J Glob Optim (2008) 40:679–695
DOI 10.1007/s10898-006-9117-x

ORGINAL PAPER

1.5-Q-superlinear convergence of an exterior-point method for constrained optimization

Igor Griva · Roman A. Polyak

Received: 13 October 2006 / Accepted: 25 October 2006 / Published online: 6 December 2006

Abstract We introduce and analyze an exterior-point method (EPM) for constrained optimization problems with both inequality constraints and equations. We show that under the standard second-order optimality conditions the EPM converges to the primal–dual solution with 1.5-Q-superlinear rate.

Keywords Nonlinear rescaling · Augmented Lagrangian · duality · Primal-dual · Multipliers method

1 Introduction

The exterior-point method (EPM) is based on the nonlinear rescaling-augmented Lagrangian (NRAL) technique, which generalizes the modified barrier-augmented Lagrangian method [2]. The NRAL method uses the nonlinear rescaling technique [8,9] for inequality constraints and the augmented Lagrangian [5,11] for equations. The NR method at each step alternates the unconstrained minimization of the augmented Lagrangian for the equivalent problem in the primal space with both the Lagrange multipliers and scaling-penalty parameter update. This is equivalent to solving the primal–dual system of equations. The application of Newton's method to the primal–dual system leads to the EPM.

Dedicated to Professor Gil Strang on the occasion on his 70th birthday.

I. Griva (✉)
Department of Mathematical Sciences and School of Computational Sciences, George Mason University, Fairfax, VA 22030, USA
e-mail: igriva@gmu.edu

R. A. Polyak
Department of SEOR and Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA
e-mail: rpolyak@gmu.edu

The EPM eliminates the necessity to find the minimizer of the augmented Lagrangian for the equivalent problem at each step. Moreover, the EPM has two basic advantages over the Newton NR method, which consists of using Newton's method for finding an approximation of the primal minimizer followed by the Lagrange multipliers update [8] (see also [2,6]). First, instead of finding the primal approximation and updating the Lagrange multipliers, the EPM performs one Newton step for solving the primal–dual system. Second, a special way to increase of the penalty-barrier parameter leads to a 1.5-Q-superlinear rate of convergence of the EPM in the neighborhood of the solution under the standard second-order optimality conditions.

The EPM is the generalization of the primal–dual NR approach (see [3,4,10]) for problems with both inequality constraints and equations.

The paper is organized as follows. In the next section, we describe the problem and the basic assumptions. In Sect. 3, we consider a class of constraint transformations and the augmented Lagrangian for the equivalent problem. In Sect. 4, we formulate the NRAL method. In Sect. 5, we consider the primal–dual system, describe the EPM and prove a 1.5-Q-superlinear rate of convergence. We conclude the paper by pointing out further directions of research.

2 Statement of the problem and basic assumptions

We consider $p + q + 1$ twice continuously differential functions f, c_i, g_j: $\mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, p, j = 1, \dots, q$ and the feasible set

$$\Omega = \left\{x : c_i(x) \geq 0,\, i = 1, \dots, p; \quad g_j(x) = 0,\, j = 1, \dots, q\right\}.$$

The problem consists of finding

$$x^* \in X^* = \mathrm{Argmin}\{f(x) | x \in \Omega\}.$$

The Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^q \to \mathbb{R}^1$ for problem $(\mathcal{P})$ is given by formula

$$L(x, \lambda, v) = f(x) - \sum_{i=1}^{p} \lambda_i c_i(x) - \sum_{j=1}^{q} v_j g_j(x).$$

We assume that $I^* = \{i : c(x^*) = 0\} = \{1, \dots, r\}$ is the active set of inequality constraints, i.e., $c_i(x^*) = 0, i = 1, \dots, r$. We consider vector functions $c^T(x) = (c_1(x), \dots, c_p(x))$, $c^T_{(r)}(x) = (c_1(x), \dots, c_r(x)), g^T(x) = (g_1(x), \dots, g_q(x))$ and their Jacobians $\nabla c(x) = J(c(x))$, $\nabla c_{(r)}(x) = J\left(c_{(r)}(x)\right)$, and $\nabla g(x) = J(g(x))$ and assume that

$$\mathrm{rank} \begin{pmatrix} \nabla c_{(r)}(x^*) \\ \nabla g(x^*) \end{pmatrix} = r + q < n, \tag{1}$$

i.e., gradients $\nabla c_i(x^*), i = 1, \dots, r$ and $\nabla g_j(x^*), i = 1, \dots, q$ are linearly independent at the solution. Then there exist two vectors $\lambda^* \in \mathbb{R}^p_+$ and $v^* \in \mathbb{R}^q$ such that the K-K-T conditions

$$\nabla_x L(x^*, \lambda^*, v^*) = \nabla f(x^*) - \nabla c^T(x^*)\lambda^* - \nabla g^T(x^*)v^* = 0, \tag{2}$$

$$\lambda^*_i c_i(x^*) = 0, \quad c_i(x^*) \geq 0, \quad \lambda^*_i \geq 0, \quad i = 1, \dots, p, \tag{3}$$

$$g_i(x^*) = 0, \quad i = 1, \dots, q \tag{4}$$

are satisfied.

We assume that

$$\lambda_i^* > 0, \quad i = 1, \dots, r. \tag{5}$$

Let us consider the Hessian of the Largangian of the problem $\mathcal{P}$

$$\nabla_{xx}^2 L(x^*, \lambda^*, v^*) = \nabla^2 f(x^*) - \sum_{i=1}^{p} \lambda_i^* \nabla^2 c_i(x^*) - \sum_{j=1}^{q} v_j^* \nabla^2 g_j(x^*).$$

The sufficient regularity conditions (1) and (5) together with the sufficient condition for a minimum x^* to be isolated

$$\left\langle \nabla_{xx}^2 L(x^*, \lambda^*, v^*)\xi, \xi \right\rangle \ge m\xi^T\xi, \quad \forall \xi : \nabla c_{(r)}(x^*)\xi = 0, \quad \nabla g(x^*)\xi = 0, \quad m > 0 \tag{6}$$

comprise the standard second-order optimality conditions for the problem $\mathcal{P}$.

Let $d : \mathbb{R}_+^p \times \mathbb{R}^q$ be the dual function defined by the formula

$$d(y) = d(\lambda, v) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, v).$$

With the primal problem ($\mathcal{P}$) is associated the dual problem

$$d(y^*) = d(\lambda^*, v^*) = \max \left\{ d(\lambda, v) | \lambda \in \mathbb{R}_+^p,\ v \in \mathbb{R}^q \right\}. \tag{$\mathcal{D}$}$$

The standard second-order optimality conditions guarantee the uniqueness of the primal–dual solution (x^*, y^*) and the absence of the duality gap, i.e., $f(x^*) = d(y^*)$.

In the following, we use the l_∞ vector norm $\|r\| = \max_{1 \le i \le s} |r_i|$, and the corresponding matrix norm $\|Q\| = \max\limits_{1 \le i \le s} \left(\sum\limits_{j=i}^{s} |q_{ij}| \right)$.

Later, we will also use the Lipschitz conditions for the Hessians $\nabla^2 f(x)$, $\nabla^2 c_i(x)$, $i = 1, \dots, p$ and $\nabla^2 g_j(x), j = 1, \dots, q$ in the neighborhood $\Omega_{\varepsilon_0}(x^*) = \{x : \|x - x^*\| \le \varepsilon_0\}$ of the primal solution x^*.

$$\begin{aligned} \|\nabla^2 f(x_1) - \nabla^2 f(x_2)\| &\le L_0 \|x_1 - x_2\|, \\ \|\nabla^2 c_i(x_1) - \nabla^2 c_i(x_2)\| &\le L_i \|x_1 - x_2\|,\ i = 1, \dots, p, \\ \|\nabla^2 g_j(x_1) - \nabla^2 g_j(x_2)\| &\le L_j \|x_1 - x_2\|,\ j = 1, \dots, q. \end{aligned} \tag{7}$$

We conclude the section with the following lemma, which is a slight modification of the Debreu theorem [1].

