -f'_{\mathcal{B}}(x^{k*}) + f'_{\mathcal{B}}(x^k) + Q_{\mathcal{B}\mathcal{B}}(z^k)f'_{\mathcal{B}}(x^k) & 0 \end{bmatrix} \begin{bmatrix} \widetilde{\Delta x}^k \\ \widetilde{\Delta t}^k_{\mathcal{B}} \end{bmatrix}.
$$

Similar to the proof given in [27, Lemma 5.2], by partitioning $\widetilde{\Delta x}^k$ into its components in $\mathrm{Ker}\, f'_{\mathcal{B}}(x^*)$ and $\mathrm{Ran}\, f'_{\mathcal{B}}(x^*)$, it follows from Assumption 5.2 that $\|\widetilde{\Delta x}^k\|$ is bounded by the right-hand side of the above inequality, i.e.,

$$\|\widetilde{\Delta x}^k\|$$

$$\leq \left\|\left[\begin{array}{c} -f'_{\mathcal{N}}(x^k)^T((\Phi(y^k))'_\lambda[N(z^k)]^{-1})_{\mathcal{N}\mathcal{N}}R(z^k)_{\mathcal{N}\mathcal{N}}f'_{\mathcal{N}}(x^k) \\ -([(\Phi(y^k))'_\lambda]^{-1}N(z^k))_{\mathcal{B}\mathcal{B}} \end{array}\right]\right\|\left\|\left[\begin{array}{c}\widetilde{\Delta x}^k \\ \widetilde{\Delta t}^k_{\mathcal{B}}\end{array}\right]\right\|$$

$$+\left\|\left[\begin{array}{c} h_0(z^k) - f'_{\mathcal{N}}(x^k)^T\left((\Phi(y^k))'_\lambda[N(z^k)]^{-1}\right)_{\mathcal{N}\mathcal{N}}(h_1(x^k))_{\mathcal{N}} \\ (h_1(x^k))_{\mathcal{B}} \end{array}\right]\right\|$$

$$+\left\|\left[\begin{array}{cc} P^0+\sum_{i\in\mathcal{B}}\lambda_i^{k*}P^i - M(z^k) & f'_{\mathcal{B}}(x^{k*})^T - f'_{\mathcal{B}}(x^k)^T \\ -f'_{\mathcal{B}}(x^{k*})+f'_{\mathcal{B}}(x^k)+Q_{\mathcal{B}\mathcal{B}}(z^k)f'_{\mathcal{B}}(x^k) & 0 \end{array}\right]\right\|\left\|\left[\begin{array}{c}\widetilde{\Delta x}^k \\ \widetilde{\Delta t}^k_{\mathcal{B}}\end{array}\right]\right\|.$$

Thus, we can further obtain that

$$\|\widetilde{\Delta x}^k\| = O(\theta^2(z^k))\left\|\left[\begin{array}{c}\widetilde{\Delta x}^k \\ \widetilde{\Delta t}^k_{\mathcal{B}}\end{array}\right]\right\| + O(\theta(z^k))\left\|\left[\begin{array}{c}\widetilde{\Delta x}^k \\ \widetilde{\Delta t}^k_{\mathcal{B}}\end{array}\right]\right\| + O(\theta^2(z^k))$$

$$= O(\theta^2(z^k)) + O(\theta(z^k))(\|\widetilde{\Delta x}^k\| + \|\widetilde{\Delta t}^k_{\mathcal{B}}\|).$$

Since $\theta(z^k) \to 0$ as $k \to \infty$, it follows from the above relation that

$$\|\widetilde{\Delta x}^k\| = O(\theta^2(z^k)) + O(\theta(z^k))\|\widetilde{\Delta t}^k_{\mathcal{B}}\|.$$

