ON THE FAST LOCAL CONVERGENCE OF INTERIOR-POINT ℓ_2-PENALTY METHODS FOR NONLINEAR PROGRAMMING

LIFENG CHEN* AND DONALD GOLDFARB†

July 27, 2006

Abstract. In [3], two interior-point ℓ_2-penalty methods with strong global convergence properties were proposed for solving nonlinear programming problems. In this paper we show that under standard assumptions, slight modifications of these methods lead to fast local convergence. Specifically, we show that for each fixed small barrier parameter μ, iterates in a small neighborhood (roughly within $o(\mu)$) of the minimizer of the barrier subproblem converge Q-quadratically to the minimizer. The overall convergence rate of the iterates to the solution of the nonlinear program is Q-superlinear. Our modifications include refinements of the rule for updating the penalty parameter and the termination criteria used by the inner algorithms, and the computation at each iteration for two correction steps that incur only modest cost. We illustrate these convergence results by some examples.

Key words. nonlinear programming, interior-point method, ℓ_2-penalty method, global convergence, local convergence, superlinear convergence

1. Introduction.

We consider the nonlinear programming problem:

$$\text{(P)}\qquad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c_i(x) \geq 0,\ i \in \mathcal{I}, \\ & g_i(x) = 0,\ i \in \mathcal{E}, \end{array}$$

where $x \in \Re^n$, $\mathcal{I} = \{1, ..., m\}$, $\mathcal{E} = \{1, ..., p\}$, and the functions f, c and g are real valued and twice continuously differentiable on $\Re^n$.

In [3], two interior-point methods, a quasi-feasible method and an infeasible method were proposed for solving problem (P). It was shown there that both methods have strong global convergence properties; namely, under fairly weak assumptions they converge to either a critical point of problem (P) or identify an infeasible stationary point of an appropriate infeasibility measure. No analysis regarding local convergence was given in [3]. The aim of this paper is to develop slightly modified versions of these methods that (i) preserve their global convergence properties, (ii) exhibit fast local convergence, and (iii) do not incur much additional cost.

We first briefly review our quasi-feasible method. By adding the barrier term $-\mu \sum_{i\in\mathcal{I}} \ln c_i(x)$ to the objective function $f(x)$, we obtain the barrier subproblem

$$(\text{P}_\mu)\qquad \begin{array}{ll} \text{minimize} & \varphi_\mu(x) = f(x) - \mu \sum_{i\in\mathcal{I}} \ln c_i(x) \\ \text{subject to} & c_i(x) > 0,\ i \in \mathcal{I}, \\ & g_i(x) = 0,\ i \in \mathcal{E}, \end{array}$$

where $\mu > 0$ is the barrier parameter. The first-order optimality conditions for problem (P_μ) give rise to the following system of nonlinear equations in $(x, u, v) \in \Re^{n+m+p}$

$$\mathcal{R}_\mu(x, u, v) = \begin{bmatrix} \nabla_x \mathcal{L}(x, u, v) \\ C(x)u - \mu e \\ g(x) \end{bmatrix} = 0, \tag{1.1}$$

where $C(x) = \text{diag}(c(x))$, $\text{diag}(\cdot)$ denotes the diagonal matrix of a vector and $\mathcal{L}(x, u, v)$ is the Lagrangian function associated with problem (P)

$$\mathcal{L}(x, u, v) = f(x) - c(x)^\top u + g(x)^\top v. \tag{1.2}$$

*IEOR Department, Columbia University, New York, NY10027 (lifeng.chen@columbia.edu). Research supported by the Presidential Fellowship of Columbia University.

†IEOR Department, Columbia University, New York, NY10027 (gold@ieor.columbia.edu). Research supported in part by NSF Grant DMS 01-04282, DOE Grant DE-FG02-92EQ25126 and DNR Grant N00014-03-0514.

Our quasi-feasible method starts with a point strictly satisfying the inequality constraints and uses a perturbed Newton method with a line search strategy to find an approximate solution $x(\mu)$ of the barrier problem (P_μ). The merit function used in the line search is the ℓ_2-penalty function:

$$\Phi_{\mu,r}(x) = \varphi_\mu(x) + r\|g(x)\|, \tag{1.3}$$

where $r > 0$ is the penalty parameter and $\|\cdot\|$ denotes the Euclidean vector norm. The barrier parameter μ is then decreased and a new barrier problem is approximately solved by the perturbed Newton method.

Each step $\triangle x$ of the perturbed Newton method is obtained by solving a modified Newton system of (1.1):

$$\mathcal{M}_r(\mathcal{H}, x, u) \begin{bmatrix} \triangle x \\ \lambda \\ y \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ \mu e \\ -g(x) \end{bmatrix}, \tag{1.4}$$

where $\mathcal{H}$ is the Hessian of the Lagrangian $\nabla^2_{xx}\mathcal{L}(x, u, v)$,

$$\mathcal{M}_r(\mathcal{H}, x, u) = \begin{bmatrix} \mathcal{H} & -\nabla c(x) & \nabla g(x) \\ \mathcal{U}\nabla c(x)^\top & C(x) & 0 \\ \nabla g(x)^\top & 0 & -\frac{\|g(x)\|}{r} I \end{bmatrix}, \tag{1.5}$$

$\mathcal{U} = \mathrm{diag}(u)$ and I is the identity. To ensure that (1.4) solvable, the inertia of the symmetric matrix obtained from (1.5) by multiplying its second row of submatrices by $\mathcal{U}^{-1}$ is checked. If this matrix does not have the correct inertia, a multiple of the identity is added to $\mathcal{H}$ until it does. The primal iterate is updated by setting $x^+ = x + t\triangle x$, where the step size $t \in (0, 1]$ is determined by a backtracking line search strategy that ensures that $c(x^+)$ is strictly positive and a sufficient reduction is made in the merit function $\Phi_{\mu,r}(x)$. A fraction-to-the-boundary rule is employed to update the dual iterates corresponding to the inequality constraints so that they are positive. The penalty parameter r is updated according to a rule that guarantees proper convergence of the iterates.

It was shown in [3] that if the penalty parameter is bounded and an infinite sequence of iterates $\{(x^k, u^k, v^k)\}$ is generated ($x^k = x(\mu)$), then any accumulation point (x^*, u^*, v^*) satisfies the first-order optimality conditions for problem (P), i.e., x^* is a KKT point of (P) and (u^*, v^*) is an associated multiplier vector. In this paper we show that slight modifications of our quasi-feasible method result in local superlinear convergence of the iterates provided the following standard nondegeneracy conditions hold at (x^*, u^*, v^*).

Assumption A.

A1. The Hessian matrices $\nabla^2 f(x)$, $\nabla^2 c_i(x)$, $i \in \mathcal{I}$, $\nabla^2 g_i(x)$, $i \in \mathcal{E}$ are locally Lipschitz continuous at x^*.

A2. The linear independence constraint qualification (LICQ) holds: the active constraint gradients $\nabla g_i(x^*)$, $i \in \mathcal{E}$ and $\nabla c_i(x^*)$, $i \in \mathcal{B} = \{i \in \mathcal{I} | c_i(x^*) = 0\}$ are linearly independent.

A3. The second-order sufficient condition (SOSC) holds: there exists a $\rho > 0$ such that

$$d^\top \nabla^2_{xx}\mathcal{L}(x^*, u^*, v^*) d \geq \rho \|d\|^2$$

for all $d \in \Re^n$ such that $\nabla c_i(x^*)^\top d = 0$, $\forall i \in \mathcal{B}$ and $\nabla g(x^*)^\top d = 0$.

A4. The strict complementarity conditions hold: $c(x^*) + u^* > 0$.

We note that under the LICQ, (u^*, v^*) is the unique multiplier vector at x^*.

Our infeasible method is essentially our quasi-feasible method applied to problem (P) with slack variables, i.e.,

$$(\mathrm{SP}) \qquad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & w_i \geq 0,\ w_i - c_i(x) = 0,\ i \in \mathcal{I}, \\ & g_i(x) = 0,\ i \in \mathcal{E}. \end{array}$$

The barrier subproblem solved by the inner algorithm is

$$(\text{SP}_\mu) \qquad \begin{array}{ll} \text{minimize} & \varphi_\mu(x,w) = f(x) - \mu \sum_{i\in\mathcal{I}} \ln w_i \\ \text{subject to} & w_i > 0,\ w_i - c_i(x) = 0,\ i \in \mathcal{I}, \\ & g_i(x) = 0,\ i \in \mathcal{E}, \end{array}$$

and the corresponding merit function is

$$\Phi_{\mu,r}(x,w) = \varphi_\mu(x,w) + r \left\| \begin{bmatrix} w - c(x) \\ g(x) \end{bmatrix} \right\|. \tag{1.6}$$

The only difference between our quasi-feasible method and our infeasible method is that the latter takes advantage of the special structure of the slack variables when correcting the inertia of certain matrices. Although this may result in different paths of the iterates, the global convergence behavior of the two methods is almost identical. On the other hand, as a solution satisfying the SOSC and LICQ is approached, both of our methods use the exact Hessians as there is no need to modify them to ensure that they have the correct inertia. Therefore, our infeasible method eventually reduces to our quasi-feasible method applied to problem (SP) and its local convergence properties are exactly the same as those of our quasi-feasible method. Hence, we only consider the latter method in this paper.

The local convergence results presented below are twofold. First, we show how to modify our quasi-feasible method so that it converges locally Q-quadratically for each barrier subproblem with a sufficiently small μ. We also show that the radius of the region in which a unit Newton step satisfies the line search conditions and fast convergence occurs is only modestly less than μ. The modifications consist of a new rule for updating the penalty parameter and a second-order correction step computed from two perturbed Newton systems that have the same coefficient matrix as the original one used in [3]. Second, we show that the overall convergence rate of the iterates toward a solution of problem (P) that satisfies the SOSC is Q-superlinear as the barrier parameter is decreased at a superlinear rate. This is achieved by refining the termination criteria of the inner algorithm for solving the barrier subproblems.

Throughout the paper, we use the same notation as [3]. In addition, for an index set $\mathcal{A} \in \mathcal{I}$, $|\mathcal{A}|$ is the cardinality of $\mathcal{A}$. We use order notation $o(\cdot)$, $\mathcal{O}(\cdot)$ and $\Omega(\cdot)$ as follows. For vectors s and scalars $\varepsilon > 0$, we write $s = o(\varepsilon)$ if there is no constant $a > 0$ such that $\|s\| \geq a\varepsilon$ for all values of s and ε that are sufficiently small or sufficiently large. We write $s = \mathcal{O}(\varepsilon)$ if there is a constant $a > 0$ such that $\|s\| \leq a\varepsilon$ for all values of s and ε. We write $s = \Omega(\varepsilon)$ if there are constants $a_2 > a_1 > 0$ such that $a_1\varepsilon \leq \|s\| \leq a_2\varepsilon$ for all values of s and ε. Similarly, we write $s = o(1)$ if $\|s\| \to 0$, $s = \mathcal{O}(1)$ if $\|s\| \leq a$ and $s = \Omega(1)$ if $a_1 \leq \|s\| \leq a_2$.

The rest of the paper is organized as follows. In the next section we carry out a local analysis of the barrier subproblem (P_μ) for a fixed μ. Local analysis of problem (P) is given in Section 3. The detailed algorithm and its global and fast local convergence properties are presented in Section 4. We conclude in Section 5 with some numerical tests that illustrate our convergence results.

2. Local analysis for a fixed μ. In this section we focus on problem (P_μ) for a fixed μ. We assume that the penalty parameter r is a fixed constant and that $\mathcal{H}$ is the exact Lagrangian Hessian, i.e., $\mathcal{H} = \nabla^2_{xx}\mathcal{L}(x,u,v)$. It is well known that under Assumption A, if μ is sufficiently small, the nonlinear system of equations (1.1) has a unique solution $z(\mu) = (x(\mu), u(\mu), v(\mu))$ in a small neighborhood of $z^* = (x^*, u^*, v^*)$ that converges to z^* as μ goes to zero. Moreover, $z(\mu)$ is locally Lipschitz continuous, i.e., there exists a constant $\bar{C} > 0$ such that $\|z(\mu) - z^*\| \leq \bar{C}\mu$ for all μ small enough. For simplicity we use the notation

$$z = (x,u,v), \quad \mathcal{M}_r(z) = \mathcal{M}_r(\mathcal{H}, x, u).$$

The pure Newton system for (1.1) is given by

$$\mathcal{R}_0'(z) \begin{bmatrix} \triangle x^N \\ \lambda^N \\ y^N \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ \mu e \\ -g(x) \end{bmatrix}, \tag{2.1}$$

where

$$\mathcal{R}_0'(z) = \begin{bmatrix} \mathcal{H} & -\nabla c(x) & \nabla g(x) \\ \mathcal{U}\nabla c(x)^\top & C(x) & 0 \\ \nabla g(x)^\top & 0 & 0 \end{bmatrix}. \tag{2.2}$$

It is easy to see that as x approaches $x(\mu)$, the matrix $\mathcal{M}_r(z)$ approaches $\mathcal{R}_0'(z)$ with a residual that is $O\left(\frac{\|g(x)\|}{r}\right)$. Therefore, since the right hand sides of (1.4) and (2.1) are identical, if the rate of convergence of $\|g(x)\|$ to zero is slow, the step computed from (1.4) may not generate iterates that converge to $x(\mu)$ at a satisfactory rate. To circumvent this problem, we propose solving another perturbed Newton system:

$$\mathcal{M}_r(z) \begin{bmatrix} \widetilde{\triangle x} \\ \widetilde{\lambda} \\ \widetilde{y} \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ \mu e \\ -g(x) - \frac{\|g(x)\|}{r} v \end{bmatrix}. \tag{2.3}$$

It will be shown that when the iterates are close to being feasible, (2.3) provides a better approximation to the pure Newton system than (1.4). Solving (2.3) instead of (2.1) has the advantage of avoiding an extra matrix factorization as (2.3) and (1.4) have the same coefficient matrix. Hence, only a modest increase in cost is incurred. Moreover, our numerical experience indicates that if the Jacobian of the active constraints is singular or nearly singular, solving (2.3) often appears to be more stable than solving (2.1) due to the perturbation of the diagonal elements. This helps overcome numerical difficulties in some irregular problems.