Lemma 1 *Let A be a symmetric matrix, $B : \mathbb{R}^n \to \mathbb{R}^r$, $\Lambda = \mathrm{diag}(\lambda_i)_{i=1}^r$ with $\lambda_i > 0$ and there is $m > 0$ that $u^T A u \ge m u^T u$, $\forall u : Bu = 0$. Then there exists $k_0 > 0$ large enough that for any $0 < \mu < m$ the inequality*

$$u^T (A + kB^T \Lambda B) u \ge \mu u^T u, \quad \forall u \in \mathbb{R}^n$$

holds for any $k \ge k_0$.

3 Constraint transformations and augmented Lagrangian for an equivalent problem

We consider a class Ψ of concave monotone, increasing and twice continuous differentiable functions $\psi : -\infty \leq t_0 < t < t_1 \leq +\infty \to \mathbb{R}$ that satisfy the following properties

1. $\psi(0) = 0$.
2. $\psi'(t) > 0$.
3. $\psi'(0) = 1$.
4. $\psi''(t) < 0$.
5. (a) $\psi'(t) \leq a(t+1)^{-1}$, (b) $-\psi''(t) \leq b(t+1)^{-2}$, $t \geq 0$, $a > 0$, $b > 0$.

Examples of $\psi \in \Psi$ can be found in [9].

We will use $\psi \in \Psi$ to transform the inequality constraints $c_i(x) \geq 0$, $i = 1, \ldots, p$ into an equivalent set of constraints.

For any fixed $k > 0$ the following problem is equivalent to the original problem $(\mathcal{P})$ due to the properties of $\psi \in \Psi$, i.e., we have

$$x^* \in X^* \\ = \text{Argmin}\{f(x) | k^{-1}\psi(kc_i(x)) \geq 0,\ i = 1, \ldots, p;\ g_j(x) = 0,\ j = 1, \ldots, q\}.$$

For a given $k > 0$, we define the augmented Lagrangian for the equivalent problem $\mathcal{L}_k : \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^q \to \mathbb{R}^1$ by the formula

$$\mathcal{L}_k(x, \lambda, v) = f(x) - k^{-1} \sum_{i=1}^{p} \lambda_i \psi(kc_i(x)) - \sum_{j=1}^{q} v_j g_j(x) + \frac{k}{2} \sum_{j=1}^{q} g_j^2(x). \tag{8}$$

The first two terms define the classical Lagrangian for the equivalent problem in the absence of equality constraints (see [8,9] and references therein). The last two terms coincide with the augmented Lagrangian terms associated with equality constraints (see [5,11]). We would like to note that for any $k > 0$ at the solution the augmented Lagrangian (8) for the equivalent problem has the following useful properties:

1. $\mathcal{L}_k(x^*, \lambda^*, v^*) = f(x^*)$.
2. $\nabla_x \mathcal{L}_k(x^*, \lambda^*, v^*) = \nabla_x L(x^*, \lambda^*, v^*) = 0$.
3. $\nabla^2_{xx} \mathcal{L}_k(x^*, \lambda^*, v^*) = \nabla^2_{xx} L(x^*, \lambda^*, v^*) - k\Psi''(0) \nabla c^T_{(r)}(x^*) \Lambda^*_{(r)} \nabla c_{(r)}(x^*) + k \nabla g^T(x^*)$ $\nabla g(x^*)$.

The following lemma is a direct consequence of the standard second-order optimality conditions (1), (5), and (6) and Lemma 1.

Lemma 2 *If the standard second-order optimality conditions are satisfied then there is* $k_0 > 0$ *large enough such that for any* $k \geq k_0$ *the matrix* $\nabla^2_{xx} \mathcal{L}_k(x^*, \lambda^*, v^*)$ *is positive definite, i.e. there is* $0 < \mu < m$ *that*

$$u^T \nabla^2_{xx} \mathcal{L}_k(x^*, \lambda^*, v^*) u \geq \mu u^T u, \quad \forall u \in \mathbb{R}^n. \tag{9}$$

Let us consider the neighborhood $\Omega_\varepsilon(z^*) = \{z = (x, \lambda, v) : \|z - z^*\| \leq \varepsilon\}$ of the primal–dual solution $z^* = (x^*, \lambda^*, v^*)$. If $f, c_i, g_j \in \mathcal{C}^2$, then the inequality (9) remains true for any $z = (x, \lambda, v) \in \Omega_\varepsilon(z^*)$. In other words, for $k \geq k_0$ the augmented Lagrangian for the equivalent problem $\mathcal{L}_k(x, \lambda, v)$ is strongly convex with respect to x for any $z \in \Omega_\varepsilon(z^*)$.

The problem $(\mathcal{P})$ is a nonconvex optimization problem in $x \in \mathbb{R}^n$, in general. Nevertheless, by Lemma 2 under the standard second-order optimality conditions the augmented Lagrangian for the equivalent problem $\mathcal{L}_k(x, \lambda, v)$ is strongly convex for any fixed $y : (x, y) \in \Omega_\varepsilon(z^*)$ and any $k \geq k_0$. This is not true, in general, for the classical Lagrangian $L(x, \lambda, v)$ for the original problem $(\mathcal{P})$ (see [8]).

The property (9) of the Hessian $\nabla^2_{xx}\mathcal{L}_k(x, \lambda, v)$ remains true in the neighborhood $\Omega_\varepsilon(z^*)$ of the primal–dual solution. Therefore, after finding the primal minimizer of $\mathcal{L}_k(x, \lambda, v)$ for $k \geq k_0$ large enough, at each step the NR method finds the primal minimizer of the strongly convex function followed by the Lagrange multipliers update by the formulas (14) and (15) described below.

In this paper, we replace the primal minimization and dual update by one step of Newton's method for solving the primal–dual system of equations. The properties of the Hessian $\nabla^2_{xx}\mathcal{L}_k(x, \lambda, v)$ in the neighborhood $\Omega_\varepsilon(z^*)$, the smoothness of $f(x)$, $c_i(x)$, $i = 1, \ldots, p$, and $g_j(x)$, $j = 1, \ldots, q$, along with the properties 1^0–5^0 of the transformation $\psi(t)$ provide important properties of the primal–dual system in the neighborhood $\Omega_\varepsilon(z^*)$, which allow to prove a 1.5-Q-superlinear rate of convergence of the EPM.

4 Nonlinear rescaling–augmented Lagrangian multipliers method

In this section we consider the NRAL method for solving problem $(\mathcal{P})$. First, we define the extended dual domain. For $k_0 > 0$ large enough and small enough $\delta > 0$ we consider the following sets

$$D(\lambda^*_{(r)}, k_0, \delta) = \left\{(\lambda_{(r)}, k) : |\lambda_i - \lambda^*_i| \leq k\delta,\ \lambda_i \geq \delta,\ i = 1, \ldots, r,\ k \geq k_0\right\},$$
$$D(\lambda^*_{(p-r)}, k_0, \delta) = \left\{(\lambda_{(p-r)}, k) : 0 \leq \lambda_i \leq k\delta,\ i = r+1, \ldots, p,\ k \geq k_0\right\},$$

and

$$D(v^*, k_0, \delta) = \left\{(v, k) : |v_i - v^*_i| \leq k\delta, \quad i = 1, \ldots, q, \quad k \geq k_0\right\},$$

We define the extended dual domain as follows

$$D(y^*, k_0, \delta) = D(\lambda^*_{(r)}, k_0, \delta) \times D(\lambda^*_{(p-r)}, k_0, \delta) \times D(v^*, k_0, \delta).$$

Theorem 1 *Let $f, c_i, g_j \in C^2$ and the standard second-order optimality conditions (1), (5), and (6) are satisfied, then there exists $k_0 > 0$ large enough and $\delta > 0$ small enough that for any $(y, k) \in D(y^*, k_0, \delta)$ the following statements hold.*

(1) *There exists a vector*

$$\hat{x} = \hat{x}(y, k) = \operatorname{argmin}\{\mathcal{L}_k(x, y) | x \in \mathbb{R}^n\}$$

such that

$$\nabla_x \mathcal{L}_k(\hat{x}, y) = 0. \tag{10}$$

(2) *Let $\hat{y} = (\hat{\lambda}, \hat{v})$ with*

$$\hat{\lambda} = \Psi'(kc(\hat{x}))\lambda \quad \text{and} \quad \hat{v} = v - kg(\hat{x}), \tag{11}$$

where $\Psi'(kc(\hat{x})) = \operatorname{diag}\left(\psi(kc_i(\hat{x}))\right)_{i=1}^p$. Then for the pair $(\hat{x}, \hat{y})$ the following bound holds

$$\max\left\{\|\hat{x} - x^*\|, \|\hat{y} - y^*\|\right\} \leq ck^{-1}\|y - y^*\|,$$

where $c > 0$ is independent of $k \geq 0$.