The proof is completed. □

Appendix V

Proof of Lemma 5.6: Let w^{k*} and t^{k*} be defined by (5.15) and (5.16), respectively. From the last equation of (5.24) we have

$$D^k\widetilde{\Delta t}^k + (D^k)^{-1}\widetilde{\Delta r}^k = 0, \tag{7.31}$$

which, together with the second equation of (5.24), implies

$$R(z^k)^{-1}(D^k)^2\widetilde{\Delta t}^k = -R(z^k)^{-1}\widetilde{\Delta r}^k = R(z^k)^{-1}h_1(x^k) + f'(x^k)\widetilde{\Delta x}^k,$$

and hence,

$$\begin{aligned}\|R(z^k)^{-1/2}D^k\widetilde{\Delta t}^k\|^2 &= (\widetilde{\Delta t}^k)^T(R(z^k)^{-1}h_1(x^k) + f'(x^k)\widetilde{\Delta x}^k) \\ &= (\widetilde{\Delta t}^k)^T R(z^k)^{-1}h_1(x^k) + (\widetilde{\Delta t}^k)^T f'(x^k)\widetilde{\Delta x}^k \\ &= (\widetilde{\Delta t}^k)^T R(z^k)^{-1}h_1(x^k) - (\widetilde{\Delta x}^k)^T M(z^k)\widetilde{\Delta x}^k + (\widetilde{\Delta x}^k)^T h_0(z^k) \\ &\leq (\widetilde{\Delta t}^k)^T R(z^k)^{-1}h_1(x^k) + (\widetilde{\Delta x}^k)^T h_0(z^k) \\ &\leq \|\widetilde{\Delta t}^k\|\|R(z^k)^{-1}\|\|h_1(x^k)\| + \|\widetilde{\Delta x}^k\|\|h_0(z^k)\| \\ &\leq O(1)\|\widetilde{\Delta t}^k\|\|h_1(x^k)\| + \|\widetilde{\Delta x}^k\|\|h_0(z^k)\|,\end{aligned}$$

where the first inequality is due to the positive semidefiniteness of $M(z^k)$ and the last inequality due to (7.26). Since from (5.18) and (5.19) it follows that

$$
\begin{aligned}
\|h_0(z^k)\| &= \|(\nabla f_0(x^{k*}) + f'(x^{k*})^T t^{k*}) - (\nabla f_0(x^k) + f'(x^k)^T t^k) \\
&\quad + M(z^k)(x^k - x^{k*}) + f'(x^k)^T (t^k - t^{k*})\| \\
&= O\left(\left\|\begin{pmatrix} x^{k*} - x^k \\ t^{k*} - t^k \end{pmatrix}\right\|^2\right) = O(\theta^2(z^k))
\end{aligned}
$$

and

$$
\|h_1(x^k)\| = \|f(x^*) - f(x) + f'(x)(x^k - x^*)\| = O(\|x^k - x^*\|^2) = O(\theta^2(z^k)),
$$

we can further obtain that

$$
\begin{aligned}
\|R(z^k)^{-1/2} D^k \widetilde{\Delta t}^k\|^2 &= O(\theta^2(z^k))\|\widetilde{\Delta t}^k\| + O(\theta^2(z^k))\|\widetilde{\Delta x}^k\| \\
&\leq O(\theta^2(z^k))\|\widetilde{\Delta t}^k\| + O(\theta^2(z^k))[O(\theta^2(z^k)) + O(\theta(z^k))\|\widetilde{\Delta t}_{\mathcal{B}}^k\|] \\
&= O(\theta^2(z^k))\|\widetilde{\Delta t}^k\| + O(\theta^4(z^k)) \\
&= O(1)\|D^k \widetilde{\Delta t}^k\| + O(\theta^4(z^k)),
\end{aligned}
$$

where the first inequality follows from Proposition 5.1 and the last equality due to $\|(D^k)^{-1}\| = O(1/\mu_k) = O(1/\theta^2(z^k))$. Thus, by (7.26) we have

$$
\|D^k \widetilde{\Delta t}^k\|^2 \leq \|R(z^k)^{1/2}\|\|R(z^k)^{-1/2} D^k \widetilde{\Delta t}^k\|^2 = O(1)\|\|D^k \widetilde{\Delta t}^k\| + O(\theta^4(z^k)).
$$