It has been known that exact penalty function based methods may suffer from the Maratos effect in that an approximate full Newton step increases the merit function and hence is rejected by the line search. Among several remedies for overcoming this are nonmonotone line search strategies and second-order corrections. In this paper, we explore the use of a second-order correction step given by

$$\mathcal{M}_r(z) \begin{bmatrix} \widehat{\triangle x} \\ \widehat{\lambda} \\ \widehat{y} \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ \mu e \\ \nabla g(x)^\top \widetilde{\triangle x} - g(x + \widetilde{\triangle x}) - \frac{\|g(x)\|}{r} v \end{bmatrix}. \tag{2.4}$$

In the remainder of this section, we will use the notation

$$z^N = (x + \triangle x^N, \lambda^N, y^N), \quad \widetilde{z} = \left(x + \widetilde{\triangle x}, \widetilde{\lambda}, \widetilde{y}\right), \quad \widehat{z} = \left(x + \widehat{\triangle x}, \widehat{\lambda}, \widehat{y}\right).$$

Note that $\widetilde{z} = \widehat{z} = z^N$ if $\|g(x)\| = 0$.

Nondegeneracy assumptions A2, A3 and A4 imply that matrices $\mathcal{R}_0'(z)$ and $\mathcal{M}_r(z)$ are uniformly nonsingular if μ is sufficiently small and z is sufficiently close to $z(\mu)$. Specifically, there exist a constant $M > 0$ and a neighborhood $\mathcal{N}(z^*)$ of z^* such that

$$\|\mathcal{R}_0'(z)^{-1}\| \le M, \quad \|\mathcal{M}_r(z)^{-1}\| \le M, \quad \forall z \in \mathcal{N}(z^*). \tag{2.5}$$

Also, it is easy to verify that the solutions of (2.1), (2.3) and (2.4) tend to $(0, u(\mu), v(\mu))$ as z tends to $z(\mu)$. After some algebra, we have from (2.3) that

$$\mathcal{R}_0'(z) \begin{bmatrix} \widetilde{\triangle x} \\ \widetilde{\lambda} \\ \widetilde{y} \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ \mu e \\ -g(x) - \frac{\|g(x)\|}{r} (v - \widetilde{y}) \end{bmatrix}. \tag{2.6}$$

Subtracting (2.1) from (2.6) yields

$$\left\|z^N - \widetilde{z}\right\| = \mathcal{O}(\|g(x)\| \|\widetilde{y} - v\|) = o(\|x - x(\mu)\|). \tag{2.7}$$

Moreover, subtracting (2.3) from (2.4) and using Taylor's theorem yields

$$\|\widetilde{z} - \widehat{z}\| = \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|^2\right). \tag{2.8}$$

Hence, it is immediate that (2.3) is a better approximation to (2.1) than (1.4) and the perturbation introduced by the second-order correction step is relatively small.

The next theorem shows that $\widetilde{z}$ and $\widehat{z}$ converge quadratically to $z(\mu)$.

THEOREM 2.1. *Suppose Assumption A holds. If $z \in \mathcal{N}(z^*)$, then*
(i) $\|\widetilde{z} - z(\mu)\| = \mathcal{O}\left(\|z - z(\mu)\|^2\right)$ *and* $\|\widehat{z} - z(\mu)\| = \mathcal{O}\left(\|z - z(\mu)\|^2\right)$;
(ii)$\|\widetilde{z} - z\| = \Omega(\|z - z(\mu)\|)$ *and* $\|\widehat{z} - z\| = \Omega(\|z - z(\mu)\|)$.

Proof. We have from (2.3) that

$$\mathcal{M}_r(z)(\widetilde{z} - z(\mu)) = \begin{bmatrix} \mathcal{H}(x - x(\mu)) - \nabla_x \mathcal{L}(x, u(\mu), v(\mu)) \\ \mathcal{U}\nabla c(x)^\top (x - x(\mu)) - C(x)u(\mu) + \mu e \\ \frac{\|g(x)\|}{r}(v(\mu) - v) - g(x) + \nabla g(x)^\top (x - x(\mu)) \end{bmatrix}. \tag{2.9}$$

Moreover, we have from Assumption A1 and Taylor's Theorem and the fact that $\nabla_x \mathcal{L}(z(\mu)) = 0$, $c_i(x(\mu))u_i(\mu) = \mu$ for all $i \in \mathcal{I}$ and $g_i(x(\mu)) = 0$ for all $i \in \mathcal{E}$ that

$$\begin{aligned} & \mathcal{H}(x - x(\mu)) - \nabla_x \mathcal{L}(x, u(\mu), v(\mu)) \\ = \; & (\nabla^2_{xx}\mathcal{L}(z) - \nabla^2_{xx}\mathcal{L}(z(\mu))(x - x(\mu)) - \nabla_x \mathcal{L}(z(\mu)) + \mathcal{O}(\|x - x(\mu)\|^2) \\ = \; & \mathcal{O}(\|z - z(\mu)\|^2), \end{aligned} \tag{2.10}$$

$$\begin{aligned} & u_i \nabla c_i(x)^\top (x - x(\mu)) - c_i(x)u_i(\mu) + \mu \\ = \; & -(u_i(\mu) - u_i)\nabla c_i(x)^\top (x - x(\mu)) + \mathcal{O}(\|x - x(\mu)\|^2) \\ = \; & \mathcal{O}(\|z - z(\mu)\|^2), \quad \forall\, i \in \mathcal{I}, \end{aligned} \tag{2.11}$$

and

$$\frac{\|g(x)\|}{r}(v(\mu) - v) - g(x) + \nabla g(x)^\top (x - x(\mu)) = \mathcal{O}(\|z - z(\mu)\|^2). \tag{2.12}$$

Since $z \in \mathcal{N}(z^*)$, the quadratic convergence of $\widetilde{z}$ to $z(\mu)$ follows from (2.5), which together with (2.8) implies the quadratic convergence of $\widehat{z}$ to $z(\mu)$. Part (ii) follows immediately from (i). ☐

The strict complementarity assumption A4 implies that

$$\begin{cases} u_i(\mu) = \Omega(1), \quad c_i(x(\mu)) = \Omega(\mu), \quad \forall i \in \mathcal{B}; \\ u_i(\mu) = \Omega(\mu), \quad c_i(x(\mu)) = \Omega(1), \quad \forall i \in \mathcal{I}\backslash\mathcal{B}. \end{cases} \tag{2.13}$$

Hence, if $\|z - z(\mu)\| = o(\mu)$, we know by the continuous differentiability of $c(\cdot)$ that

$$\begin{cases} u > 0, c(x) > 0, \\ u_i = \Omega(1), \quad c_i(x) = \Omega(\mu), \quad \forall i \in \mathcal{B}; \\ u_i = \Omega(\mu), \quad c_i(x) = \Omega(1), \quad \forall i \in \mathcal{I}\backslash\mathcal{B}. \end{cases} \tag{2.14}$$

The next lemma shows that if μ is sufficiently small and z is very close to $z(\mu)$ then a full second-order correction step of (2.4) guarantees feasibility of $x + \widehat{\triangle x}$ with respect to the inequality constraints and dual feasibility of $\widehat{\lambda}$.

LEMMA 2.2. *Suppose Assumption A holds. If μ is sufficiently small and $\|z - z(\mu)\| = o(\mu)$, then $c(x + \widehat{\triangle x}) > 0$ and $\min\left\{\theta u_i, \frac{\mu}{c_i(x)}\right\} \le \widehat{\lambda}_i \le \frac{\mu\gamma}{c_i(x)}$, where $\theta \in (0, 1)$ and $\gamma > 1$.*

Proof. First, we have $z \in \mathcal{N}(z^*)$ for μ small enough. It then follows from $\|z-z(\mu)\| = o(\mu)$ and Theorem 2.1 (ii) that $\|\widehat{z} - z\| = o(\mu)$. From (2.4) and Theorem 2.1, we have that for each $i \in \mathcal{I}$

$$\begin{aligned} \widehat{\lambda}_i c_i(x + \widehat{\triangle x}) &= \widehat{\lambda}_i c_i(x) + \widehat{\lambda}_i \nabla c_i(x)^\top \widehat{\triangle x} + \mathcal{O}\left(\left\|\widehat{\triangle x}\right\|^2\right) \\ &= \mu + (\widehat{\lambda}_i - u_i)\nabla c_i(x)^\top \widehat{\triangle x} + \mathcal{O}\left(\left\|\widehat{\triangle x}\right\|^2\right) \\ &= \mu + \mathcal{O}\left(\|\widehat{z} - z\|^2\right). \end{aligned} \tag{2.15}$$

Hence, it follows that $c(x + \widehat{\triangle x}) > 0$ and $\widehat{\lambda} > 0$. By (2.15) and Assumption A4, we know $\widehat{\lambda}_i = \Omega(1)$ for $i \in \mathcal{B}$ and $\widehat{\lambda}_i = \Omega(\mu)$ for $i \in \mathcal{I}\backslash\mathcal{B}$. It then follows from Theorem 2.1 (ii) that

$$\widehat{\lambda} - \theta u = (1-\theta)\widehat{\lambda} + \theta(\widehat{\lambda} - u) = (1-\theta)\widehat{\lambda} + o(\mu) > 0.$$

Finally, we again have from (2.4) that for each $i \in \mathcal{I}$

$$\mu\gamma - \widehat{\lambda}_i c_i(x) = (\gamma - 1)\mu + u_i \nabla c_i(x)^\top \widehat{\triangle x} = (\gamma-1)\mu + o(\mu) > 0.$$

□

To prove that our second-order correction step can be eventually accepted by the line search, we first show in the next lemma that $\left\|\widehat{\triangle x}\right\| = \mathcal{O}(\|x - x(\mu)\|)$. We use an argument similar to one used by Wright in [15], which considers primal barrier methods, whereas our method is primal-dual.

LEMMA 2.3. *Suppose Assumption A holds. If μ is sufficiently small and $\|z - z(\mu)\| = o(\mu)$, then $\left\|\widetilde{\triangle x}\right\| = \mathcal{O}(\|x - x(\mu)\|)$ and $\left\|\widehat{\triangle x}\right\| = \mathcal{O}(\|x - x(\mu)\|)$.*

Proof. By eliminating $\widetilde{\lambda}$ and subtracting $(\nabla g(x)v(\mu), 0)^\top$ from both sides of (2.3) we have

$$\begin{bmatrix} \widehat{\mathcal{H}} & \nabla g(x) \\ \nabla g(x)^\top & 0 \end{bmatrix} \begin{bmatrix} \widetilde{\triangle x} \\ \widetilde{y} - v(\mu) \end{bmatrix} = \begin{bmatrix} -\nabla \varphi_\mu(x) - \nabla g(x) v(\mu) \\ -g(x) - \frac{\|g(x)\|}{r}(v - \widetilde{y}) \end{bmatrix}, \tag{2.16}$$

where

$$\widehat{\mathcal{H}} = \nabla^2_{xx}\mathcal{L}(z) + \nabla c(x) C(x)^{-1} \mathcal{U} \nabla c(x)^\top. \tag{2.17}$$

From Taylor series, (2.14) and $\|z - z(\mu)\| = o(\mu)$, we have for all $i \in \mathcal{I}$,

$$\frac{1}{c_i(x(\mu))} = \frac{1}{c_i(x)} + \frac{1}{c_i(x)^2}\mathcal{O}(\|x - x(\mu)\|). \tag{2.18}$$

Hence from the continuous differentiability of $\nabla c(\cdot)$, we obtain for all $i \in \mathcal{I}$,

$$\frac{\mu}{c_i(x(\mu))}\nabla c_i(x(\mu)) = \left(\frac{\mu}{c_i(x)} + \frac{\mu}{c_i(x)^2}\mathcal{O}(\|x - x(\mu)\|)\right)(\nabla c_i(x) + \mathcal{O}(\|x - x(\mu)\|)). \tag{2.19}$$

It then follows from (2.14) that

$$\frac{\mu}{c_i(x(\mu))}\nabla c_i(x(\mu)) = \begin{cases} \frac{\mu}{c_i(x)}\nabla c_i(x) + \mathcal{O}\left(\frac{\|x - x(\mu)\|}{\mu}\right)\nabla c_i(x) + \mathcal{O}(\|x - x(\mu)\|), & \forall i \in \mathcal{B}, \\ \frac{\mu}{c_i(x)}\nabla c_i(x) + \mathcal{O}(\mu\|x - x(\mu)\|), & \forall i \in \mathcal{I}\backslash\mathcal{B}. \end{cases}$$