(3) *The augmented Lagrangian for the equivalent problem $\mathcal{L}_k(x, y)$ is strongly convex in the neighborhood of $\hat{x}$.*

Theorem 1, which can be proven by modifying the technique used in [2] and [8], is the foundation for the following NR method. The method alternates the unconstrained minimization of the augmented Lagrangian $\mathcal{L}_k(x, y)$ in the primal space with Lagrange multipliers update.

Let the primal–dual approximation $(x^s, y^s) = (x^s, \lambda^s, v^s)$ be found already. We find the next approximation $(x^{s+1}, y^{s+1}) = (x^{s+1}, \lambda^{s+1}, v^{s+1})$ by the following formulas

$$x^{s+1} = \arg\min_{x \in \mathbb{R}^n} \mathcal{L}_k(x, \lambda^s, v^s) \tag{12}$$

or, equivalently, we find x^{s+1} as a solution of the following system of equations

$$\nabla f(x) - \sum_{i=1}^{p} \lambda_i^s \psi' (kc_i(x)) \nabla c_i(x) - \sum_{j=1}^{q} \left(v_j^s - kg_j(x)\right) \nabla g_j(x) = 0. \tag{13}$$

We find the new Lagrange multipliers by the formulas

$$\lambda_i^{s+1} = \lambda_i^s \psi'(kc_i(x^{s+1})), \quad i = 1, \ldots, p, \tag{14}$$

$$v_j^{s+1} = v_j^s - kg_j(x^{s+1}), \quad j = 1, \ldots, q. \tag{15}$$

The unconstrained minimization (12) is an infinite procedure. In the next section we replace the minimization (12) and the Lagrange multipliers update (14) and (15) by solving a primal–dual system of equations. The application of Newton's method for solving the primal–dual system leads to the EPM. The EPM reduces the computational complexity at each step as compared with the Newton NR method (see [2,6,8]) and improves the rate of convergence in the neighborhood of the primal–dual solution.

5 Exterior-point method

In this section, we introduce and analyze the EPM.

The important component of the EPM is the merit function, which measures the distance between the current approximation (x, λ, v) and the solution:

$$\begin{aligned} \nu(x, y) = \nu(x, \lambda, v) = \max \Bigg\{ & \|\nabla_x L(x, \lambda, v)\|, -\min\nolimits_{1 \leq i \leq p} c_i(x), \\ & \max\nolimits_{1 \leq i \leq q} |g_i(x)|, \sum_{i=1}^{p} |\lambda_i||c_i(x)|, -\min\nolimits_{1 \leq i \leq p} \lambda_i, \Bigg\} \end{aligned} \tag{16}$$

For a given scaling parameter $k > 0$ and a starting point $z = (x, \lambda, v)$ one step of the NRAL method is equivalent to solving the following primal–dual system for $\hat{x}$, $\hat{\lambda}$, and $\hat{v}$

$$\begin{aligned} \nabla_x \mathcal{L}_k(\hat{x}, \lambda, v) = \nabla f(\hat{x}) & - \sum_{i=1}^{p} \psi' \left(kc_i(\hat{x})\right) \lambda_i \nabla c_i(\hat{x}) \\ & - \sum_{j=1}^{q} \left(v_j - kg_j(\hat{x})\right) \nabla g_j(\hat{x}) = 0, \end{aligned} \tag{17}$$

$$\hat{\lambda} - \Psi'\left(kc(\hat{x})\right)\lambda = 0, \tag{18}$$

$$\hat{v} - v + kg(\hat{x}) = 0, \tag{19}$$

where $\Psi'(kc(\hat{x})) = \text{diag}(\psi'(kc_i(\hat{x})))_{i=1}^{p}$. After replacing in (17) the terms $\psi'\left(kc_i(\hat{x})\right)\lambda_i$ by $\hat{\lambda}_i$ and $\left(v_j - kg_j(\hat{x})\right)$ by $\hat{v}_j$ we obtain an equivalent primal–dual system

$$\nabla_x L(\hat{x}, \hat{\lambda}, \hat{v}) = \nabla f(\hat{x}) - \sum_{i=1}^{m} \hat{\lambda}_i \nabla c_i(\hat{x}) - \sum_{j=1}^{q} \hat{v}_j \nabla g_j(\hat{x}) = 0, \tag{20}$$

$$\hat{\lambda} - \Psi'\left(kc(\hat{x})\right)\lambda = 0, \tag{21}$$

$$\hat{v} - v + kg(\hat{x}) = 0. \tag{22}$$

Let us consider one Newton step for solving the systems (20)–(22) for $\hat{x}$, $\hat{\lambda}$, and $\hat{v}$ from the starting point $(x, y) = (x, \lambda, v)$. By linearizing the systems (20)–(22) and ignoring the terms of the second and higher order we obtain the following system for finding the primal–dual Newton direction $(\Delta x, \Delta y) = (\Delta x, \Delta\lambda, \Delta v)$

$$\begin{bmatrix} \nabla^2_{xx} L(\cdot) & -\nabla c^T(\cdot) & -\nabla g^T(\cdot) \\ -k\Lambda\Psi''(\cdot)\nabla c(\cdot) & I_p & 0 \\ k\nabla g(\cdot) & 0 & I_q \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta v \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ \bar{\lambda} - \lambda \\ -kg(\cdot) \end{bmatrix}, \tag{23}$$

where $\nabla c(\cdot) = \nabla c(x)$, $\nabla g(\cdot) = \nabla g(x)$, $\Psi''(\cdot) = \Psi''(kc(x)) = \text{diag}\left(\psi''(kc_i(x))\right)_{i=1}^{p}$, $\bar{\lambda} = \Psi'(kc(x))\lambda$, $\Lambda = \text{diag}(\lambda_i)_{i=1}^{p}$ and I_p, I_q are the identity matrices in $\mathbb{R}^{p,p}$ and $\mathbb{R}^{q,q}$, respectively. By introducing

$$N_k(\cdot) = \begin{bmatrix} \nabla^2_{xx} L(\cdot) & -\nabla c^T(\cdot) & -\nabla g^T(\cdot) \\ -k\Lambda\Psi''(\cdot)\nabla c(\cdot) & I_p & 0 \\ k\nabla g(\cdot) & 0 & I_q, \end{bmatrix}$$

we can rewrite system (23) as follows

$$N_k(\cdot) \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta v \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ \bar{\lambda} - \lambda \\ -kg(\cdot) \end{bmatrix}.$$

Another important component of the EPM is the relation between the scaling parameter and the merit function value. We define the relation by the following formula

$$k = \nu(x, \lambda, v)^{-0.5}. \tag{24}$$

Now we can describe the EPM step. For a given $x \in \mathbb{R}^n$, Lagrange multipliers vectors $\lambda \in \mathbb{R}^p_{++}$, $v \in \mathbb{R}^q$, and a scaling parameter $k > 0$ one step of the EPM consists of the following operations:

1. Calculate the scaling parameter

$$k = \nu(x, \lambda, v)^{-0.5}. \tag{25}$$

2. Find the primal–dual Newton direction from the system

$$N_k(\cdot)\begin{bmatrix}\Delta x\\ \Delta\lambda\\ \Delta v\end{bmatrix}=\begin{bmatrix}-\nabla_x L(\cdot)\\ \bar{\lambda}-\lambda\\ -kg(\cdot)\end{bmatrix}. \quad (26)$$

3. Find the new primal–dual vector

$$\hat{x} := x+\Delta x, \quad \hat{\lambda} := \lambda+\Delta\lambda, \quad \hat{v} := v+\Delta v \quad (27)$$

Very often matrix $N_k(\cdot)$ is sparse, so numerical linear algebra techniques developed for the interior-point method (see [12]) can be used for solving (26). The following lemma guarantees that the methods (25)–(27) is well defined.