Furthermore, we obtain from (7.31) that

$$
\|(D^k)^{-1} \widetilde{\Delta r}^k\| = \|D^k \widetilde{\Delta t}^k\| = O(1),
$$

which further implies

$$
\begin{aligned}
\|\widetilde{\Delta t}_{\mathcal{N}}^k\| &\leq \|((D^k)_{\mathcal{N}\mathcal{N}})^{-1}\|\|(D^k)_{\mathcal{N}\mathcal{N}} \widetilde{\Delta t}_{\mathcal{N}}^k\| \leq \|((D^k)_{\mathcal{N}\mathcal{N}})^{-1}\|\|D^k \widetilde{\Delta t}^k\| \\
&= O(\mu_k)O(1) = O(\theta(z^k)),
\end{aligned}
$$

and similarly,

$$
\|\widetilde{\Delta r}_{\mathcal{B}}^k\| \leq \|(D^k)_{\mathcal{B}\mathcal{B}}\|\|((D^k)_{\mathcal{B}\mathcal{B}})^{-1} \widetilde{\Delta r}_{\mathcal{B}}^k\| \leq \|(D^k)_{\mathcal{B}\mathcal{B}}\|\|(D^k)^{-1} \widetilde{\Delta r}^k\| = O(\theta(z^k)).
$$

This completes the proof. $\qquad \square$

Appendix VI

Proof of Proposition 5.2: From the first two equations of (5.24), it is easy to see that $(\widetilde{\Delta x}^k, \widetilde{\Delta t}_{\mathcal{B}}^k, \widetilde{\Delta r}_{\mathcal{N}}^k)$ is a feasible solution to the problem (5.26). Since this problem is

a convex problem with linear constraints, $(\widetilde{\Delta x}^k, \widetilde{\Delta t}_{\mathcal{B}}^k, \widetilde{\Delta r}_{\mathcal{N}}^k)$ is an optimal solution to (5.26) if and only if it satisfies the KKT conditions of (5.26), which can be written as

$$\begin{bmatrix} 0 \\ (D^k)_{\mathcal{BB}}^2 \widetilde{\Delta t}_{\mathcal{B}}^k \\ (D^k)_{\mathcal{NN}}^{-2} \widetilde{\Delta r}_{\mathcal{N}}^k \end{bmatrix} \in \mathrm{Ran} C(z^k),$$

where $C(\cdot)$ is defined by (5.1). By (5.24), it follows that

$$(D^k)_{\mathcal{NN}}^{-2} \widetilde{\Delta r}_{\mathcal{N}}^k = -\widetilde{\Delta t}_{\mathcal{N}}^k,$$

$$\begin{aligned}(D^k)_{\mathcal{BB}}^2 \widetilde{\Delta t}_{\mathcal{B}}^k &= -\widetilde{\Delta r}_{\mathcal{B}}^k = R(z^k)_{\mathcal{BB}} f_{\mathcal{B}}'(x^k) \widetilde{\Delta x}^k + (h_1(x^k))_{\mathcal{B}} \\ &= R(z^k)_{\mathcal{BB}} f_{\mathcal{B}}'(x^k) \widetilde{\Delta x}^k - R(z^k)_{\mathcal{BB}} f_{\mathcal{B}}'(x^k)(x^k - x^{k*}) \\ &\quad - \int_0^1 f_{\mathcal{B}}'(x^k + \alpha(x^{k*} - x^k))(x^{k*} - x^k)d\alpha,\end{aligned}$$

and

$$\begin{aligned}0 &= -[M(z^k)\widetilde{\Delta x}^k + f'(x^k)^T \widetilde{\Delta t}^k] + h_0(z^k) \\ &= -[M(z^k)\widetilde{\Delta x}^k + f'(x^k)^T \widetilde{\Delta t}^k] + M(z^k)(x^k - x^{k*}) + f'(x^k)^T(t^k - t^{k*}) \\ &\quad + \int_0^1 \{M^0(z^k + \alpha(z^{k*} - z^k))(x^{k*} - x^k) + f'(x^k + \alpha(x^{k*} - x^k))(t^{k*} - t^k)\}d\alpha.\end{aligned}$$