Using the smoothness of $\nabla f(\cdot)$, $c(\cdot)$, $g(\cdot)$ and their derivatives, we can now estimate the magnitude of the right hand side of (2.16),

$$
\begin{aligned}
& \nabla\varphi_\mu(x) + \nabla g(x)v(\mu) \\
= \; & \nabla\varphi_\mu(x) + \nabla g(x)v(\mu) - \nabla\varphi_\mu(x(\mu)) - \nabla g(x(\mu))v(\mu) \\
= \; & \nabla f(x) - \nabla f(x(\mu)) - \textstyle\sum_{i\in\mathcal{B}} \left(\frac{\mu}{c_i(x)}\nabla c_i(x) - \frac{\mu}{c_i(x(\mu))}\nabla c_i(x(\mu)) \right) \\
& - \textstyle\sum_{i\in\mathcal{I}\setminus\mathcal{B}} \left(\frac{\mu}{c_i(x)}\nabla c_i(x) - \frac{\mu}{c_i(x(\mu))}\nabla c_i(x(\mu)) \right) + \left(\nabla g(x) - \nabla g(x(\mu))\right) v(\mu) \\
= \; & \textstyle\sum_{i\in\mathcal{B}} \mathcal{O}\left(\frac{\|x - x(\mu)\|}{\mu} \right) \nabla c_i(x) + \mathcal{O}(\|x - x(\mu)\|),
\end{aligned} \tag{2.20}
$$

Since $\|g(x)\| = \|g(x) - g(x(\mu))\| = \mathcal{O}(\|x - x(\mu)\|)$ and $\|v - \widetilde{y}\| = o(\mu)$, it follows that

$$
g(x) + \frac{\|g(x)\|}{r}(v - \widetilde{y}) = \mathcal{O}(\|x - x(\mu)\|). \tag{2.21}
$$

For the matrix in (2.16), we first consider the following decomposition

$$
\sum_{i\in\mathcal{B}} \frac{u_i}{c_i(x)} \nabla c_i(x)\nabla c_i(x)^\top = \widehat{\mathcal{Q}}(x)\mathcal{S}(x)\widehat{\mathcal{Q}}^\top(x), \tag{2.22}
$$

where $\mathcal{S}(x) \in \Re^{|\mathcal{B}|\times|\mathcal{B}|}$ is a diagonal matrix and $\widehat{\mathcal{Q}}(x) \in \Re^{n\times|\mathcal{B}|}$ is an orthonormal matrix. Note that by the linear independence assumption A2 and (2.14), all diagonal elements of $\mathcal{S}(x)$ are of order $\Omega(\mu^{-1})$ and the columns of $\widehat{\mathcal{Q}}(x)$ span the range space of $\nabla c_{\mathcal{B}}(x)$, where $\nabla c_{\mathcal{B}}(x) = [\nabla c_{i_1}(x), ..., \nabla c_{i_l}] \in \Re^{n\times|\mathcal{B}|}$ and $\mathcal{B} \equiv \{i_1, ..., i_l\}$. Let $\widetilde{\mathcal{Q}}(x) \in \Re^{n\times(n-|\mathcal{B}|)}$ be an orthonormal matrix that spans the null space of $\nabla c_{\mathcal{B}}(x)^\top$. Hence, $\nabla c_{\mathcal{B}}(x)^\top\widetilde{\mathcal{Q}}(x) = 0$ and $\mathcal{Q}(x) = \left[\widehat{\mathcal{Q}}(x), \widetilde{\mathcal{Q}}(x)\right]$ is orthogonal.

If we define

$$
\mathcal{Q}_1(x) = \begin{bmatrix} \widehat{\mathcal{Q}}(x) \\ 0 \end{bmatrix} \in \Re^{(n+p)\times|\mathcal{B}|}, \quad \mathcal{Q}_2(x) = \begin{bmatrix} \widetilde{\mathcal{Q}}(x) & 0 \\ 0 & I \end{bmatrix} \in \Re^{(n+p)\times(n+p-|\mathcal{B}|)},
$$

then $[\mathcal{Q}_1(x), \mathcal{Q}_2(x)]$ is an orthogonal matrix and we have

$$
\begin{aligned}
& \begin{bmatrix} \widehat{\mathcal{H}} & \nabla g(x) \\ \nabla g(x)^\top & 0 \end{bmatrix} = [\mathcal{Q}_1(x), \mathcal{Q}_2(x)] \begin{bmatrix} \widehat{\mathcal{M}}_{11}(x) & \widehat{\mathcal{M}}_{12}(x) \\ \widehat{\mathcal{M}}_{12}(x)^\top & \widehat{\mathcal{M}}_{22}(x) \end{bmatrix} \begin{bmatrix} \mathcal{Q}_1(x)^\top \\ \mathcal{Q}_2(x)^\top \end{bmatrix} \\
& = \begin{bmatrix} \widehat{\mathcal{Q}}(x) & \widetilde{\mathcal{Q}}(x) & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \widehat{\mathcal{M}}_{11}(x) & \widehat{\mathcal{M}}_{12}(x) \\ \widehat{\mathcal{M}}_{12}(x)^\top & \widehat{\mathcal{M}}_{22}(x) \end{bmatrix} \begin{bmatrix} \widehat{\mathcal{Q}}(x)^\top & 0 \\ \widetilde{\mathcal{Q}}(x)^\top & 0 \\ 0 & I \end{bmatrix},
\end{aligned} \tag{2.23}
$$

where

$$
\begin{cases}
\widehat{\mathcal{M}}_{11}(x) = \widehat{\mathcal{Q}}(x)^\top\widehat{\mathcal{H}}\widehat{\mathcal{Q}}(x) = \mathcal{S}(x) + \mathcal{O}(1), \\
\widehat{\mathcal{M}}_{22}(x) = \begin{bmatrix} \widetilde{\mathcal{Q}}(x)^\top\mathcal{H}\widetilde{\mathcal{Q}}(x) & \widetilde{\mathcal{Q}}(x)^\top\nabla g(x) \\ \nabla g(x)^\top\widetilde{\mathcal{Q}}(x) & 0 \end{bmatrix} + \mathcal{O}(\mu), \\
\widehat{\mathcal{M}}_{12}(x) = \left[\widehat{\mathcal{Q}}(x)^\top\widehat{\mathcal{H}}\widetilde{\mathcal{Q}}(x), \widehat{\mathcal{Q}}(x)^\top\nabla g(x)\right] = \mathcal{O}(1).
\end{cases} \tag{2.24}
$$

In deriving the above, we have used (2.14) to obtain

$$
\widehat{\mathcal{H}}\widetilde{\mathcal{Q}}(x) = \mathcal{H}\widetilde{\mathcal{Q}}(x) + \sum_{i\in\mathcal{I}\setminus\mathcal{B}} \frac{u_i}{c_i(x)} \nabla c_i(x)\nabla c_i(x)^\top\widetilde{\mathcal{Q}}(x) = \mathcal{H}\widetilde{\mathcal{Q}}(x) + \mathcal{O}(\mu).
$$

By Assumption A and (2.24), it can be readily verified that the $|\mathcal{B}| \times |\mathcal{B}|$ matrix $\widehat{\mathcal{M}}_{11}(x)$ and its $(n+p-|\mathcal{B}|) \times (n+p-|\mathcal{B}|)$ Schur complement

$$\mathcal{G}(x) = \widehat{\mathcal{M}}_{22}(x) - \widehat{\mathcal{M}}_{12}(x)^\top \widehat{\mathcal{M}}_{11}(x)^{-1}\widehat{\mathcal{M}}_{12}(x) \tag{2.25}$$

are nonsingular. Hence,

$$\begin{bmatrix} \widehat{\mathcal{M}}_{11}(x) & \widehat{\mathcal{M}}_{12}(x) \\ \widehat{\mathcal{M}}_{12}(x)^\top & \widehat{\mathcal{M}}_{22}(x) \end{bmatrix}^{-1} = \begin{bmatrix} \widetilde{\mathcal{M}}_{11}(x) & \widetilde{\mathcal{M}}_{12}(x) \\ \widetilde{\mathcal{M}}_{12}(x)^\top & \widetilde{\mathcal{M}}_{22}(x) \end{bmatrix},$$

where

$$\begin{cases} \widetilde{\mathcal{M}}_{11}(x) = \widehat{\mathcal{M}}_{11}(x)^{-1} + \widehat{\mathcal{M}}_{11}(x)^{-1}\widehat{\mathcal{M}}_{12}(x)\mathcal{G}(x)^{-1}\widehat{\mathcal{M}}_{12}(x)^\top \widehat{\mathcal{M}}_{11}(x)^{-1}, \\ \widetilde{\mathcal{M}}_{12}(x) = -\widehat{\mathcal{M}}_{11}(x)^{-1}\widehat{\mathcal{M}}_{12}(x)\mathcal{G}(x)^{-1}, \\ \widetilde{\mathcal{M}}_{22}(x) = \mathcal{G}(x)^{-1}. \end{cases} \tag{2.26}$$

Since all the diagonal elements of $\mathcal{S}(x)$ are of order $\Omega(\mu^{-1})$, we have from (2.24) that $\widehat{\mathcal{M}}_{11}(x)^{-1} = \mathcal{O}(\mu)$. Note that if μ is sufficiently small, the matrix

$$\begin{bmatrix} \widetilde{\mathcal{Q}}(x)^\top \mathcal{H}\widetilde{\mathcal{Q}}(x) & \widetilde{\mathcal{Q}}(x)^\top \nabla g(x) \\ \nabla g(x)^\top \widetilde{\mathcal{Q}}(x) & 0 \end{bmatrix}$$

is uniformly nonsingular by Assumption A and the fact that the columns of $\widetilde{\mathcal{Q}}(x)$ span the null space of $\nabla c_{\mathcal{B}}(x)^\top$. Hence, we have from (2.24), (2.25) and (2.26) that $\widetilde{\mathcal{M}}_{22}(x) = \mathcal{G}(x)^{-1} = \mathcal{O}(1)$ and thus $\widetilde{\mathcal{M}}_{11}(x) = \mathcal{O}(\mu)$ and $\widetilde{\mathcal{M}}_{12}(x) = \mathcal{O}(\mu)$.

Now we can estimate the magnitude of the solution of (2.16) using (2.20), (2.21), (2.23), (2.24) and the fact $\nabla c_{\mathcal{B}}(x)^\top \widetilde{\mathcal{Q}}(x) = 0$,

$$\begin{aligned} \begin{bmatrix} \widetilde{\triangle x} \\ \widetilde{y} - v(\mu) \end{bmatrix} &= \begin{bmatrix} \widehat{\mathcal{H}} & \nabla g(x) \\ \nabla g(x)^\top & 0 \end{bmatrix}^{-1} \begin{bmatrix} -\nabla \varphi_\mu(x) - \nabla g(x) v(\mu) \\ -g(x) - \frac{\|g(x)\|}{r}(v - \widetilde{y}) \end{bmatrix} \\ &= [\mathcal{Q}_1(x), \mathcal{Q}_2(x)] \begin{bmatrix} \widetilde{\mathcal{M}}_{11}(x) & \widetilde{\mathcal{M}}_{12}(x) \\ \widetilde{\mathcal{M}}_{12}(x)^\top & \widetilde{\mathcal{M}}_{22}(x) \end{bmatrix} \begin{bmatrix} -\widehat{\mathcal{Q}}(x)^\top (\nabla \varphi_\mu(x) + \nabla g(x) v(\mu)) \\ \begin{bmatrix} \widetilde{\mathcal{Q}}(x)^\top & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} -\nabla \varphi_\mu(x) - \nabla g(x) v(\mu) \\ -g(x) - \frac{\|g(x)\|}{r}(v - \widetilde{y}) \end{bmatrix} \end{bmatrix} \\ &= [\mathcal{O}(1)), \mathcal{O}(1)] \begin{bmatrix} \mathcal{O}(\mu) & \mathcal{O}(\mu) \\ \mathcal{O}(\mu) & \mathcal{O}(1) \end{bmatrix} \begin{bmatrix} \mathcal{O}(\mu^{-1}\|x - x(\mu)\|) \\ \mathcal{O}(\|x - x(\mu)\|) \end{bmatrix} \\ &= \mathcal{O}(\|x - x(\mu)\|). \end{aligned}$$

From (2.8), it then follows that $\left\|\widehat{\triangle x}\right\| = \mathcal{O}(\|x - x(\mu)\|)$. ◻

The next lemma shows that $\widetilde{\triangle x}$ is a descent direction for the merit function $\Phi_{\mu,r}(\cdot)$.

LEMMA 2.4. *Suppose Assumption A holds. If μ is sufficiently small, $z - z(\mu) = o(\mu)$, $\nu \in (0, 1 - \frac{1}{\eta})$ and $r \geq \max\{\eta\|v(\mu)\|, r_0\}$ with $\eta > 1$ and $r_0 > 0$, then*

$$\Phi'_{\mu,r}(x; \widetilde{\triangle x}) \leq -\nu \widetilde{\triangle x}^\top \widetilde{\mathcal{H}} \widetilde{\triangle x}, \tag{2.27}$$

where

$$\widetilde{\mathcal{H}} = \begin{cases} \widehat{\mathcal{H}} + \frac{r}{\|g(x)\|}\nabla g(x) \nabla g(x)^\top, & \textit{if} \quad \|g(x)\| > 0, \\ \widehat{\mathcal{H}}, & \textit{if} \quad \|g(x)\| = 0. \end{cases} \tag{2.28}$$

Proof. First note that $\widetilde{\triangle x}^\top \widetilde{\mathcal{H}} \widetilde{\triangle x} \geq 0$ by assumption A3. When $\|g(x)\| = 0$, it is easy to verify that $\Phi'_{\mu,r}(x; \widetilde{\triangle x}) = \nabla\varphi_\mu(x)^\top \widetilde{\triangle x}$. See the proof of Lemma 3.1 in [3]. Since in this case $\nabla g(x)^\top \widetilde{\triangle x} = 0$ and $\widehat{\mathcal{H}} = \widetilde{\mathcal{H}}$, it follows from (2.16) that $\Phi'_{\mu,r}(x; \widetilde{\triangle x}) = -\widetilde{\triangle x}^\top \widetilde{\mathcal{H}} \widetilde{\triangle x}$. Hence, (2.27) holds trivially.