Lemma 3 *If the standard second-order optimality conditions* (1)*,* (5)*,* (6)*, and the Lipschitz conditions* (7) *are satisfied then there exists* $\varepsilon_0 > 0$ *small enough that for any* $(x,\lambda,v) \in \Omega_{\varepsilon_0}(x^*, l^*, v^*) = \Omega_{\epsilon 0}$ *the matrix*

$$M_k(x,\lambda,v) = \nabla^2_{xx}L(x,\lambda,v) - k\nabla c^T(x)\Psi''(kc(x))\,\Lambda\nabla c(x) + k\nabla g^T(x)\nabla g(x)$$

is positive definite and therefore the matrix $N_k(x,\lambda,v) = N_k(\cdot)$ *is nonsingular.*

Proof Note that $M_k(x^*,y^*) = \nabla^2_{xx}\mathcal{L}_k(x^*,y^*)$, therefore from Lemma follows the existence of $\mu > 0$ such that

$$u_1^T M_k(x^*,\lambda^*,v^*)u_1 \geq \mu u_1^T u_1, \quad \forall k \geq k_0, \forall u_1 \in \mathbb{R}^n$$

It follows from the Lipschitz conditions (7) that there exists $\varepsilon_0 > 0$ such that for any triple $(x,\lambda,v) \in \Omega_{\varepsilon_0}$ the matrix $M_k(x,\lambda,v)$ is positive definite.

To prove that $N_k(x,\lambda,v) = N(\cdot)$ is nonsingular for all $\forall\,(x,\lambda,v) \in \Omega_\varepsilon$ and $k \geq k_0$ we show that the equation $N_k(\cdot)u = 0$ implies $u = 0$, where $u = (u_1,u_2,u_3)$. We can rewrite the system

$$\begin{bmatrix}\nabla^2_{xx}L(\cdot) & -\nabla c^T(\cdot) & -\nabla g^T(\cdot)\\ -k\Lambda\Psi''(\cdot)\,\nabla c(\cdot) & I_p & 0\\ k\nabla g(\cdot) & 0 & I_q\end{bmatrix}\begin{bmatrix}u_1\\ u_2\\ u_3\end{bmatrix}=\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}$$

as follows

$$\left(\nabla^2_{xx}L(x,\lambda,v)\right)u_1 - \nabla c(x)^T u_2 - \nabla g(x)^T u_3 = 0, \quad (28)$$

$$-\,k\Lambda\Psi''(kc(x))\,\nabla c(x)u_1 + u_2 = 0, \quad (29)$$

$$-\,k\nabla g(x)u_1 + u_3 = 0. \quad (30)$$

By substituting the value of u_2 and u_3 from (29) and (30) into (28) we obtain the following system

$$M_k(x,\lambda,v)u_1 = \big(\nabla^2_{xx}L(x,\lambda,v) - k\nabla c^T(x)\Psi''(kc(x))\,\Lambda\nabla c(x) + k\nabla g(x)^T\nabla g(x)\big)\,u_1 = 0. \quad (31)$$

Since the matrix $M_k(x,\lambda,v)$ is positive definite then from (31) follows $u_1 = 0$ and, consequently, due to (29) and (30) we obtain $u_2 = u_3 = 0$.

The lemma is proven. □

We recall that $I^* = \{1, \ldots, r\}$ and $I^0 = \{r+1, \ldots, p\}$ are the sets of the active and the passive inequality constraints, respectively. Let $c_{(r)}(x)$, $\nabla c_{(r)}(x)$, $\lambda_{(r)}$, $c_{(p-r)}(x)$, $\nabla c_{(p-r)}(x)$, and $\lambda_{(p-r)}$ be the vector-functions, their Jacobians and the vector of the Lagrange multipliers corresponding to the active and passive sets, respectively. Let $L_{(r+q)}(x, \lambda_{(r)}, v) = f(x) - \lambda_{(r)}^T c_{(r)}(x) - v^T g(x)$ be the Lagrangian corresponding to both the active set and the equations.

We need the following auxiliary lemmas.

Lemma 4 *Let the matrix* $A \in \mathbb{R}^{n,n}$ *be nonsingular,* $\|A^{-1}\| \leq M$ *and* $\varepsilon > 0$ *small enough. Then any matrix* $B \in \mathbb{R}^{n,n}$ *such that* $\|A - B\| \leq \varepsilon$ *is nonsingular and* $\|B^{-1}\| \leq 2M$.

The proof of Lemma can be found for example in [4].

It follows from the standard second-order optimality conditions (see [7]) that the matrix

$$A = A(x^*, \lambda^*, v^*) = \begin{bmatrix} \nabla^2_{xx} L_{(r+q)}(x^*, \lambda^*_{(r)}, v^*) & -\nabla c^T_{(r)}(x^*) & -\nabla g^T(x^*) \\ \nabla c_{(r)}(x^*) & 0 & 0 \\ \nabla g(x^*) & 0 & 0 \end{bmatrix}$$

has an inverse and there is $M > 0$ such that

$$\|A^{-1}\| \leq M. \tag{32}$$

We will use (32) and Lemma 4 to prove the following lemma.

Lemma 5 *If the standard second-order optimality conditions* (1), (5), (6), *and the Lipschitz conditions* (7), *are satisfied then there exists* $\varepsilon_0 > 0$ *small enough such that for any primal–dual vector* $z = (x, y) = (x, \lambda, v) \in \Omega_{\varepsilon_0}$ *the following hold true*

(1) *There exist* $0 < L_1 < L_2$ *such that the merit function* $\nu(z)$ *yields*

$$L_1\|z - z^*\| \leq \nu(z) \leq L_2\|z - z^*\|. \tag{33}$$

(2) *For any* $z \in \Omega_{\varepsilon_0}$ *the matrix*

$$A(x, \lambda_{(r)}, v) = \begin{bmatrix} \nabla^2_{xx} L_{(r+q)}(x, \lambda_{(r)}, v) & -\nabla c^T_{(r)}(x) & -\nabla g^T(x) \\ \nabla c_{(r)}(x) & 0 & 0 \\ \nabla g(x) & 0 & 0 \end{bmatrix}$$

is nonsingular and there exists $M > 0$ *such that the following bound holds*

$$\|A^{-1}(x, \lambda_{(r)}, v)\| \leq 2M. \tag{34}$$

(3) *Let* $D_r = diag(d_i)_{i=1}^r$ *be diagonal matrices with bounded from above elements, i.e.,* $\max\{d_i\}_{i=1}^r = \bar{d} < \infty$. *Then there exists* $k_0 > 0$ *such that for any* $k \geq k_0$ *and any* $z \in \Omega_{\varepsilon_0}$ *the matrix*

$$B_k(x, \lambda, v) = \begin{bmatrix} \nabla^2_{xx} L(x, \lambda, v)) & -\nabla c^T_{(r)}(x) & -\nabla g^T(x) \\ \nabla c_{(r)}(x) & \frac{1}{k} D_r & 0 \\ \nabla g(x) & 0 & \frac{1}{k} I_q \end{bmatrix}$$

is nonsingular and there exists $M > 0$ *such that the following bound holds*

$$\|B_k^{-1}(x, \lambda, v)\| \leq 2M. \tag{35}$$

Proof Keeping in mind that $\nu(z^*) = 0$ the right inequality (33) follows from the Lipschitz conditions (7) and the boundedness of Ω_{ε_0}, i.e., there exists $L_2 > 0$ such that

$$\nu(z) \leq L_2\|z - z^*\|.$$

From a definition of the merit function (16) we have

$$\|\nabla_x L(x, \lambda, v)\| \leq \nu(z), \tag{36}$$

$$-\min_{1\leq i\leq p} c_i(x) \leq \nu(z), \tag{37}$$

$$\max_{1\leq j\leq q} |g_i(x)| \leq \nu(z), \tag{38}$$

$$|\lambda_i||c_i(x)| \leq \nu(z), \quad i = 1, \ldots, p. \tag{39}$$

Due to the standard second-order optimality conditions there exists $\tau_1 > 0$ such that $c_i(x) \geq \tau_1$, $i \in I^0$, if $z \in \Omega_{\varepsilon_0}$. Therefore, from (39) we get

$$|\lambda_i| \leq \frac{1}{\tau_1}\nu(z) = C_1\nu(z), \quad i \in I^0. \tag{40}$$