Let $A(\cdot)$ and $B(\cdot)$ be defined by (5.1). Then

$$\begin{aligned}\begin{bmatrix} 0 \\ (D^k)_{\mathcal{BB}}^2 \widetilde{\Delta t}_{\mathcal{B}}^k \\ (D^k)_{\mathcal{NN}}^{-2} \widetilde{\Delta r}_{\mathcal{N}}^k \end{bmatrix} &= A(z^k) \begin{bmatrix} \widetilde{\Delta x}^k \\ \widetilde{\Delta t}_{\mathcal{N}}^k \end{bmatrix} + A(z^k) \begin{bmatrix} x^{k*} - x^k \\ t^{k*} - t^k \end{bmatrix} \\ &\quad + \int_0^1 B(z^k + \alpha(z^{k*} - z^k)) \begin{bmatrix} x^k - x^{k*} \\ t^k - t^{k*} \end{bmatrix} d\alpha.\end{aligned}$$

Thus, by using Assumption 5.3, we have

$$\begin{bmatrix} 0 \\ (D^k)_{\mathcal{BB}}^2 \widetilde{\Delta t}_{\mathcal{B}}^k \\ (D^k)_{\mathcal{NN}}^{-2} \widetilde{\Delta r}_{\mathcal{N}}^k \end{bmatrix} \in \mathrm{Ran} A(z^k) = \mathrm{Ran} B(z^k + \alpha(z^{k*} - z^k)) = \mathrm{Ran} C(z^k)$$

for all z^k sufficiently close z^*. This completes the proof. $\qquad\qquad\square$

Appendix VII

Proof of Lemma 5.7: By Lemma 5.6, we have that $\max\{\|\widetilde{\Delta t}_{\mathcal{N}}^k\|, \|\widetilde{\Delta r}_{\mathcal{B}}^k\|\} = O(\theta(z^k))$ holds for all z^k sufficiently close to z^*. Since (5.26) is always feasible, there must be a feasible solution to the problem, $\bar{u}$ and $\bar{v}$, such that for all z^k sufficiently close to z^*,

$$\max\{\|\bar{u}\|, \|\bar{v}\|\} = O(\|h_0(z^k)\|) + O(\widetilde{\Delta t}_{\mathcal{N}}^k) + O(\|h_1(x^k)\|) + O(\widetilde{\Delta r}_{\mathcal{B}}^k) = O(\theta(z^k)).$$

Since for all z^k sufficiently close to z^*, $C(z^k)$ is invariable, it follows that, for all z^k sufficiently close to z^*, $\mathrm{Ker}(C(z^k)^T)$ is invariable, which indicates that this matrix has constant rank for all z^k sufficiently close to z^*. Thus, by using Lemma 5.9 in [28] and Proposition 5.2, we have

$$\left\|(D^k)_{\mathcal{BB}}\widetilde{\Delta t}_{\mathcal{B}}^k\right\| + \left\|(D^k)_{\mathcal{NN}}^{-1}\widetilde{\Delta r}_{\mathcal{N}}^k\right\| \leq \left\|(D^k)_{\mathcal{BB}}\bar{u}\right\| + \left\|(D^k)_{\mathcal{NN}}^{-1}\bar{v}\right\|$$
$$\leq \left\|(D^k)_{\mathcal{BB}}\right\|\|\bar{u}\| + \left\|(D^k)_{\mathcal{NN}}^{-1}\right\|\|\bar{v}\|$$
$$= O(\mu_k\theta(z^k)),$$

which further implies that, for all z^k sufficiently close to z^*,

$$\left\|\widetilde{\Delta t}_{\mathcal{B}}^k\right\| + \left\|\widetilde{\Delta r}_{\mathcal{N}}^k\right\| \leq \left\|(D^k)_{\mathcal{BB}}^{-1}\right\|\left\|(D^k)_{\mathcal{BB}}\widetilde{\Delta t}_{\mathcal{B}}^k\right\| + \left\|(D^k)_{\mathcal{NN}}\right\|\left\|(D^k)_{\mathcal{NN}}^{-1}\widetilde{\Delta r}_{\mathcal{N}}^k\right\|$$
$$= O(\mu_k)O(\mu_k\theta(z^k)) = O(\theta(z^k)).$$