Now suppose $\|g(x)\| > 0$. Then it follows from (2.3) that

$$\begin{aligned} \Phi'_{\mu,r}(x; \widetilde{\triangle x}) &= \nabla\Phi_{\mu,r}(x)^\top \widetilde{\triangle x} \\ &= \nabla\varphi_\mu(x)^\top \widetilde{\triangle x} + \tfrac{r}{\|g(x)\|} g(x)^\top \nabla g(x)^\top \widetilde{\triangle x} \\ &= -\widetilde{\triangle x}^\top \widehat{\mathcal{H}} \widetilde{\triangle x} - \widetilde{y}^\top \nabla g(x)^\top \widetilde{\triangle x} + \tfrac{r}{\|g(x)\|} g(x)^\top \nabla g(x)^\top \widetilde{\triangle x} \\ &= -\widetilde{\triangle x}^\top \widetilde{\mathcal{H}} \widetilde{\triangle x} - v^\top \nabla g(x)^\top \widetilde{\triangle x}. \end{aligned} \tag{2.29}$$

To prove (2.27), it suffices to show that $(1-\nu)\widetilde{\triangle x}^\top \widetilde{\mathcal{H}} \widetilde{\triangle x} + v^\top \nabla g(x)^\top \widetilde{\triangle x} \geq 0$. Since $\nu \in (0, 1 - \frac{1}{\eta})$, there is a constant $\bar{\nu}$ such that $\bar{\nu} \in (\frac{1}{\eta(1-\nu)}, 1)$. Moreover, if μ is small enough, we have $r > \tilde{\nu}\eta\|v\|$ for some $\tilde{\nu} \in (\frac{1}{\eta\bar{\nu}(1-\nu)}, 1)$ as $\|z - z(\mu)\| = o(\mu)$ and $r \geq \max\{\eta\|v(\mu)\|, r_0\}$. By assumption A3, for small enough $\|g(x)\|$, we know that

$$\widehat{\mathcal{H}} + \frac{r(1-\bar{\nu})}{\|g(x)\|} \nabla g(x) \nabla g(x)^\top \succ 0, \quad \text{if} \;\; \|g(x)\| > 0. \tag{2.30}$$

Now we can derive from (2.3) that if $\|g(x)\| > 0$,

$$\begin{aligned} & (1-\nu)\widetilde{\triangle x}^\top \widetilde{\mathcal{H}} \widetilde{\triangle x} + v^\top \nabla g(x)^\top \widetilde{\triangle x} \\ = \;& (1-\nu)\widetilde{\triangle x}^\top \left(\widehat{\mathcal{H}} + \tfrac{r(1-\bar{\nu})}{\|g(x)\|} \nabla g(x) \nabla g(x)^\top\right) \widetilde{\triangle x} \\ & +(1-\nu)\tfrac{r\bar{\nu}}{\|g(x)\|} \widetilde{\triangle x}^\top \nabla g(x) \nabla g(x)^\top \widetilde{\triangle x} + v^\top \nabla g(x)^\top \widetilde{\triangle x} \\ \geq \;& (1-\nu)\tfrac{r\bar{\nu}}{\|g(x)\|} \widetilde{\triangle x}^\top \nabla g(x) \nabla g(x)^\top \widetilde{\triangle x} + v^\top \nabla g(x)^\top \widetilde{\triangle x} \\ = \;& (1-\nu)\tfrac{r\bar{\nu}}{\|g(x)\|} \left\| g(x) - \tfrac{\|g(x)\|}{r}(\widetilde{y} - v) \right\|^2 - v^\top \left(g(x) - \tfrac{\|g(x)\|}{r}(\widetilde{y} - v) \right) \\ = \;& r\bar{\nu}(1-\nu)\|g(x)\| - v^\top g(x) + \mathcal{O}\left(\|g(x)\|\|\widetilde{y} - v\|\right) \\ \geq \;& (r - \tilde{\nu}\eta\|v\| + \tilde{\nu}\eta\|v\|)\bar{\nu}(1-\nu)\|g(x)\| - \|v\|\|g(x)\| + o(\|g(x)\|) \\ = \;& (\tilde{\nu}\eta\bar{\nu}(1-\nu) - 1)\|v\|\|g(x)\| + (r - \tilde{\nu}\eta\|v\|)\bar{\nu}(1-\nu)\|g(x)\| \\ & + o(\|g(x)\|) \geq 0, \end{aligned}$$

where the last inequality follows from the fact $\tilde{\nu}\eta\bar{\nu}(1-\nu) > 1$ and $r > \tilde{\nu}\eta\|v\|$. Hence, the lemma follows. ∎

We now show that once we are close to a solution to problem (P_μ), a full second-order correction step yields a sufficient decrease in the penalty function.

THEOREM 2.5. *Suppose Assumption A holds. For $\sigma \in (0, \frac{1}{2})$ and $r \geq \eta\|v(\mu)\|$ with $\eta > 1$, if μ is sufficiently small and $z - z(\mu) = o(\mu)$, then*

$$\Phi_{\mu,r}(x + \widehat{\triangle x}) - \Phi_{\mu,r}(x) \leq \sigma \Phi'_{\mu,r}(x; \widetilde{\triangle x}). \tag{2.31}$$

Proof. By (2.14) and Lemma 2.2, $c(x) > 0$ and $c(x + \widehat{\triangle x}) > 0$. Then by Taylor's theorem we have

$$\begin{aligned} \varphi_\mu(x + \widehat{\triangle x}) = \varphi_\mu(x) + \nabla\varphi_\mu(x)^\top \widehat{\triangle x} + \tfrac{1}{2}\widehat{\triangle x}^\top \nabla^2 f(x + \xi\widehat{\triangle x})\widehat{\triangle x} \\ + \tfrac{\mu}{2}\widehat{\triangle x}^\top \textstyle\sum_{i\in\mathcal{I}} \left(\frac{\nabla c_i(x + \xi\widehat{\triangle x}) \nabla c_i(x + \xi\widehat{\triangle x})^\top}{c_i(x + \xi\widehat{\triangle x})^2} - \frac{\nabla^2 c_i(x + \xi\widehat{\triangle x})}{c_i(x + \xi\widehat{\triangle x})} \right) \widehat{\triangle x} \end{aligned} \tag{2.32}$$

for some $\xi \in [0,1]$. Lemma 2.3 implies that $\left\|\widehat{\triangle x}\right\| = \mathcal{O}(\|(x - x(\mu))\|) = o(\mu)$, which together with (2.14) give that $\frac{\left\|\widehat{\triangle x}\right\|}{c_i(x)} = o(1)$. Hence, similar to (2.18), we have

$$\frac{1}{c_i(x+\xi\widehat{\triangle x})} = \frac{1}{c_i(x)} + \mathcal{O}\left(\frac{\left\|\widehat{\triangle x}\right\|}{c_i(x)^2}\right), \forall\, i \in \mathcal{I}. \tag{2.33}$$

This together with the fact from (2.8) that $\left\|\widehat{\triangle x} - \widetilde{\triangle x}\right\| = \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|^2\right)$ yields the following estimates,

$$\widehat{\triangle x}^\top \nabla^2 f(x+\xi\widehat{\triangle x})\widehat{\triangle x} = \widehat{\triangle x}^\top \nabla^2 f(x)\widehat{\triangle x} + \mathcal{O}\left(\left\|\widehat{\triangle x}\right\|^3\right) = \widetilde{\triangle x}^\top \nabla^2 f(x)\widetilde{\triangle x} + \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|^3\right), \tag{2.34}$$

$$\begin{aligned} \frac{\nabla c_i(x+\xi\widehat{\triangle x})^\top \widehat{\triangle x}}{c_i(x+\xi\widehat{\triangle x})} &= \nabla c_i(x)^\top \widehat{\triangle x}\left(\frac{1}{c_i(x)} + \mathcal{O}\left(\frac{\left\|\widehat{\triangle x}\right\|}{c_i(x)^2}\right)\right) + \mathcal{O}\left(\frac{\left\|\widehat{\triangle x}\right\|^2}{c_i(x)}\right) \\ &= \nabla c_i(x)^\top \widetilde{\triangle x}\left(\frac{1}{c_i(x)} + \mathcal{O}\left(\frac{\left\|\widetilde{\triangle x}\right\|}{c_i(x)^2}\right)\right) + \mathcal{O}\left(\frac{\left\|\widetilde{\triangle x}\right\|^2}{c_i(x)}\right), \quad \forall i \in \mathcal{I}, \end{aligned} \tag{2.35}$$

$$\begin{aligned} \frac{\widehat{\triangle x}^\top \nabla^2 c_i(x+\xi\widehat{\triangle x})\widehat{\triangle x}}{c_i(x+\xi\widehat{\triangle x})} &= \frac{\widehat{\triangle x}^\top \nabla^2 c_i(x)\widehat{\triangle x}}{c_i(x)} + \mathcal{O}\left(\frac{\left\|\widehat{\triangle x}\right\|^3}{c_i(x)^2}\right) \\ &= \frac{\widetilde{\triangle x}^\top \nabla^2 c_i(x)\widetilde{\triangle x}}{c_i(x)} + \mathcal{O}\left(\frac{\left\|\widetilde{\triangle x}\right\|^3}{c_i(x)^2}\right), \quad \forall\, i \in \mathcal{I}. \end{aligned} \tag{2.36}$$

Combining (2.32), (2.34), (2.35) and (2.36) and using the fact that $\left\|\widehat{\triangle x} - \widetilde{\triangle x}\right\| = \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|^2\right)$ from (2.8), we obtain

$$\begin{aligned} &\varphi_\mu(x+\widehat{\triangle x}) - \varphi_\mu(x) = \nabla\varphi_\mu(x)^\top\widehat{\triangle x} + \tfrac{1}{2}\widetilde{\triangle x}^\top\left(\nabla^2 f(x) - \textstyle\sum_{i\in\mathcal{I}}\frac{\mu}{c_i(x)}\nabla^2 c_i(x)\right)\widetilde{\triangle x} \\ &+\tfrac{\mu}{2}\textstyle\sum_{i\in\mathcal{I}}\left(\frac{\nabla c_i(x)^\top\widetilde{\triangle x}}{c_i(x)}\right)^2 + \mathcal{O}\left(\mu^{-2}\left\|\widetilde{\triangle x}\right\|\right)\sum_{i\in\mathcal{I}}\left(\nabla c_i(x)^\top\widetilde{\triangle x}\right)^2 + \mathcal{O}\left(\mu^{-1}\left\|\widetilde{\triangle x}\right\|^3\right). \end{aligned} \tag{2.37}$$

From (2.8) and (2.4) we have

$$\begin{aligned} & g_i(x+\widehat{\triangle x}) \\ = \;& g_i(x+\widetilde{\triangle x}) + \nabla g_i(x+\widehat{\triangle x})^\top\left(\widehat{\triangle x} - \widetilde{\triangle x}\right) + \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|^4\right) \\ = \;& g_i(x+\widetilde{\triangle x}) + \nabla g_i(x)^\top(\widehat{\triangle x} - \widetilde{\triangle x}) + \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|^3\right) \\ = \;& \tfrac{\|g(x)\|}{r}(\widehat{y}_i - v_i) + \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|^3\right), \quad \forall i \in \mathcal{E}. \end{aligned} \tag{2.38}$$

On the other hand, we have from (2.3), (2.8) and the fact $\|g(x)\| = \mathcal{O}(\|x - x(\mu)\|) = \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|\right)$ from Lemma 2.3 that

$$
\begin{aligned}
& g_i(x + \widehat{\triangle x}) \\
= \;& \left(g_i(x) + \nabla g_i(x)^\top \widetilde{\triangle x}\right) + \nabla g_i(x)^\top \left(\widehat{\triangle x} - \widetilde{\triangle x}\right) + \tfrac{1}{2}(\widehat{\triangle x})^\top \nabla^2 g_i(x) \widehat{\triangle x} + \mathcal{O}\left(\left\|\widehat{\triangle x}\right\|^3\right) \\
= \;& \tfrac{\|g(x)\|}{r}(\widetilde{y}_i - v_i) + \nabla g_i(x)^\top (\widehat{\triangle x} - \widetilde{\triangle x}) + \tfrac{1}{2}(\widetilde{\triangle x})^\top \nabla^2 g_i(x) \widetilde{\triangle x} + \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|^3\right) \\
= \;& \tfrac{\|g(x)\|}{r}(\widehat{y}_i - v_i) + \nabla g_i(x)^\top (\widehat{\triangle x} - \widetilde{\triangle x}) + \tfrac{1}{2}(\widetilde{\triangle x})^\top \nabla^2 g_i(x) \widetilde{\triangle x} + \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|^3\right) \quad \forall i \in \mathcal{E}.
\end{aligned} \tag{2.39}
$$