Due to the boundedness of Ω_{ε_0} there exists also $\tau_2 > 0$ such that $\|\nabla c_{(p-r)}(x)\| \leq \tau_2$ for all $z \in \Omega_{\varepsilon_0}$. Thus, taking into account (36) we have

$$\begin{aligned}\|\nabla_x L_{(r+q)}(x, \lambda_{(r)}, v)\| &\leq \|\nabla_x L(x, \lambda, v)\| + \|\nabla c^T_{(m-r)}(x)\lambda_{(p-r)}\| \\ &\leq \|\nabla_x L(x, \lambda, v)\| + \textstyle\sum_{i=p-r+1}^{p} \|\nabla c_i(x)\||\lambda_i| \leq C_2\nu(z),\end{aligned} \tag{41}$$

where $C_2 = 1 + (p - r)C_1\tau_2$. Also due to the standard second-order optimality conditions there exists $\tau_3 > 0$ such that $\lambda_i \geq \tau_3$ for $i \in I^*$ and $z \in \Omega_{\varepsilon_0}$. Combining (37)–(39) we obtain

$$\max\left\{\|c_{(r)}(x)\|,\ \|g(x)\|\right\} \leq C_3\nu(z), \tag{42}$$

where $C_3 = \min\{1, \frac{1}{\tau_3}\}$.

After linearizing $\nabla_x L_{(r+q)}(x, \lambda_{(r)}, v)$, $c_{(r)}(x)$, and $g(x)$ at the solution $(x^*, \lambda^*_{(r)}, v^*)$, we obtain

$$\begin{aligned}\nabla_x L_{(r+q)}(x, \lambda_{(r)}, v) = \ &\nabla_x L_{(r+q)}(x^*, \lambda^*_{(r)}, v^*) + \nabla^2_{xx} L_{(r+q)}(x^*, \lambda^*_{(r)}, v^*)(x - x^*) \\ &-\nabla c^T_{(r)}(x^*)(\lambda_{(r)} - \lambda^*_{(r)}) \\ &-\nabla g^T(x^*)(v - v^*) + \mathcal{O}_{(n)}\|x - x^*\|^2,\end{aligned} \tag{43}$$

$$c_{(r)}(x) = c_{(r)}(x^*) + \nabla c_{(r)}(x^*)(x - x^*) + \mathcal{O}_{(r)}\|x - x^*\|^2, \tag{44}$$

$$g(x) = g(x^*) + \nabla g(x^*)(x - x^*) + \mathcal{O}_{(q)}\|x - x^*\|^2. \tag{45}$$

Keeping in mind K-K-T conditions we can rewrite (43)–(45) in a matrix form

$$\begin{bmatrix} \nabla^2_{xx} L_{(r+q)}(x^*, \lambda^*_{(r)}, v^*) & -\nabla c^T_{(r)}(x^*) & -\nabla g^T(x^*) \\ \nabla c_{(r)}(x^*) & 0 & 0 \\ \nabla g(x^*) & 0 & 0 \end{bmatrix} \begin{bmatrix} x - x^* \\ \lambda_{(r)} - \lambda^*_{(r)} \\ v - v^* \end{bmatrix} \tag{46}$$

$$= \begin{bmatrix} \nabla_x L_{(r+q)}(x, \lambda_{(r)}, v) & +\mathcal{O}_{(n)}\|x - x^*\|^2 \\ c_{(r)}(x) & +\mathcal{O}_{(r)}\|x - x^*\|^2 \\ g(x) & +\mathcal{O}_{(q)}\|x - x^*\|^2 \end{bmatrix}.$$

Due to the standard second-order optimality conditions the matrix

$$A(x^*, \lambda^*_{(r)}, v^*) = \begin{bmatrix} \nabla^2_{xx} L_{(r+q)}(x^*, \lambda^*_{(r)}, v^*) & -\nabla c^T_{(r)}(x^*) & -\nabla g^T(x^*) \\ \nabla c_{(r)}(x^*) & 0 & 0 \\ \nabla g(x^*) & 0 & 0 \end{bmatrix}$$

is nonsingular (see [7], p.231) and there exists $M > 0$ such that

$$\|A^{-1}(x^*, \lambda^*_{(r)}, v^*)\| \leq M. \tag{47}$$

Hence from (46) we have

$$\left\| \begin{matrix} x - x^* \\ \lambda_{(r)} - \lambda^*_{(r)} \\ v - v^* \end{matrix} \right\| \leq M \max\{C_2, C_3\} \nu(z) + \mathcal{O}\|z - z^*\|^2.$$

Using (40) and assuming $1/L_1 = \max\{C_1, 2M \max\{C_2, C_3\}\}$ we obtain the left inequality (33), i.e.,

$$L_1\|z - z^*\| \leq \nu(z).$$

The bounds (34) and (35) follow from Lemma 4 and the Lipschitz conditions (7). Lemma 5 is proven. □

The NR methods (12), (14), and (15) requires finding an unconstrained minimizer at each step. The Newton NR method replaces the unconstrained minimization by finding an approximation of the primal minimizer using Newton's method [2,6,8]. Several Newton steps are required to find the primal approximation and the updated Lagrange multipliers. Due to Theorem 1 finding the primal approximation followed by the Lagrange multipliers update reduces the distance between the current primal–dual approximation and the solution by a factor $0 < \gamma < 1, \gamma = ck^{-1}$, i.e., the Newton NR method has a linear rate of convergence.

The EPM improves the Newton NR method in two directions. First, each step of the EPM requires only one Newton step for solving the primal–dual system (26). Second, instead of a linear rate, the EPM converges to the primal–dual solution with 1.5-Q-superlinear rate.

Now we are ready to establish our main result. For the methods (25)–(27) the following theorem holds.

Theorem 2 *If the standard second-order optimality conditions* (1), (5)*, and* (6) *and the Lipschitz conditions* (7) *are satisfied then there exists* $\varepsilon_0 > 0$ *small enough such that for any primal–dual triple* $z = (x, \lambda, v) \in \Omega_{\varepsilon_0}$ *only one step of EPM (25)–(27) is enough to obtain a new primal–dual approximation* $\hat{z} = (\hat{x}, \hat{\lambda}, \hat{v})$ *such that the following estimation holds*

$$\|\hat{z} - z^*\| \leq C\|z - z^*\|^{3/2}, \tag{48}$$

where $C > 0$ *is a constant depending only on the problem's data.*

Proof From Lemmas 3 and 5 follows the existence of $\varepsilon_0 > 0$ small enough such that the matrix $N_k(\cdot)$ is nonsingular for any $x \in \Omega_{\varepsilon_0}$. Therefore, the methods (25)–(27) is executable for any starting point $z \in \Omega_{\varepsilon_0}$. Let $z = (x, \lambda, v) \in \Omega_{\varepsilon_0}$ be such that $\|z - z^*\| = \varepsilon \leq \varepsilon_0$.