This implies that the result of the lemma holds. □

Acknowledgments

References

1. A. Bakushinsky and A. Goncharsky, Ill-Posed Problems: Theory and Applications, Kluwer Academic Publishers, Boston, 1989.
2. F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
3. F. Cole, J.G. Ecker, and W. Gochet, "A reduced gradient method for quadratic programs with quadratic constraints and l_p-constrained l_p-approximation problems," European J. Oper. Res., vol. 9, pp. 194–203, 1982.
4. S. Engelke and C. Kanzow, "Predictor-corrector smoothing methods for the solution of linear programming," Preprint, Department of Mathematics, University of Hamburg, Germany, 2000.
5. S. Engelke and C. Kanzow, "Improved smoothing-type methods for the solution of linear programming," Numer. Math., vol. 90, pp. 487–507, 2002.
6. F. Facchnei, A. Fischer, and C. Kanzow, "On the accurate identification of active constrains," SIAM J. Optim., vol. 9, pp. 14–32, 1998.
7. F. Facchnei and C. Kanzow, "Beyond monotonicity in regularization methods for nonlinear complementarity problems," SIAM J. Control Optim., vol. 37, pp. 1150–1161, 1999.

8. F. Facchinei and J.-S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume II. Springer-Verlag, New York, 2003.

9. A. Fischer, "Solution of monotone complementarity problems with Lipschitzian functions," Math. Program., vol. 76, pp. 513–532, 1997.

10. A. Fischer, "Local beharior of an iterative framework for generalized equations with nonisolated solutions," Math. Program., vol. 94, pp. 91–124, 2002.

11. M.S. Gowda and J.J. Sznajder, "Weak univalence and connectedness of inverse images of continuous functions," Math. Oper. Res., vol. 24, pp. 255–261, 1999.

12. Z.H. Huang, "Sufficient conditions on nonemptiness and boundedness of the solution set of the P_0 function nonlinear complementarity problem," Oper. Res. Letters, vol. 30, pp. 202–210, 2002.

13. Z.H. Huang, J. Han, and Z. Chen, "A predictor-corrector smoothing Newton algorithm, based on a new smoothing function, for solving the nonlinear complementarity problem with a P_0 function," J. Optim. Theory Appl., vol. 117, pp. 39–68, 2003.

14. Z.H. Huang, L. Qi, and D. Sun, "Sub-quadratic convergence of a smoothing Newton algorithm for the P_0- and monotone LCP," Math. Program., vol. 99, pp. 423–441, 2004.

15. S. Karamardian, "Complementarity problems over cones with monotone and pseudomonotone maps," J. Optim. Theory Appl., vol. 18, pp. 445–454, 1976.

16. M. Kojima, "Strongly stable stationary solutions in nonlinear programs," in S.M. Robinson (ed.), Analysis and Computation of Fixed Points, Academic Press, New York, pp. 93–138, 1980.

17. M.S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebert, "Applications of second-order cone programming," Linear Algebra Appl., vol. 284, pp. 193–228, 1998.

18. L. Mclinden, "Stable monotone variational inequalities," Math. Program., vol. 48, pp. 303–338, 1990.

19. R. Mifflin, " Semismooth and semiconvex functions in constrained optimization," SIAM J. Control Optim., vol. 15, pp. 957–972, 1977.

20. S. Mizuno, "A superlinearly convergent infeasible-interior-point algorithm for geometrical LCPs without strictly complementary condition," Math. Oper. Res., vol. 21, pp. 382–400, 1996.