Comparing the last equalities of (2.38) and (2.39), we obtain that

$$
\nabla g_i(x)^\top (\widetilde{\triangle x} - \widehat{\triangle x}) = \frac{1}{2}(\widetilde{\triangle x})^\top \nabla^2 g_i(x) \widetilde{\triangle x} + \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|^3\right), \quad \forall i \in \mathcal{E}. \tag{2.40}
$$

From (2.16) we have that $\nabla \varphi_\mu(x) = -\widehat{\mathcal{H}}\widetilde{\triangle x} + \nabla g(x)\widetilde{y}$, where $\widehat{\mathcal{H}}$ is defined by (2.17). This together with (2.40), (2.8) and (2.14) give that for any constant ϱ

$$
\begin{aligned}
& \nabla \varphi_\mu(x)^\top \widehat{\triangle x} = \nabla \varphi_\mu(x)^\top \widetilde{\triangle x} - \nabla \varphi_\mu(x)^\top \left(\widetilde{\triangle x} - \widehat{\triangle x}\right) \\
= \;& \left(\tfrac{1}{2} - \varrho\right) \nabla \varphi_\mu(x)^\top \widetilde{\triangle x} - \left(\tfrac{1}{2} + \varrho\right) \left(\widetilde{\triangle x}^\top \widehat{\mathcal{H}} + \widetilde{y}^\top \nabla g(x)^\top\right) \widetilde{\triangle x} \\
& + \left(\widetilde{\triangle x}^\top \widehat{\mathcal{H}} + \widetilde{y}^\top \nabla g(x)^\top\right) \left(\widetilde{\triangle x} - \widehat{\triangle x}\right) \\
= \;& \left(\tfrac{1}{2} - \varrho\right) \nabla \varphi_\mu(x)^\top \widetilde{\triangle x} - \left(\tfrac{1}{2} + \varrho\right) \left(\widetilde{\triangle x}^\top \widehat{\mathcal{H}} + \widetilde{y}^\top \nabla g(x)^\top\right) \widetilde{\triangle x} \\
& + \tfrac{1}{2} \widetilde{\triangle x}^\top \textstyle\sum_{i \in \mathcal{E}} \widetilde{y}_i \nabla^2 g_i(x) \widetilde{\triangle x} + \mathcal{O}\left(\mu^{-1} \left\|\widetilde{\triangle x}\right\|^3\right).
\end{aligned} \tag{2.41}
$$

If $\|g(x)\| > 0$, we have from the third equation of (2.3) that

$$
\begin{aligned}
& \left(\tfrac{1}{2} + \varrho\right) \widetilde{y}^\top \nabla g(x)^\top \widetilde{\triangle x} \\
= \;& \tfrac{1}{2} \widetilde{y}^\top \left(\tfrac{\|g(x)\|}{r}(\widetilde{y} - v) - g(x)\right) + \tfrac{\varrho r}{\|g(x)\|} \left(\nabla g(x)^\top \widetilde{\triangle x} + g(x) + \tfrac{\|g(x)\|}{r} v\right)^\top \nabla g(x)^\top \widetilde{\triangle x} \\
= \;& -\tfrac{1}{2} g(x)^\top \widetilde{y} + \tfrac{\varrho r}{\|g(x)\|} \left\|\widetilde{\triangle x}^\top \nabla g(x)\right\|^2 + \tfrac{\varrho r}{\|g(x)\|} \left(g(x) + \tfrac{\|g(x)\|}{r} v\right) \left(\tfrac{\|g(x)\|}{r}(\widetilde{y} - v) - g(x)\right) + o(\|g(x)\|) \\
= \;& -\tfrac{1}{2} g(x)^\top \widetilde{y} + \tfrac{\varrho r}{\|g(x)\|} \left\|\widetilde{\triangle x}^\top \nabla g(x)\right\|^2 - \varrho r \|g(x)\| - \varrho g(x)^\top v + o(\|g(x)\|).
\end{aligned} \tag{2.42}
$$

Since $\|z - z(\mu)\| = o(\mu)$ and strict complementarity holds, if μ is sufficiently small, then $\frac{u_i}{2c_i(x)}$ is sufficiently large for all $i \in \mathcal{B}$, and moreover, if $\|g(x)\| > 0$, $\frac{r}{\|g(x)\|}$ is also sufficiently large. Hence, we know from assumption A3 that

$$
\mathcal{H} + \sum_{i \in \mathcal{B}} \frac{u_i}{2c_i(x)} \nabla c_i(x) \nabla c_i(x)^\top + \frac{r}{\|g(x)\|} \nabla g(x) \nabla g(x)^\top \succeq 0, \quad \text{if} \quad \|g(x)\| > 0. \tag{2.43}
$$

If $\|g(x)\| = 0$, assumption A3 and the fact $\nabla g(x)^\top \widetilde{\triangle x} = 0$ gives that

$$
(\widetilde{\triangle x})^\top \left(\mathcal{H} + \sum_{i \in \mathcal{B}} \frac{u_i}{2c_i(x)} \nabla c_i(x) \nabla c_i(x)^\top\right) \widetilde{\triangle x} \geq 0, \quad \text{if} \quad \|g(x)\| = 0. \tag{2.44}
$$

Moreover, from (2.14), the fact $\|z - z(\mu)\| = o(\mu)$ and Lemma 2.3, we have $u_i - \frac{\mu}{c_i(x)} = o(1)$ and thus for $\varrho > 0$,

$$\left(\frac{1}{2} + \frac{\varrho}{2}\right) \frac{u_i}{c_i(x)} - \frac{1}{2}\left(\frac{\mu}{c_i(x)^2} + \mathcal{O}\left(\mu^{-2}\left\|\widetilde{\triangle x}\right\|\right)\right) > 0, \quad \forall i \in \mathcal{B}. \tag{2.45}$$

Now let us require that $\varrho \in \left(0, \min\left(\frac{1}{2} - \sigma, \frac{\eta - 1}{2}\right)\right)$. Combining (2.37), (2.41), (2.42), (2.43), (2.45) and (2.14), we obtain for the case $\|g(x)\| > 0$ that

$$\begin{aligned}
& \varphi_\mu(x + \widehat{\triangle x}) - \varphi_\mu(x) \\
= \;& \left(\tfrac{1}{2} - \varrho\right) \nabla\varphi_\mu(x)^\top \widetilde{\triangle x} - \left(\tfrac{1}{2} + \varrho\right) \widetilde{\triangle x}^\top \left(\mathcal{H} + \textstyle\sum_{i \in \mathcal{I}} \frac{u_i}{c_i(x)} \nabla c_i(x) \nabla c_i(x)^\top\right) \widetilde{\triangle x} \\
& + \tfrac{1}{2} g(x)^\top \widetilde{y} - \tfrac{\varrho r}{\|g(x)\|} \widetilde{\triangle x}^\top \nabla g(x) \nabla g(x)^\top \widetilde{\triangle x} + \varrho r \|g(x)\| + \varrho g(x)^\top v \\
& + \tfrac{1}{2} \widetilde{\triangle x}^\top \textstyle\sum_{i \in \mathcal{E}} \widetilde{y}_i \nabla^2 g_i(x) \widetilde{\triangle x} + \tfrac{1}{2} \widetilde{\triangle x}^\top \left(\nabla^2 f(x) - \sum_{i \in \mathcal{B}} \frac{\mu}{c_i(x)} \nabla^2 c_i(x)\right) \widetilde{\triangle x} \\
& + \tfrac{1}{2} \widetilde{\triangle x}^\top \textstyle\sum_{i \in \mathcal{B}} \left(\left(\frac{\mu}{c_i(x)^2} + \mathcal{O}\left(\mu^{-2}\left\|\widetilde{\triangle x}\right\|\right)\right) \nabla c_i(x) \nabla c_i(x)^\top\right) \widetilde{\triangle x} \\
& + o(\|g(x)\|) + \mathcal{O}\left(\mu \left\|\widetilde{\triangle x}\right\|^2\right) + \mathcal{O}\left(\mu^{-1}\left\|\widetilde{\triangle x}\right\|^3\right).
\end{aligned}$$

Then using the fact that $\mathcal{H} = \nabla^2 f(x) - \sum_{i \in \mathcal{I}} u_i \nabla^2 c_i(x) + \sum_{i \in \mathcal{E}} v_i \nabla^2 g_i(x)$, we obtain

$$\begin{aligned}
& \varphi_\mu(x + \widehat{\triangle x}) - \varphi_\mu(x) \\
= \;& \left(\tfrac{1}{2} - \varrho\right) \nabla\varphi_\mu(x)^\top \widetilde{\triangle x} + \tfrac{1}{2} g(x)^\top \widetilde{y} + \varrho r \|g(x)\| + \varrho g(x)^\top v \\
& - \varrho \widetilde{\triangle x}^\top \left(\mathcal{H} + \textstyle\sum_{i \in \mathcal{B}} \frac{u_i}{2 c_i(x)} \nabla c_i(x) \nabla c_i(x)^\top + \frac{r}{\|g(x)\|} \nabla g(x) \nabla g(x)^\top\right) \widetilde{\triangle x} \\
& + \tfrac{1}{2} \widetilde{\triangle x}^\top \textstyle\sum_{i \in \mathcal{E}} (\widetilde{y}_i - v_i) \nabla^2 g_i(x) \widetilde{\triangle x} + \tfrac{1}{2} \widetilde{\triangle x}^\top \sum_{i \in \mathcal{B}} \left(u_i - \frac{\mu}{c_i(x)}\right) \nabla^2 c_i(x) \widetilde{\triangle x} \\
& + \tfrac{1}{2} \widetilde{\triangle x}^\top \textstyle\sum_{i \in \mathcal{I} \setminus \mathcal{B}} u_i \left(\nabla^2 c_i(x) - \frac{1 + 2\varrho}{c_i(x)} \nabla c_i(x) \nabla c_i(x)^\top\right) \widetilde{\triangle x} \\
& - \widetilde{\triangle x}^\top \textstyle\sum_{i \in \mathcal{B}} \left(\left(\frac{1}{2} + \frac{\varrho}{2}\right) \frac{u_i}{c_i(x)} - \frac{1}{2}\left(\frac{\mu}{c_i(x)^2} + \mathcal{O}\left(\mu^{-2}\left\|\widetilde{\triangle x}\right\|\right)\right)\right) \nabla c_i(x) \nabla c_i(x)^\top \widetilde{\triangle x} \\
& + o(\|g(x)\|) + \mathcal{O}\left(\mu \left\|\widetilde{\triangle x}\right\|^2\right) + \mathcal{O}\left(\mu^{-1}\left\|\widetilde{\triangle x}\right\|^3\right) \\
\le \;& \left(\tfrac{1}{2} - \varrho\right) \nabla\varphi_\mu(x)^\top \widetilde{\triangle x} + \tfrac{1}{2} g(x)^\top \widetilde{y} + \varrho r \|g(x)\| + \varrho g(x)^\top v + o(\|g(x)\|) + o\left(\left\|\widetilde{\triangle x}\right\|^2\right).
\end{aligned} \tag{2.46}$$

In a similar way to Lemma 3.2 in [3], it can be readily verified from (2.3) that

$$\Phi'_{\mu,r}(x; \widetilde{\triangle x}) = \nabla\varphi_\mu(x)^\top \widetilde{\triangle x} - r\|g(x)\| + g(x)^\top (\widetilde{y} - v). \tag{2.47}$$

Let $\bar{\varrho} \in (0, 1)$ be a constant such that $\bar{\varrho}\eta > 2\varrho + 1$. Since $r \ge \eta \|y(\mu)\|$, it follows that if μ is sufficiently small, $r > \bar{\varrho}\eta \max\{\|\widetilde{y}\|, \|v\|\}$. Now we can obtain from (2.38), (2.46), (2.47) and Lemma 2.4 that

$$\begin{aligned}
& \Phi_{\mu,r}(x + \widehat{\triangle x}) - \Phi_{\mu,r}(x) = \varphi_\mu(x + \widehat{\triangle x}) - \varphi_\mu(x) + r(\|g(x + \widehat{\triangle x})\| - g(x)) \\
= \;& \left(\tfrac{1}{2} - \varrho\right) \Phi'_{\mu,r}(x; \widetilde{\triangle x}) + g(x)^\top \left(\varrho \widetilde{y} + \tfrac{1}{2} v\right) - \tfrac{r}{2} \|g(x)\| + o(\|g(x)\|) + o\left(\left\|\widetilde{\triangle x}\right\|^2\right) \\
\le \;& \sigma \Phi'_{\mu,r}(x; \widetilde{\triangle x}) - \left(\tfrac{\bar{\varrho}\eta}{2} - \left(\varrho + \tfrac{1}{2}\right)\right) \max\{\|\widetilde{y}\|, \|v\|\} \\
& - \left(\tfrac{r}{2} - \tfrac{\bar{\varrho}\eta}{2} \max\{\|\widetilde{y}\|, \|v\|\}\right) \|g(x)\| + o(\|g(x)\|) + o\left(\left\|\widetilde{\triangle x}\right\|^2\right) \\
\le \;& \sigma \Phi'_{\mu,r}(x; \widetilde{\triangle x}).
\end{aligned}$$

The case that $\|g(x)\| = 0$ is straightforward by (2.47) since $\nabla g(x)^\top \widetilde{\triangle x} = 0$. ◻

3. Local analysis for the overall algorithm. Suppose the current iterate z satisfies the following criteria for terminating the inner algorithm,

$$\|\mathcal{R}_\mu(z)\| \le \epsilon_\mu,\ c(x) > 0,\ u > 0. \tag{3.1}$$

We study in this section the outcome of applying our quasi-feasible method when μ and ϵ_μ are decreased to μ^+ and ϵ_{μ^+}, respectively. At this time, the μ in the right hand sides of (1.4), (2.3) and (2.4) is replaced by μ^+.