Due to the formulas (24) for the scaling parameter update and (33), we have

$$\frac{1}{\sqrt{L_2}}\varepsilon^{-\frac{1}{2}} \leq k \leq \frac{1}{\sqrt{L_1}}\varepsilon^{-\frac{1}{2}}. \tag{49}$$

We rewrite the system (23) specifying the active and the passive constraints sets

$$\begin{bmatrix} \nabla^2_{xx}L(\cdot) & -\nabla c^T_{(r)}(\cdot) & -\nabla c^T_{(p-r)}(\cdot) & -\nabla g^T(\cdot) \\ -k\Lambda_{(r)}\Psi''_{(r)}(\cdot)\,\nabla c_{(r)}(\cdot) & I_r & 0 & 0 \\ -k\Lambda_{(p-r)}\Psi''_{(p-r)}(\cdot)\,\nabla c_{(p-r)}(\cdot) & 0 & I_{p-r} & 0 \\ k\nabla g(\cdot) & 0 & 0 & I_q \end{bmatrix}$$

$$\times \begin{bmatrix} \Delta x \\ \Delta\lambda_{(r)} \\ \Delta\lambda_{(p-r)} \\ \Delta v \end{bmatrix} = \begin{bmatrix} -\nabla_x L(\cdot) \\ \bar{\lambda}_{(r)} - \lambda_{(r)} \\ \bar{\lambda}_{(p-r)} - \lambda_{(p-r)} \\ \bar{v} - v \end{bmatrix}. \tag{50}$$

First, we consider separately the system corresponding to the passive constraints. After rearranging the terms we obtain

$$\hat{\lambda}_{(p-r)} := \lambda_{(p-r)} + \Delta\lambda_{(p-r)} = \bar{\lambda}_{(p-r)} + k\Lambda_{(p-r)}\Psi''_{(p-r)}(\cdot)\,\nabla c_{(p-r)}(\cdot)\Delta x.$$

Therefore, for any $i \in I^0$ we have

$$\hat{\lambda}_i = \lambda_i + \Delta\lambda_i = \psi'(kc_i(x))\lambda_i + k\psi''(kc_i(x))\lambda_i\nabla c_i(x)^T\Delta x.$$

We recall that $\psi'(t) \leq a(t+1)^{-1}$, $-\psi''(t) \leq b(t+1)^{-2}$, $t \geq 0$, $a > 0$, $b > 0$. Also due to the standard second-order optimality conditions and the boundedness Ω_{ε_0} there exists $\eta_1 > 0$, $\eta_2 > 0$, $\eta_3 > 0$ such that $c_i(x) \geq \eta_1$, $\|\nabla c_i(x)\| \leq \eta_2$, $\|\Delta x\| \leq \eta_3$, $i \in I^0$ for any $(x, \lambda, v) \in \Omega_{\varepsilon_0}$. Using the formula (24) for the scaling parameters update, keeping in mind that $|\lambda_i| \leq \varepsilon$ for $i \in I^0$ and (49) we obtain

$$|\hat{\lambda}_i| \leq \frac{a}{k\eta_1}\lambda_i + \frac{b\eta_2\eta_3}{k\eta_1^2}\lambda_i \leq C_4\varepsilon^{\frac{3}{2}}, \quad i \in I^0, \tag{51}$$

where $C_4 = \frac{a\sqrt{L_2}}{\eta_1} + \frac{b\sqrt{L_2}\eta_2\eta_3}{\eta_1^2}$.

Now we concentrate on the analysis of the primal–dual system that corresponds to the active inequality constraints and equations. The first, the second, and the fourth rows of the system (50) are equivalent to

$$\begin{bmatrix} \nabla^2_{xx}L(\cdot) & -\nabla c^T_{(r)}(\cdot) & -\nabla g^T(\cdot) \\ -k\Lambda_{(r)}\Psi''_{(r)}(\cdot)\,\nabla c_{(r)}(\cdot) & I_r & 0 \\ k\nabla g(\cdot) & 0 & I_q \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda_{(r)} \\ \Delta v \end{bmatrix}$$

$$= \begin{bmatrix} -\nabla_x L(\cdot) + \nabla c^T_{(p-r)}(\cdot)\Delta\lambda_{(p-r)} \\ \bar{\lambda}_{(r)} - \lambda_{(r)} \\ \bar{v} - v \end{bmatrix}.$$

After multiplying the second row of the system by $\left[-k\Lambda_{(r)}\Psi''(\cdot)\right]^{-1}$ and dividing the third one by k we obtain

$$\begin{bmatrix} \nabla^2_{xx}L(\cdot) & -\nabla c^T_{(r)}(\cdot) & -\nabla g^T(\cdot) \\ \nabla c_{(r)}(\cdot) & \left[-k\Lambda_{(r)}\Psi''_{(r)}(\cdot)\right]^{-1} & 0 \\ \nabla g(\cdot) & 0 & \frac{1}{k}I_q \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda_{(r)} \\ \Delta v \end{bmatrix} \tag{52}$$

$$= \begin{bmatrix} -\nabla_x L(\cdot) + \nabla c^T_{(p-r)}(\cdot)\Delta\lambda_{(p-r)} \\ \left[-k\Lambda_{(r)}\Psi''_{(r)}(\cdot)\right]^{-1}(\bar{\lambda}_{(r)} - \lambda_{(r)}) \\ -g(\cdot) \end{bmatrix}.$$

Keeping in mind that $c_i(x^*) = 0$ for $i \in I^*$ and using the Lagrange formula we have

$$\begin{aligned} (\bar{\lambda}_i - \lambda_i)(-k\lambda_i\psi''(\cdot))^{-1} &= \left(\lambda_i\psi'(kc_i(\cdot)) - \lambda_i\psi'(kc_i(x^*))\right)(-k\lambda_i\psi''(\cdot))^{-1} \\ &= \lambda_i k\psi''(\xi_i)(c_i(\cdot) - c_i(x^*))(-k\lambda_i\psi''(\cdot))^{-1} = -\psi''(\xi_i)(\psi''(\cdot))^{-1}c_i(\cdot), \end{aligned}$$

where $\xi_i = k\theta_i c_i(\cdot) + k(1-\theta_i)c_i(x^*) = k\theta_i c_i(\cdot)$, $0 < \theta_i < 1$. Therefore, the system (52) is equivalent to

$$\begin{bmatrix} \nabla^2_{xx}L(\cdot) & -\nabla c^T_{(r)}(\cdot) & -\nabla g^T(\cdot) \\ \nabla c_{(r)}(\cdot) & \left[-k\Lambda_{(r)}\Psi''_{(r)}(\cdot)\right]^{-1} & 0 \\ \nabla g(\cdot) & 0 & \frac{1}{k}I_q \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda_{(r)} \\ \Delta v \end{bmatrix}$$

$$= \begin{bmatrix} -\nabla_x L(\cdot) + \nabla c^T_{(p-r)}(\cdot)\Delta\lambda_{(p-r)} \\ -\Psi''_{(r)}(\xi)\left[\Psi''_{(r)}(\cdot)\right]^{-1}c_{(r)}(\cdot) \\ -g(\cdot) \end{bmatrix},$$

where $\Psi''_{(r)}(\xi) = \text{diag}(\psi''(\xi_i))_{i=1}^r$, or

$$B(\cdot)\Delta z_b = b(\cdot),$$

where

$$B(\cdot) = \begin{bmatrix} \nabla^2_{xx}L(\cdot) & -\nabla c^T_{(r)}(\cdot) & -\nabla g^T(\cdot) \\ \nabla c_{(r)}(\cdot) & \left[-k\Lambda_{(r)}\Psi''_{(r)}(\cdot)\right]^{-1} & 0 \\ \nabla g(\cdot) & 0 & \frac{1}{k}I_q \end{bmatrix},$$

$$b(\cdot) = \begin{bmatrix} -\nabla_x L(\cdot) + \nabla c^T_{(p-r)}(\cdot)\Delta\lambda_{(p-r)} \\ -\Psi''_{(r)}(\xi)\left[\Psi''_{(r)}(\cdot)\right]^{-1}c_{(r)}(\cdot) \\ -g(\cdot) \end{bmatrix}$$

and $\Delta z_b = (\Delta x, \Delta\lambda_{(r)}, \Delta v)$.