21. Y.E. Nesterov and A.S. Nemirovskii, Interior Point Polynomial Methods in Convex Programming: Theory and Applications, SIAM, Philadelphia, PA, 1994.

22. E. Polak, Optimization: Algorithms and Consistent Approximations, New York, Springer-Verlag, 1997.

23. E. Polak, L. Qi, and D. Sun, "Second-order algorithm for generalized finite and semi-infinite min-max problems," SIAM J. Optim., vol. 11, pp. 937–961, 2001.

24. H.D. Qi, "A regularized smoothing Newton method for box constrained variational inequality problems with P_0-functions," SIAM J. Optim., vol. 10, pp. 315–330, 2000.

25. L. Qi and J. Sun, "A nonsmooth version of Newton's method," Math. Program., vol. 58, pp. 353–367, 1993.

26. L. Qi, D. Sun, and G. Zhou, "A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequality problems," Math. Program., vol. 87, pp. 1–35, 2000.

27. D. Ralph and S.J. Wright, "Superlinear convergence of an interior point method for monotone variational inequalities," Complementarity and Variational Problems: State of the Art, SIAM, pp. 345–385, 1997.

28. D. Ralph and S.J. Wright, "Superlinear convergence of an interior point method despite dependent constraints," Math. Oper. Res., vol. 25, pp. 179–194, 2000.

29. G. Ravindran and M.S. Gowda, "Regularization of P_0-functions in box variational inequality problems," SIAM J. Optim., vol. 11, pp. 748–760, 2001.

30. S.M. Robinson, "Normal maps induced by linear transformation," Math. Oper. Res., vol. 17, pp. 691–714, 1992.

31. J. Stoer, M. Weches, and S. Mizuno, "Hight order infeasible-interior-point methods for solving sufficient linear complementarity problems," Math. Oper. Res., vol. 23, pp. 832–862, 1998.

32. J.F. Sturm, "Superlinear convergence of an algorithm for monotone linear complementarity problems, when no strictly complementary solution exists," Math. Oper. Res., vol. 24, pp. 72–94, 1999.

33. D. Sun, "A regularization Newton method for solving nonlinear complementarity problems," Appl. Math. Optim., vol. 36, pp. 315–339, 1999.

34. J. Sun and G. Zhao, "Quadratic convergence of a long-step interior point method for nonlinear monotone variational inequality problems," J. Optim. Theory Appl., vol. 97, pp. 471–491, 1998.

35. P. Tseng, "Error bounds and superlinear convergence analysis of some Newton-type methods in optimization," in Nonlinear Optimization and Related Topics, Di G. Pillo and F. Giannessi (eds.), Kluwer Academic Publishers, Boston, pp. 445–462, 2000.
36. S.J. Wright, Primal-Dual Interior Methods. SIAM, Philadelphia, 1997.
37. N. Yamashita and M. Fukushima, "On the rate of convergence of the Levenberg-Marquardt method," Computing[Suppl], vol. 15, pp. 239–249, 2001.
38. Y. Ye and K. Anstreicher, "On quadratic and $O(\sqrt{n}L)$ convergence of a predictor-corrector algorithm for LCP," Math. Program., vol. 62, pp. 537–552, 1993.
39. Y. Ye, O. Güler, R.A. Tapia, and Y. Zhang, "A quadratically convergent $O(\sqrt{n}L)$-iteration algorithm for linear programming," Math. Program., vol. 59, pp. 151–162, 1993.
40. Y. Zhang, R.A. Tapia, and J.E. Dennis, "On superlinear and quadratic convergence of primal-dual interior point linear programming algorithms." SIAM J. Optim., vol. 2, pp. 304–324, 1992.
41. G. Zhao and J. Sun, "On the rate of local convergence of high-order infeasible-path-following algorithms for P_*-LCP with or without strictly complementary solutions," Comput. Optim. Appl., vol. 14, pp. 293–307, 1999.