THEOREM 3.1. *Suppose conditions (3.1) and Assumption A hold. If $z \in \mathcal{N}(z^*)$ and $\widetilde{z}$ is computed by solving (2.3) with μ replaced by μ^+, then $\|\widetilde{z} - z^*\| = \mathcal{O}(\|z - z^*\|^2) + \mathcal{O}(\mu^+)$.*

Proof. Analogously to (2.9), the following linear system can be derived from (2.3)

$$\mathcal{M}_r(z)(\widetilde{z} - z^*) = \begin{bmatrix} \mathcal{H}(x - x^*) - \nabla_x \mathcal{L}(x, u^*, v^*) \\ \mathcal{U}\nabla c(x)^\top (x - x^*) - C(x)u^* + \mu^+ e \\ \frac{\|g(x)\|}{r}(v^* - v) - g(x) + \nabla g(x)^\top (x - x^*) \end{bmatrix}. \tag{3.2}$$

Also, analogously to (2.10) and (2.12), we have

$$\begin{array}{l} \mathcal{H}(x - x^*) - \nabla_x \mathcal{L}(x, u^*, v^*) = \mathcal{O}(\|z - z^*\|^2), \\ \frac{\|g(x)\|}{r}(v^* - v) - g(x) + \nabla g(x)^\top (x - x^*) = \mathcal{O}(\|z - z^*\|^2). \end{array}$$

Moreover, we have for $i \in \mathcal{B}$

$$\begin{array}{rl} & u_i \nabla c(x)^\top (x - x^*) - c_i(x)u_i^* + \mu^+ \\ = & u_i(\nabla c(x)^\top (x - x^*) - c_i(x)) - c_i(x)(u_i^* - u_i) + \mu^+ \\ = & \mathcal{O}(\|z - z^*\|^2) + \mathcal{O}(\mu^+), \end{array}$$

and for $i \in \mathcal{I}\backslash\mathcal{B}$,

$$\begin{array}{rl} & u_i \nabla c(x)^\top (x - x^*) - c_i(x)u_i^* + \mu^+ \\ = & (u_i - u_i^*)\nabla c(x)^\top (x - x^*) + \mu^+ \\ = & \mathcal{O}(\|z - z^*\|^2) + \mathcal{O}(\mu^+). \end{array}$$

The result then follows from the fact that $\|\mathcal{M}_r(z)^{-1}\|$ is uniformly bounded when $z \in \mathcal{N}(z^*)$. ∎

THEOREM 3.2. *Suppose conditions (3.1) and Assumption A hold. If z is sufficiently close to z^*, μ is sufficiently small and $\mu^+ \le \mu$, then $\|\mathcal{R}_{\mu^+}(\widetilde{z})\| = \mathcal{O}((\epsilon_\mu + \mu)^2)$.*

Proof. Since z is sufficiently close to z^*, by Taylor's theorem, we have

$$\|z - z(\mu)\| = \left\|\mathcal{R}_0'(z(\mu))^{-1}\left(\mathcal{R}_\mu(z) + o(\|z - z(\mu)\|)\right)\right\| = \mathcal{O}(\|\mathcal{R}_\mu(z)\|).$$

Theorem 2.1 gives that $\|\widetilde{z} - z(\mu^+)\| = \mathcal{O}(\|z - z(\mu^+)\|^2)$, where $\tilde{z}$ is computed by solving (2.3) with μ replaced by μ^+. Since $z(\mu)$ is locally Lipschitz continuous if μ is sufficiently small, we have $\|z(\mu) - z(\mu^+)\| = \mathcal{O}(\|z(\mu) - z^*\|) = \mathcal{O}(\mu)$. Using (3.1) we obtain

$$\begin{array}{rl} & \|\mathcal{R}_{\mu^+}(\widetilde{z})\| = \|\mathcal{R}_{\mu^+}(\widetilde{z}) - \mathcal{R}_{\mu^+}(z(\mu^+))\| \\ = & \mathcal{O}(\|\widetilde{z} - z(\mu^+)\|) = \mathcal{O}(\|z - z(\mu^+)\|^2) \\ = & \mathcal{O}(\|z - z(\mu) + z(\mu) - z(\mu^+)\|^2) \\ = & \mathcal{O}((\epsilon_\mu + \mu)^2). \end{array}$$

∎

THEOREM 3.3. *Suppose conditions (3.1) and Assumption A hold. If z is sufficiently close to z^*, μ is sufficiently small, $\mu^+ \le \mu$ and $(\mu + \epsilon_\mu)^2 = o(\mu^+)$, then $c(x + \widetilde{\triangle x}) > 0$ and $\widetilde{\lambda} > 0$.*

Proof. We have from (3.2) with z^* replaced by z that

$$\begin{aligned} & \|\widetilde{z} - z\| = \|\mathcal{M}_r(z)^{-1}\mathcal{R}_{\mu^+}(z)\| \\ = \; & \left\|\mathcal{M}_r(z)^{-1}\left(\mathcal{R}_\mu(z) + (\mu^+ - \mu)(0, e, 0)^\top\right)\right\| \\ = \; & \mathcal{O}(\epsilon_\mu) + \mathcal{O}(\mu), \end{aligned}$$

where $\tilde{z}$ is computed by solving (2.3) with μ replaced by μ^+ and the last equation uses the fact that $\mu^+ \leq \mu$ and $\|\mathcal{M}_r(z)^{-1}\|$ is uniformly bounded if z is sufficiently close to z^*. Also from (2.3) we have

$$\begin{aligned} & \widetilde{\lambda}_i c_i(x + \widetilde{\triangle x}) \\ = \; & \widetilde{\lambda}_i \left(c_i(x) + \nabla c_i(x)^\top \widetilde{\triangle x} + \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|^2 \right) \right) \\ = \; & \widetilde{\lambda}_i c_i(x) + u_i \nabla c_i(x)^\top \widetilde{\triangle x} + (\widetilde{\lambda}_i - u_i)\nabla c_i(x)^\top \widetilde{\triangle x} + \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|^2 \right) \\ = \; & \mu^+ + \mathcal{O}(\|\widetilde{z} - z\|^2) = \mu^+ + \mathcal{O}((\epsilon_\mu + \mu)^2) > 0, \quad \forall \, i \in \mathcal{I}. \end{aligned} \tag{3.3}$$

For $i \in \mathcal{B}$, $\widetilde{\lambda}_i > 0$ by strict complementarity and thus $c_i(x + \widetilde{\triangle x}) > 0$ by (3.3). Also, for $i \in \mathcal{I} \backslash \mathcal{B}$, $c_i(x + \widetilde{\triangle x}) > 0$ and hence $\widetilde{\lambda}_i > 0$ by (3.3). ☐

4. Algorithm and convergence. In this section we describe our modified interior-point ℓ_2-penalty method and establish its global and local superlinear convergence.

4.1. Algorithm. We use indexes k and j to denote an inner iterate and an outer iterate, respectively. As in the method in [3], the Hessian of the Lagrangian $\mathcal{H}^k = \nabla^2_{xx}\mathcal{L}(x^k, u^k, v^k)$ at the current iterate x^k is modified if necessary so that the following condition holds:

$$\begin{cases} d^\top \widetilde{\mathcal{H}^k} d \geq \bar{\nu}\|d\|^2, \; \forall d \in \Re^n, & \text{if } \|g(x^k)\| > 0, \\ d^\top \widetilde{\mathcal{H}^k} d \geq \bar{\nu}\|d\|^2, \; \forall d \text{ s.t. } \nabla g(x^k)^\top d = 0, & \text{if } \|g(x^k)\| = 0, \end{cases} \tag{4.1}$$

for some $\bar{\nu} > 0$ and all k, where

$$\widetilde{\mathcal{H}^k} = \begin{cases} \widehat{\mathcal{H}^k}, & \text{if } \|g(x^k)\| = 0, \\ \widehat{\mathcal{H}^k} + \frac{r_k}{\|g(x^k)\|}\nabla g(x^k)\nabla g(x^k)^\top, & \text{if } \|g(x^k)\| > 0, \end{cases} \tag{4.2}$$

and

$$\widehat{\mathcal{H}^k} = \mathcal{H}^k + \nabla c(x^k) C(x^k)^{-1} \mathcal{U}^k \nabla c(x^k)^\top. \tag{4.3}$$

We note that condition (4.1) holds in a neighborhood of a local minimizer of problem (P) that satisfies the second-order sufficient conditions. However, outside of such a neighborhood, it may be necessary to modify $\mathcal{H}^k$ so that (4.1) holds. One approach for doing so is a delicate inertia controlling strategy described in [6]. Alternatively, one can simply add a suitable multiple of the identity to $\mathcal{H}^k$ as done in the software package IPOPT [13].

Our modified interior-point method solves at each iteration the linear systems (2.3) and (2.4) in addition to (1.4). Since, as shown in the last section, the step $\widetilde{\triangle x^k}$ determined by (2.3) is able to generate fast local convergence, one would expect this step to be a good choice for a line search direction. Unfortunately, however, the direction $\widetilde{\triangle x^k}$ is not necessarily a descent direction for the merit function $\Phi_{\mu,r}(\cdot)$. To guarantee global convergence, the following conditions are checked to determine if the solution of (2.3) is acceptable as a search direction,

$$\begin{cases} \text{(i)} & \left\|\widetilde{\triangle x^k} - \triangle x^k\right\| \leq \bar{\kappa} \max\left\{ \left\|\widetilde{\triangle x^k}\right\|^\vartheta, \left\|\widetilde{\triangle x^k}\right\|^{1/\vartheta} \right\}, \\ \text{(ii)} & \Phi'_{\mu_j, r_k}(x^k; \widetilde{\triangle x^k}) \leq -\nu \widetilde{\triangle x^k}^\top \widetilde{\mathcal{H}^k} \widetilde{\triangle x^k}, \end{cases} \tag{4.4}$$

where $\bar{\kappa} > 0$, $\vartheta \in (0,1)$, $\nu \in (0,1)$ and $\Phi'_{\mu_j,r_k}(x^k;\widetilde{\triangle x^k})$ can be computed from (2.47). Condition (i) of (4.4) ensures that $\widetilde{\triangle x^k}$ does not differ too much from $\triangle x^k$ since along the latter direction, global convergence can be guaranteed. Condition (ii) of (4.4) ensures that $\widetilde{\triangle x^k}$ is a descent direction for $\Phi_{\mu_j,r_k}(\cdot)$. If (4.4) holds, we set $(d^x,d^u,d^v) = \left(\widetilde{\triangle x^k},\widetilde{\lambda^k},\widetilde{y^k}\right)$; otherwise, we set $(d^x,d^u,d^v) = \left(\triangle x^k,\lambda^k,y^k\right)$. From Lemma 4.1 in [3] we have

$$\Phi'_{\mu_j,r_k}(x^k;\triangle x^k) = \nabla\varphi(x^k)^\top\widetilde{\triangle x^k} - r_k\|g(x^k)\| + g(x^k)^\top y^k. \tag{4.5}$$

Given a step d^x, we obtain the next iterate

$$x^{k+1} = x^k + t_k d^x \tag{4.6}$$

by a basic backtracking line search procedure; i.e., we consider a decreasing sequence of step sizes $t_{k,l} = \beta^l$ $(l = 0,1,...)$ with $\beta \in (0,1)$ until the following criteria are satisfied,

$$\begin{cases} \text{(i)} & c(x^k + t_{k,l}d^x) > 0; \\ \text{(ii)} & \Phi_{\mu_j,r_k}(x^k + t_{k,l}d^x) - \Phi_{\mu_j,r_k}(x^k) \le \sigma t_{k,l}\Phi'_{\mu_j,r_k}(x^k;d^x), \end{cases} \tag{4.7}$$

where $\sigma \in (0,\frac{1}{2})$. Since the second-order correction step involves extra function evaluations, it may not be invoked at the beginning of the optimization process. However, when a KKT solution is approached, the following condition is checked to determine whether to accept the step,

$$\Phi_{\mu_j,r_k}(x^k + \widehat{\triangle x^k}) - \Phi_{\mu_j,r_k}(x^k) \le \sigma\Phi'_{\mu_j,r_k}(x^k;\widetilde{\triangle x^k}). \tag{4.8}$$

The dual iterate is set as follows: for $i \in \mathcal{I}$

$$u_i^{k+1} = \begin{cases} d_i^u, & \text{if } \min\{\theta u_i^k, \frac{\mu_j}{c_i(x^k)}\} \le d_i^u \le \frac{\mu_j\gamma}{c_i(x^k)}, \\ \min\{\theta u_i^k, \frac{\mu_j}{c_i(x^k)}\}, & \text{if } d_i^u < \min\{\theta u_i^k, \frac{\mu_j}{c_i(x^k)}\}, \\ \frac{\mu_j\gamma}{c_i(x^k)}, & \text{if } d_i^u > \frac{\mu_j\gamma}{c_i(x^k)}, \end{cases} \tag{4.9}$$

where $\theta \in (0,1)$ and $\gamma > 1$.

To update the penalty parameter, we use a different rule from that used in [3]. In particular, we check the following conditions at each iteration,

$$\begin{cases} \text{(i)} & \|\triangle x^k\| \le \pi_k \\ \text{(ii)} & \kappa_1\mu_j e \le C(x^k)\lambda^k \le \kappa_2\mu_j e, \\ \text{(iii)} & r_k < \eta\|y^k\|, \end{cases} \tag{4.10}$$

where $\pi_k > 0$, $0 < \kappa_1 < 1 < \kappa_2$ and $\eta > 1$. If all conditions (i)-(iii) hold, we increase the penalty parameter.