We compare the Newton directions Δz_b with those generated by Newton's method applied to the Lagrange system of equations that corresponds to the active constraints and equations

$$\nabla L_{(r+q)}(x, \lambda_{(r)}, v) = \nabla f(x) - \nabla c^T_{(r)}(x)\lambda_{(r)} - \nabla g^T(x)v = 0, \tag{53}$$

$$c_{(r)}(x) = 0, \tag{54}$$

$$g(x) = 0. \tag{55}$$

By linearizing the equations (53) and (54) the linear system to find the primal–dual Newton direction is given by

$$\begin{bmatrix} \nabla^2_{xx} L_{(r+q)}(\cdot) & -\nabla c^T_{(r)}(\cdot) & -\nabla g^T(\cdot) \\ \nabla c_{(r)}(\cdot) & 0 & 0 \\ \nabla g(\cdot) & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x' \\ \Delta \lambda'_{(r)} \\ \Delta v' \end{bmatrix} = \begin{bmatrix} -\nabla_x L_{(r+q)}(\cdot) \\ -c_{(r)}(\cdot) \\ -g(\cdot) \end{bmatrix}$$

or

$$A(\cdot)\Delta z'_a = a(\cdot),$$

where

$$A(\cdot) = \begin{bmatrix} \nabla^2_{xx} L_{(r+q)}(\cdot) & -\nabla c^T_{(r)}(\cdot) & -\nabla g^T(\cdot) \\ \nabla c_{(r)}(\cdot) & 0 & 0 \\ \nabla g(\cdot) & 0 & 0 \end{bmatrix}, \quad a(\cdot) = \begin{bmatrix} -\nabla_x L_{(r+q)}(\cdot) \\ -c_{(r)}(\cdot) \\ -g(\cdot) \end{bmatrix}$$

and $\Delta z'_a = (\Delta x', \Delta\lambda'_{(r)}, \Delta v')$. The new primal–dual approximation is obtained by the formulas

$$\hat{x}' = x + \Delta x', \quad \hat{\lambda}'_{(r)} = \lambda_{(r)} + \Delta\lambda'_{(r)}, \quad \hat{v}' = v + \Delta v' \tag{56}$$

or

$$\hat{z}' = z + \Delta z'_a.$$

Let us estimate $\|\hat{z}_{(r+q)} - z^*_{(r+q)}\|$, where $\hat{z}_{(r+q)} = (\hat{x}, \hat{\lambda}_{(r)}, \hat{v})$ is generated by (25)–(27).

$$\begin{aligned} \hat{z}_{(r+q)} - z^*_{(r+q)} &= z_{(r+q)} + \Delta z_b - z^*_{(r+q)} = z_{(r+q)} + \Delta z'_a + \Delta z_b - \Delta z'_a - z^*_{(r+q)} \\ &= \hat{z}'_{(r+q)} - z^*_{(r+q)} - \Delta z'_a + \Delta z_b. \end{aligned}$$

Therefore,

$$\|\hat{z}_{(r+q)} - z^*_{(r+q)}\| \le \|z'_{(r+q)} - z^*_{(r+q)}\| + \|\Delta z'_a - \Delta z_b\|. \tag{57}$$

First let us estimate $\|\Delta z'_a - \Delta z_b\|$. Due to Lemma 5 there exist inverse matrices $A^{-1} = A^{-1}(\cdot)$ and $B^{-1} = B^{-1}(\cdot)$ and for $a = a(\cdot)$, $b = b(\cdot)$ we have

$$\begin{aligned} &\|\Delta z'_a - \Delta z_b\| = \|A^{-1}a - B^{-1}b\| = \|A^{-1}a - B^{-1}a + B^{-1}a - B^{-1}b\| \\ &\quad = \|\left(A^{-1} - B^{-1}\right)a + B^{-1}(a-b)\| \le \|A^{-1} - B^{-1}\|\|a\| + \|B^{-1}\|\|a-b\| \\ &\quad \le \|A^{-1}\|\|A - B\|\|B^{-1}\|\|a\| + \|B^{-1}\|\|a-b\|. \end{aligned} \tag{58}$$

We consider the following matrix

$$A - B = \begin{bmatrix} \sum_{i=r+1}^{p} \lambda_i \nabla^2 c_i(x) & 0 & 0 \\ 0 & -\frac{1}{k}\left[\Psi''(\cdot)\right]^{-1} & 0 \\ 0 & 0 & \frac{1}{k} I_q \end{bmatrix}.$$

Due to the formulas (24), (42) and (49) we obtain

$$|kc_i(\cdot)| \le \frac{C_3 L_2}{\sqrt{L_1}} \varepsilon^{\frac{1}{2}}, \quad i \in I^* \tag{59}$$

and hence there is $\eta_4 > 0$ such that

$$|\psi''(kc_i(\cdot))| \geq \frac{1}{\eta_4}. \tag{60}$$

Due to the boundedness of Ω_{ε_0} there exists $\tau_4 > 0$ such that for $z \in \Omega_{\varepsilon_0}$ we have

$$\|\nabla^2 c_i(x)\| < \tau_4, \quad i \in I_0. \tag{61}$$

Therefore, keeping in mind the formulas (24), (49), and (61) we have

$$\|A - B\| \leq \max\left\{(\tau_4(p-r))\varepsilon, \quad \sqrt{L_2}\eta_4\varepsilon^{\frac{1}{2}}, \varepsilon^{\frac{1}{2}}\right\} = \max\left\{\sqrt{L_2}\eta_4, 1\right\}\varepsilon^{\frac{1}{2}} \tag{62}$$

for $0 < \varepsilon \leq \varepsilon_0$ small enough.

Let's now estimate $\|a - b\|$:

$$\|a - b\| = \left\| \begin{array}{c} -\nabla_x L_{(r+q)}(\cdot) + \nabla_x L(\cdot) - \nabla c^T_{(p-r)}(\cdot)\Delta\lambda_{(p-r)} \\ -c_{(r)}(\cdot) + \left(\Psi''(\xi)\right)\left[\Psi''(kc_{(r)}(\cdot)\right]^{-1} c_{(r)}(\cdot) \\ 0 \end{array} \right\|. \tag{63}$$

For the first component we obtain using (51)

$$\begin{aligned} &\| - \nabla_x L_{(r+q)}(\cdot) + \nabla_x L(\cdot) - \nabla c^T_{(p-r)}(\cdot)\Delta\lambda_{(p-r)}\| \\ &\quad = \| - \nabla_x L_{(r+q)}(\cdot) + \nabla_x L_{(r+q)}(\cdot) - \nabla c^T_{(p-r)}(\cdot)\lambda_{(p-r)} - \nabla c^T_{(p-r)}(\cdot)\Delta\lambda_{(p-r)}\| \\ &\quad = \|\nabla c^T_{(p-r)}(\cdot)(\lambda_{(p-r)} + \Delta\lambda_{(p-r)})\| = \|\nabla c^T_{(p-r)}(\cdot)\hat{\lambda}_{(p-r)}\| \leq \eta_2 C_4 \varepsilon^{\frac{3}{2}}. \end{aligned}$$

Next we estimate the second component of (63). Using the Lagrange formula for $i \in I^*$ we have

$$\begin{aligned} &\left|\left(\frac{\psi''(\xi_i)}{\psi''(kc_i(\cdot))} - 1\right)c_i(\cdot)\right| \leq \left|\frac{\psi''(\xi_i) - \psi''(kc_i(\cdot))}{\psi''(kc_i(\cdot))}\right| |c_i(\cdot)| \\ &\quad \leq \frac{|\psi'''(\bar{\xi}_i)||\xi_i - kc_i(\cdot)|}{|\psi''(kc_i(\cdot))|}|c_i(\cdot)| \leq \frac{|\psi'''(\bar{\xi}_i)||kc_i(\cdot)(\theta_i - 1)|}{|\psi''(kc_i(\cdot))|}|c_i(\cdot)|, \end{aligned}$$

where $\bar{\xi}_i = \bar{\theta}_i\xi_i + k(1-\bar{\theta}_i)c_i(\cdot) = kc_i(\cdot)(\bar{\theta}_i\theta_i + 1 - \bar{\theta}_i)$. Due to (59) there exist $\eta_5 > 0$ such that for $i \in I^*$

$$|\psi'''(\bar{\xi}_i)| \leq \eta_5.$$

Thus, taking into consideration the formulas (24), (41), (49), (59), and (60) we obtain for $i \in I_+$

$$\frac{|\psi'''(\bar{\xi}_i)||kc_i(\cdot)(1-\theta_i)|}{|\psi''(kc_i(\cdot))|}|c_i(\cdot)| \leq \eta_4\eta_5(1-\theta)C_3^2 L_2^2 L_1^{-\frac{1}{2}}\varepsilon^{\frac{3}{2}} = C_5\varepsilon^{\frac{3}{2}},$$

where $\theta = \min_{1 \leq i \leq r} \theta_i$.