We now give an outline of our modified quasi-feasible interior-point ℓ_2-penalty method.

Algorithm 3.1. Quasi-Feasible Interior-Point ℓ_2-Penalty Method.

Parameters: $\epsilon_{\text{tol}} > 0$, $\chi > 1$, $\sigma \in (0,\frac{1}{2})$, $\gamma > 1$, $\theta \in (0,1)$, $\eta > 1$, $\nu \in \left(0, 1-\frac{1}{\eta}\right)$, $\varsigma \in (0,1)$,
$\beta \in (0,1)$, $\vartheta \in (0,1)$, $\bar{\kappa} > 0$, $0 < \kappa_1 < 1 < \kappa_2$.

Initial data: $x^0 \in \Re^n$ with $c(x^0) > 0$, $(u^0,v^0) \in \Re^{m+p}$ with $u^0 > 0$, $r_0 > 0$,
$\pi_0 > 0$, $\mu_0 > 0$, $\epsilon_{\mu_0} > 0$.

$k \leftarrow 0$, $j \leftarrow 0$

Step 1. Check convergence for the overall problem.
If $\|\mathcal{R}_0(x^k,u^k,v^k)\| \le \epsilon_{\text{tol}}$, STOP with SUCCESS.

Step 2. Check convergence for the barrier problem.
If $\|\mathcal{R}_{\mu_j}(x^k,u^k,v^k)\| \le \epsilon_{\mu_j}$, decrease μ_j and ϵ_{μ_j} respectively to μ_{j+1}

and $\epsilon_{\mu_{j+1}}$. Set $j \leftarrow j+1$.

Step 3. Compute search direction.
Set $\mathcal{H}^k \leftarrow \nabla^2_{xx}\mathcal{L}(x^k, u^k, v^k)$. Modify $\mathcal{H}^k$, if necessary, so that (4.1) holds.
If $\mathcal{M}_{r_k}(\mathcal{H}^k, x^k, u^k)$ is singular, STOP with MFCQ FAILURE.
Compute $(\triangle x^k, \lambda^k, y^k)$ from (1.4).
Compute $(\widetilde{\triangle x^k}, \widetilde{\lambda^k}, \widetilde{y^k})$ from (2.3).

Step 4. Check fast convergence.
If $\left\|\mathcal{R}_{\mu_j}(x^k + \widetilde{\triangle x^k}, \widetilde{\lambda^k}, \widetilde{y^k})\right\| \le \epsilon_{\mu_j}$, $c(x^k + \widetilde{\triangle x^k}) > 0$
and $\widetilde{y^k} > 0$, set $(x^{k+1}, u^{k+1}, v^{k+1}) \leftarrow (x^k + \widetilde{\triangle x^k}, \widetilde{\lambda^k}, \widetilde{y^k})$,
$k \leftarrow k+1$ and go to Step 1.

Step 5. Second-order correction.
If (4.4) holds, compute $(\widehat{\triangle x^k}, \widehat{\lambda^k}, \widehat{y^k})$ from (2.4) and set $(d^x, d^u, d^v) \leftarrow \left(\widetilde{\triangle x^k}, \widetilde{\lambda^k}, \widetilde{y^k}\right)$;
otherwise, set $(d^x, d^u, d^v) \leftarrow \left(\triangle x^k, \lambda^k, y^k\right)$ and go to Step 6.
If (4.8) holds, set $x^{k+1} \leftarrow x^k + \widehat{\triangle x^k}$, $v^{k+1} \leftarrow \widehat{y^k}$, $d^u \leftarrow \widehat{\lambda^k}$, compute u^{k+1} from (4.9)
and go to Step 7; otherwise, go to Step 6.

Step 6. Line search and update.
Compute x^{k+1} by (4.6) and (4.7).
Set u^{k+1} as in (4.9) and $v^{k+1} \leftarrow d^v$.

Step 7. Penalty parameter update.
If all conditions in (4.10) hold, set $r_{k+1} \leftarrow \chi r_k$, $\pi_{k+1} \leftarrow \varsigma\pi_k$;
otherwise, set $r_{k+1} \leftarrow r_k$, $\pi_{k+1} \leftarrow \pi_k$. Set $k \leftarrow k+1$ and go to Step 1.

PROPOSITION 4.1. *Algorithm 3.1 is well defined.*

Proof. First, the second-order correction step is well defined. If (4.4) holds and (4.8) fails at some iteration k, we know from condition (ii) of (4.4) that the line search step of Algorithm 3.1 is well defined as f, c and g are twice continuously differentiable; otherwise, Algorithm 3.1 reduces to the method described in [3] except for the termination criteria. Hence, we conclude from Proposition 4.2 in [3] that Algorithm 3.1 is well defined. ◻

4.2. Global convergence. To guarantee global convergence, the following assumptions used in [3] are also needed here.

Assumption B.

B1. The primal iterates $\{x^k\}$ generated by Algorithm 3.1 lie in a bounded set.

B2. The Hessian estimates $\{\mathcal{H}^k\}$ generated by Algorithm 3.1 are bounded.

Using a different rule for updating the penalty parameter, it is proved in [3] that under Assumption B, if the penalty parameter r_k tends to infinity, there is an accumulation point of the iterates that is either a Fritz-John (FJ) point of problem (P) or a FJ point of the feasibility problem:

$$\text{minimize} \quad \|g(x)\|^2 \quad \text{subject to} \quad c(x) \ge 0. \tag{4.11}$$

The next result shows that this property holds as well for the new rule (4.10).

THEOREM 4.2. *Suppose Assumption B holds. If r_k is increased infinitely many times, then there exists a limit point of $\{x^k\}$ generated by Algorithm 3.1 that is either a FJ point of problem (P) at which the Mangasarian-Fromovitz constraint qualification (MFCQ) fails to hold or a FJ point of problem (4.11).*

Proof. Since r_k is increased infinitely many times, there exists an infinite index set $\mathcal{K}$ such that $\pi_{k+1} = \varsigma\pi_k$ and $r_{k+1} = \chi r_k$. This implies that $\{\pi_k\} \to 0$ and $\{r_k\} \to \infty$ as $\varsigma \in (0,1)$ and $\chi > 1$. The conditions that trigger the increase of r_k must be satisfied. Hence, we know from (4.10) that $\{\|\triangle x^k\|\}_{\mathcal{K}} \to 0$ and $\{\|(\lambda^k, y^k)\|\}_{\mathcal{K}} \to \infty$. By assumption B1, there exists an infinite set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{x^k\}_{\bar{\mathcal{K}}} \to \bar{x}$ and

$$\left\{\frac{1}{\|(\lambda^k, y^k)\|}(\lambda^k, y^k)\right\}_{\bar{\mathcal{K}}} \to (\bar{\lambda}, \bar{y})$$

with $\|(\bar\lambda, \bar y)\| = 1$ and $\bar\lambda \geq 0$ by (ii) of (4.10). Since $c(x^k) > 0$ for all k, $c(\bar x) \geq 0$. There are two cases.

Case 1. $\|g(\bar x)\| = 0$. Dividing the first equation in (1.4) and (ii) of (4.10) by $\|(\lambda^k, y^k)\|$ and letting $k \in \bar{\mathcal{K}} \to \infty$ yields that $\bar\lambda_i = 0$ if $c_i(\bar x) > 0$ and

$$-\sum_{i\in\mathcal{I}, c_i(\bar x)=0} \bar\lambda_i \nabla c_i(\bar x) + \nabla g(\bar x)\bar y = 0.$$

This implies that $\bar x$ is a FJ point of problem (P) failing to satisfy the MFCQ.

Case 2. $\|g(\bar x)\| > 0$. We have from the third equation of (1.4) that

$$\|\nabla g(x^k)^\top \triangle x^k\| = \|g(x^k)\| \left\| \frac{1}{r_k} y^k - \frac{1}{\|g(x^k)\|} g(x^k) \right\|.$$

Since $\{\|\triangle x\|\}_{\bar{\mathcal{K}}} \to 0$ and $\|g(\bar x)\| > 0$, it follows that $\{\frac{1}{r_k} y^k\}_{\bar{\mathcal{K}}} \to \frac{1}{\|g(\bar x)\|} g(\bar x)$. Hence, dividing the first equation in (1.4) and (ii) of (4.10) by $\|(\lambda^k, y^k)\|$ and letting $k \in \bar{\mathcal{K}} \to \infty$ yields that $\bar\lambda_i = 0$ if $c_i(\bar x) > 0$ and

$$\frac{K}{\|g(\bar x)\|} g(\bar x)\nabla g(\bar x) = \sum_{i\in\mathcal{I}, c_i(\bar x)=0} \bar\lambda_i \nabla c_i(\bar x),$$

where $K = \lim_{k\in\bar{\mathcal{K}}} \left(r_k/\|(\lambda^k, y^k)\|\right) \geq 0$. Hence, we conclude that $\bar x$ is a FJ point of problem (4.11). ∎

LEMMA 4.3. *Suppose condition (4.1) and Assumption B hold and $r_k = \bar r$ for all k large enough. For any fixed $\mu_j > 0$, if $\epsilon_{\mu_j} = 0$ and the solutions of (1.4), (2.3) and (2.4) are unbounded for $k \in \mathcal{K}$, where $\mathcal{K}$ is an infinite index set, then any accumulation point of $\{x^k\}_{\mathcal{K}}$ is a FJ point of problem (P) at which the MFCQ fails to hold.*

Proof. Under condition (4.1) and Assumption B, the proof is identical to the proof of Lemma 3.8 in [3]. ∎

LEMMA 4.4. *Suppose condition (4.1) and Assumption B hold and $r_k = \bar r$ for all k large enough. For any fixed $\mu_j > 0$, if $\epsilon_{\mu_j} = 0$ and the solutions of (1.4), (2.3) and (2.4) are bounded for all k, then $\{\triangle x^k\} \to 0$.*

Proof. There are two cases to consider.

Case 1. There exists an infinite index set $\mathcal{K}$ such that (4.4) holds for all $k \in \mathcal{K}$. First consider any infinite set $\mathcal{K}' \subseteq \mathcal{K}$ such that (4.8) holds for all $k \in \mathcal{K}'$. Then we have from (4.1) that

$$\Phi_{\mu_j,\bar r}(x^{k+1}) - \Phi_{\mu_j,\bar r}(x^k) \leq -\nu\bar\nu\sigma \left\|\widetilde{\triangle x^k}\right\|^2.$$

This implies that $\left\{\widetilde{\triangle x^k}\right\}_{\mathcal{K}'} \to 0$ by assumption B1.

Now consider any infinite set $\mathcal{K}' \subseteq \mathcal{K}$ such that (4.8) fails for all $k \in \mathcal{K}'$. Since $\{\Phi_{\mu_j,\bar r}(x^k)\}$ is monotone decreasing by Algorithm 3.1, $\{c(x^k)\}$ is componentwise bounded greater than zero and $\{u^k\}$ is bounded above (see Lemma 3.7 in [3]). Note that the solutions of (1.4) and (2.3) are bounded. Consider an arbitrary infinite set $\bar{\mathcal{K}} \subseteq \mathcal{K}'$ for which $\left\{\widetilde{\triangle x^k}\right\}_{\bar{\mathcal{K}}} \to \overline{\triangle x}$ and $\{x^k\}_{\bar{\mathcal{K}}} \to \bar x$. Clearly, since $c(\bar x) > 0$, there exists a $\bar t_1 \in (0,1]$ such that for all $t \in (0, \bar t_1]$, $c(\bar x + t\overline{\triangle x}) > 0$. By continuity, we know from (2.47) that

$$\left\{\Phi'_{\mu_j,r_k}(x^k; \widetilde{\triangle x^k})\right\}_{\bar{\mathcal{K}}} \to \Phi'_{\mu_j,\bar r}(\bar x; \overline{\triangle x}), \tag{4.12}$$

and then from condition (ii) of (4.4) and (4.1) that

$$\Phi'_{\mu_j,\bar r}(\bar x; \overline{\triangle x}) \leq -\nu\bar\nu \left\|\overline{\triangle x}\right\|^2 \leq 0. \tag{4.13}$$

Suppose $\overline{\triangle x} \neq 0$. Then there exists a $\bar t \in (0, \bar t_1]$ such that for all $t \in (0, \bar t]$,

$$\Phi_{\mu_j,\bar r}(\bar x + t\overline{\triangle x}) - \Phi_{\mu_j,\bar r}(\bar x) \leq 1.1\sigma t \Phi'_{\mu_j,\bar r}(\bar x; \overline{\triangle x}), \tag{4.14}$$

as $\sigma \in (0, \frac{1}{2})$. Let $\bar{l} = \min\{l | \beta^l \in (0, \bar{t}], l = 0, 1, 2, ...\}$. From the continuity of $\Phi_{\mu_j, \bar{r}}(\cdot)$, we know from (4.12) and (4.14) that for $t_{\bar{l}} = \beta^{\bar{l}}$ and $k \in \bar{\mathcal{K}}$ large enough,

$$\text{(4.15)} \qquad \begin{cases} \text{(i)} & c_i(x^k + t_{\bar{l}}\widetilde{\triangle x^k}) > 0, \ \forall i \in \mathcal{I}; \\ \text{(ii)} & \Phi_{\mu_j, \bar{r}}(x^k + t_{\bar{l}}\widetilde{\triangle x^k}) - \Phi_{\mu_j, \bar{r}}(x^k) \le \sigma t_{\bar{l}} \Phi'_{\mu_j, r_k}(x^k; \widetilde{\triangle x^k}). \end{cases}$$

This implies that the step size $t_k \ge t_{\bar{j}}$ for all $k \in \bar{\mathcal{K}}$ large enough according to the line search criteria (4.7). Thus, we have from (4.13) and (ii) of (4.15) that $\{\Phi_{\mu_j, \bar{r}}(x^k)\} \to -\infty$, a contradiction to assumption B1 and the continuity of $f(\cdot)$, $c(\cdot)$ and $g(\cdot)$. Hence, we conclude that $\left\{\widetilde{\triangle x^k}\right\}_{\mathcal{K}} \to 0$. Furthermore, it follows from condition (i) of (4.4) that $\{\triangle x^k\}_{\mathcal{K}} \to 0$.