Finally combining the formulas (24), (34), (35) (41), (42), (49), (58), and (62) we have

$$\begin{aligned} \|\Delta y'_a - \Delta y_b\| &\leq \|A^{-1}\|\|A - B\|\|B^{-1}\|\|a\| + \|B^{-1}\|\|a - b\| \\ &\leq 4M^2 \max\left\{\sqrt{L_2}\eta_4, 1\right\}\max\{C_2, C_3\}L_2\varepsilon^{\frac{3}{2}} \\ &\quad + 2M\max\{\eta_2 C_4, C_5\}\varepsilon^{\frac{3}{2}} = C_6\varepsilon^{\frac{3}{2}}. \end{aligned} \tag{64}$$

Due to the quadratic convergence of Newton's method for solving Lagrange system of equations that corresponds to the active constraints and equations (see [7], Theorem 9, p.247), we obtain

$$\|\hat{z}'_{(r+q)} - z^*_{(r+q)}\| \le C_0\varepsilon^2, \tag{65}$$

where $\hat{z}'_{(r+q)} = (\hat{x}', \hat{\lambda}'_{(r)}, \hat{v}')$ defined by (56) and $z^*_{(r+q)} = (x^*, \lambda^*_{(r)}, v^*)$.

Therefore, combining (57), (64), and (65) we obtain

$$\|\hat{z}_{(r+q)} - z^*_{(r+q)}\| \le \|\hat{z}'_{(r+q)} - z^*_{(r+q)}\| + \|\Delta y'_a - \Delta y_b\|$$

$$\le C_0\varepsilon^2 + C_6\varepsilon^{\frac{3}{2}} \le C_7\varepsilon^{\frac{3}{2}}. \tag{66}$$

Finally combining (51) and (66) for $\hat{z} = (\hat{x}, \hat{\lambda}, \hat{v})$ we have

$$\|\hat{z} - z^*\| \le \max\{C_4, C_7\}\varepsilon^{\frac{3}{2}} = C\varepsilon^{\frac{3}{2}} = C\|z - z^*\|^{3/2}.$$

Proof of Theorem 2 is complete. □

Remark 1 To make the matrix $N(\cdot)$ nonsingular for any (x, λ, v) we can regularize the Hessian of the Lagrangian $L(x, \lambda, v)$.

$$N_\alpha(\cdot) = \begin{bmatrix} \nabla^2_{xx}L(\cdot) + \alpha I_n & -\nabla c^T(\cdot) & -\nabla g^T(\cdot) \\ -k\Lambda\Psi''(\cdot)\nabla c(\cdot) & I_p & 0 \\ k\nabla g(\cdot) & 0 & I_q \end{bmatrix}, \tag{67}$$

where I_n is an identity matrix in $\mathbb{R}^{n,n}$. It is possible to show that a certain choice of regularization parameter α does not compromise the rate of convergence and at the same time guarantee that the method is well defined for any (x, λ, v).

6 Concluding remarks

The local convergence analysis of the EPM emphasizes the fundamental difference between the primal–dual NR approach (25)–(27) and Newton NR method (see [2, 6,8]), which is based on sequential unconstrained minimization of $\mathcal{L}_k(x, \lambda, v)$ in x by Newton's method followed by the Lagrange multipliers update. The latter method converges with a fixed scaling parameter, keeps stable the Newton area for the unconstrained minimization and allows the observation of the "hot start" phenomenon [2,6,8], when from some point on one Newton step for primal minimization is enough for the Lagrange multipliers update. To improve the rate of convergence one has to increase the scaling parameter from step to step. However, the unbounded increase of the scaling parameter leads to substantial numerical difficulties, since the Newton area for unconstrained minimization degenerates to a point. Moreover, in the framework of the NR method, any rapid increase of the scaling parameter after the Lagrange multipliers update leads to a substantial increase of the computational work per update because several Newton steps are required to get back to the NR trajectory.

The situation is fundamentally different with the EPM (25)–(27) in the neighborhood of the primal–dual solution. The rapid increase of the scaling parameter does not increase the computational work per step. Just the opposite: by using (25) for the scaling parameter update we obtain the Newton direction for the primal–dual system (26) close to the Newton direction for the Lagrange system of equations that corresponds to the active inequality and equality constraints. This enables us to prove 1.5-Q-superlinear rate of convergence of the EPM. At the same time, the EPM requires solving only one linear system (26) per step. Therefore, the EPM is more efficient in the neighborhood of the solution than Newton NR method.

We would like to emphasize the importance of the standard second-order optimality conditions for performance of the EPM. They are critical for the efficiency of the EPM and enable us to prove a 1.5-Q-superlinear rate of convergence. Preliminary numerical results obtained so far are encouraging [3,4,10]. The EPM for NLP with inequality constraints was numerically implemented and a number of NLP problems from COPS and CUTE sets have been solved with high accuracy (see [4,10]). For all solved problems the "hot start" phenomenon predicted in [8], has been systematically observed. For most problems just a few Lagrange multipliers updates are required before each Newton step of EPM improves the accuracy by at least one digit. Recently, the EPM was implemented using linear algebra developed in [12]. The numerical results show that the EPM can find solutions with very high accuracy in certain cases when an interior-point method experiences difficulties [3].

The next important step is to analyze the global convergence of the EPM for nonconvex problems. This requires a modification of the methods (25)–(27). We also plan to conduct extensive numerical experiments and work on implementation issues.

Acknowledgements The research of the authors was partially supported by the NSF-Grant CCF-0324999.

References

1. Debreu, G.: Definite and semidefinite quadratic forms. Econometrica **20**, 295–300 (1952)
2. Goldfarb, D., Polyak, R., Scheinberg, K., Yuzefovich, I.: A modified barrier-augmented Lagrangian method for constrained minimization. Comput. Optim. Appl. **14** (1), 55–74 (1999)
3. Griva, I.: Numerical experiments with an interior-exterior point method for nonlinear programming. Comput. Optim. Appl. **29** (2), 173–195 (2004)
4. Griva, I., Polyak, R.: Primal–dual nonlinear rescaling method with dynamic scaling parameter update. Math. Prog. **106** (2), 237–259 (2006)
5. Hestenes, M.R.: Multipliers and gradient methods. J. Optim. Theory Appl. **4** (5), 303–320 (1969)
6. Melman, A., Polyak, R.: The Newton modified barrier method for QP Problems. Ann. Oper. Res. **62**, 465–519 (1996)
7. Polyak, B.T.: Introduction to Optimization. Software Inc., New York (1987)
8. Polyak, R.: Modified barrier functions theory and methods. mathe program **54**, 177–222 (1992)
9. Polyak, R., Teboulle, M.: Nonlinear rescaling and proximal-like methods in convex optimization. Math. Program **76**, 265–284 (1997)
10. Polyak, R., Griva, I.: Primal–dual nonlinear rescaling method for convex optimization, J. Optim. Theory Appl. **122** (1), 111–156 (2004)
11. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher Optimization, (ed) pp. 283–298. Academic Press, London (1969)
12. Vanderbei, R.: Symmetric quasidefinite matrices. SIAM J. Opti. **5**, 100-113 (1995)