Case 2. There exists an infinite index set $\mathcal{K}$ such that (4.4) fails for all $k \in \mathcal{K}$. Then Algorithm 3.1 reduces to the method described in [3]. Using an argument identical to the one in the proof of Lemma 3.9 in [3], we have that $\{\triangle x^k\}_{\mathcal{K}} \to 0$. □

LEMMA 4.5. *Suppose condition (4.1) and Assumption B hold and $r_k = \bar{r}$ for all k large enough. For any fixed $\mu_j > 0$, if $\epsilon_{\mu_j} = 0$ and the solutions of (1.4), (2.3) and (2.4) are bounded for all k, then any accumulation point of $\{x^k\}$ is a KKT point of problem (P_μ). Moreover, if the MFCQ holds at these accumulation points, the whole sequence $\{(x^k, u^k, v^k)\}$ converges to a point $(x(\mu), u(\mu), v(\mu))$ satisfying the KKT conditions (1.1).*

Proof. Since $\{\triangle x^k\} \to 0$ by Lemma 4.4 and $r_k = \bar{r}$, it follows from (1.4) that (i) and (ii) of (4.10) are satisfied for all large k, while (iii) is eventually violated. Hence, from the third equation of (1.4) we have $\|\nabla g(x^k)^\top \triangle x^k\| \ge \left(1 - \frac{1}{\eta}\right) \|g(x^k)\|$, which implies $\|g(x^k)\| \to 0$ as $\eta > 1$. Since the solutions of (1.4) are bounded, letting $k \to \infty$ in (1.4) gives that any accumulation point of $\{(x^k, \lambda^k, y^k)\}$ satisfies (1.1). Moreover, if the MFCQ holds at these accumulation points, the coefficient matrices of (1.4) are nonsingular in the limit. Therefore, (2.3) and (2.4) have the same limit as (1.4) since $\|g(x^k)\| \to 0$. Thus, we conclude that $\left\{\widetilde{\triangle x^k}\right\} \to 0$, $\left\{\widehat{\triangle x^k}\right\} \to 0$, $\{x^k\} \to x(\mu)$, $\{(\lambda^k, y^k)\} \to (u(\mu), v(\mu))$, $\left\{\left(\widetilde{\lambda^k}, \widetilde{y^k}\right)\right\} \to (u(\mu), v(\mu))$, $\left\{\left(\widehat{\lambda^k}, \widehat{y^k}\right)\right\} \to (u(\mu), v(\mu))$, where $x(\mu)$ is a KKT point of (P_μ) and $(u(\mu), v(\mu))$ is the unique multiplier vector under the MFCQ. Finally, it is easy to see from (4.9) that $\{(u^k, v^k)\} \to (u(\mu), v(\mu))$. □

To analyze the overall algorithm, we assume that Algorithm 3.1 succeeds for every μ_j. We associate index k_j with the iteration at which μ_j is decreased. The following global convergence result is an analog of Theorem 3.13 in [3].

THEOREM 4.6. *Suppose that condition (4.1) and Assumption B hold, $\epsilon_{\text{tol}} = 0$ and Algorithm 3.1 generates an infinite sequence of iterates. Then when $\{\mu_j\} \to 0$ and $\{\epsilon_{\mu_j}\} \to 0$, if $\{(x^{k_j}, u^{k_j}, v^{k_j})\}$ is bounded, any of its limit points satisfies the first-order optimality conditions for problem (P); otherwise, there exists a limit point of $\{x^{k_j}\}$ that is a FJ point of problem (P) at which the MFCQ fails.*

4.3. Fast local convergence. In this subsection we assume that the termination tolerance is set zero and Algorithm 3.1 generates an infinite sequence of iterates. Moreover, we assume that $r_k = \bar{r}$ eventually and a subsequence of the iterates $\{(x^{k_j}, u^{k_j}, v^{k_j})\}$ converges (x^*, u^*, v^*), which satisfies the first-order optimality conditions for problem (P) and Assumption A.

THEOREM 4.7. *If μ_j is sufficiently small, $\epsilon_{\mu_j} = 0$ and $\|z^k - z(\mu)\| = o(\mu)$, then z^k converges to $z(\mu)$ Q-quadratically, i.e., $\|z^{k+1} - z(\mu)\| = \mathcal{O}(\|z^k - z(\mu)\|^2)$.*

Proof. First, $\mathcal{H}^k = \nabla^2_{xx}\mathcal{L}(z^k)$ since (4.1) is eventually met under Assumption A. Subtracting (1.4) from (2.3) gives that $\left\|\widetilde{\triangle x} - \triangle x\right\| = \mathcal{O}(\|g(x)\|) = \mathcal{O}(\|x - x(\mu)\|)$. Then from 2.3 we have $\left\|\widetilde{\triangle x} - \triangle x\right\| = \mathcal{O}\left(\left\|\widetilde{\triangle x}\right\|\right)$. Hence, condition (i) of (4.4) eventually holds since $\vartheta \in (0, 1)$. Clearly, the first two conditions in (4.10) hold for all large k and hence (iii) of (4.10) is violated as $r^k = \bar{r}$ eventually. This implies that the condition $\bar{r} \ge \max\{\eta\|v(\mu)\|, r_0\}$ in Lemma 2.4 is met. Hence, by Lemma 2.4, condition (ii) of (4.4) holds eventually. Thus, it follows that $\widetilde{\triangle x^k}$ is used as the search direction for all large k. By Lemma 2.2 and Theorem 2.6, we know from Step 5 of Algorithm 3.1 that $z^{k+1} = \left(x^k + \widehat{\triangle x^k}, \widehat{\lambda^k}, \widehat{y^k}\right)$. Hence, the quadratic convergence follows from Theorem 2.1. □

THEOREM 4.8. *Suppose ϵ_{μ_j} and μ_j are decreased so that*

$$\begin{cases} (\epsilon_{\mu_j} + \mu_j)^2 = o(\epsilon_{\mu_{j+1}}), \\ \mu_{j+1} = o(\|\mathcal{R}_0(x^{k_j}, u^{k_j}, v^{k_j}\|). \end{cases} \tag{4.16}$$

Then for all j large enough,

$$(x^{k_{j+1}}, u^{k_{j+1}}, v^{k_{j+1}}) = (x^{k_j+1}, u^{k_j+1}, v^{k_j+1}), \tag{4.17}$$

and

$$\left\|(x^{k_{j+1}}, u^{k_{j+1}}, v^{k_{j+1}}) - (x^*, u^*, v^*)\right\| = o\left(\left\|(x^{k_j}, u^{k_j}, v^{k_j}) - (x^*, u^*, v^*)\right\|\right).$$

Proof. First, $\mathcal{H}^k = \nabla^2_{xx}\mathcal{L}(z^k)$ as (4.1) is eventually met under Assumption A. Since there exists a subsequence of iterates converging to (x^*, u^*, v^*), we know $\{(x^{k_j}, u^{k_j}, v^{k_j})\}$ is sufficiently close to (x^*, u^*, v^*) for some sufficiently large j. For such j, μ_j and ϵ_{μ_j} are respectively decreased to μ_{j+1} and $\epsilon_{\mu_{j+1}}$ at Step 2 of Algorithm 3.1. When the algorithm further proceeds to Step 4, we know from the first condition of (4.16) and Theorems 3.2 and 3.3 that the conditions in Step 4 are satisfied for $j+1$. Then Algorithm 3.1 proceeds to Step 2 with $j+1$ being increased to $j+2$. Hence, $k_j + 1 = k_{j+1}$ and (4.17) follows. Moreover, since

$$\mathcal{R}_0(x^{k_j}, u^{k_j}, v^{k_j}) = \mathcal{O}(\|(x^{k_j}, u^{k_j}, v^{k_j}) - (x^*, u^*, v^*)\|),$$

the superlinear convergence then follows from the second condition of (4.16) and Theorem 3.1. □

Several strategies for updating μ_j and ϵ_{μ_j} have been discussed in [1]. In particular, two strategies are given that guarantee (4.16) and thus superlinear convergence. Both strategies assume that ϵ_{μ_j} is some constant proportional to μ_j, i.e., $\epsilon_{\mu_j} = \alpha\mu_j$ for all j with $\alpha \in [0, \sqrt{m})$. The first strategy chooses $\mu_{j+1} = \mu_j^{1+\delta}$ with $\delta \in (0,1)$, while second strategy chooses $\mu_{j+1} = \|\mathcal{R}_0(x^{k_j}, u^{k_j}, v^{k_j})\|^{1+\delta}$ for all large j. It can be readily verified that the two strategies can be used in our method as well to generate superlinear convergence.

5. Examples and conclusion. In this section we give some examples that illustrate the local convergence properties established in the previous sections. For simplicity we will not present all the implementational details and parameter settings. We will present such details and extensive computational results in a forthcoming paper. We shall refer to a quasi-feasible or an infeasible version of Algorithm 3.1 as "Algorithm I" and the original algorithms in [3] as "Algorithm II". We note that in our implementation the only difference between Algorithms I and II is that the latter never computes or uses the correction steps $\left(\widetilde{\triangle x}, \widetilde{\lambda}, \widetilde{y}\right)$ and $\left(\widehat{\triangle x}, \widehat{\lambda}, \widehat{y}\right)$. All other aspects are the same including the line search and the penalty parameter update.

Our first example is Example 1 in [3], first used in [12]. This example illustrates that the global convergence properties in [3] are preserved by Algorithm I and that Algorithm I does not necessarily improve the convergence rate of Algorithm II (both exhibited fast convergence as shown in Tables 5.5 and 5.6).
Example 1.

$$\begin{array}{ll} \min & x_1 \\ \text{s.t.} & x_1^2 - x_2 - 1 = 0, \\ & x_1 - x_3 - 0.5 = 0, \\ & x_2 \geq 0, x_3 \geq 0. \end{array}$$

Our second example is problem 42 in [9]. This example clearly shows the effect of the local correction steps on the convergence rate , i.e. superlinear convergence of Algorithm I compared with linear convergence of Algorithm II.
Example 2.

$$\begin{array}{ll} \min & (x_1-1)^2 + (x_2-2)^2 + (x_3-3)^2 + (x_4-4)^2 \\ \text{s.t.} & x_1 = 2, \\ & x_3^2 + x_4^2 = 2, \\ & x_1, x_2, x_3, x_4 \geq 0. \end{array}$$

In our first set of tests we chose $\mu = 0.001$ and ran the algorithms until the optimality error $\|\mathcal{R}_\mu(z)\|_\infty$ was less than $1\mathrm{E}-14$, where $\|\cdot\|_\infty$ denotes the infinity norm. Here we are essentially forcing the algorithms to solve the log-barrier problems for a fixed value of μ to optimality to illustrate the quadratic convergence of the iterates of Algorithm I to the solution of the barrier problem (see Theorem 4.7). Note that $\|\mathcal{R}_\mu(z)\|_\infty$ can be used to measure the magnitude of the distance of the point z to the solution $z(\mu)$, i.e., $\|z - z(\mu)\| = \Omega(\|\mathcal{R}_\mu(z)\|_\infty)$, provided $z(\mu)$ satisfies certain second-order sufficient conditions, see, e.g., [14]. It is obvious from Tables 5.1 and 5.3 that the optimality error $\|\mathcal{R}_\mu(\cdot)\|_\infty$ converges to zero quadratically for Algorithm I. However, Tables 5.2 and 5.4 show that for some iterations, Algorithm II only converges linearly. Moreover, Tables 5.2 and 5.4 show that $\|\mathcal{R}_\mu(\cdot)\|_\infty = \|g(\cdot)\|_\infty$ on most iterations, where $\|g(\cdot)\|_\infty$ denotes the infeasibility measure, indicating that slow convergence of the iterates might be caused by the slow decrease of the constraint violation.

Now we look at the overall behavior of Algorithms I and II with a sequence of superlinearly decreasing barrier parameters μ. In the implementation we set the inner algorithm tolerance ϵ_μ and updated μ as follows,

$$\epsilon_{\mu_j} = 10\mu_j, \quad \mu_{j+1} = \min\{0.2\mu_j, \mu_j^{1.5}\}, \quad j = 1, 2, ...$$

The algorithms terminated when the KKT error $\|\mathcal{R}_0(z)\|_\infty$ was less than $1\mathrm{E}-10$. Note that we again have $\|z - z(0)\| = \Omega(\|\mathcal{R}_0(z)\|_\infty)$ under the SOSC [14]. Tables 5.5 and 5.6 show that both Algorithms I and II exhibit fast local convergence on Example 1, although Algorithm II takes more than one iteration to solve a barrier problem when μ is small. The benefit of the correction steps introduced in this paper becomes clearer in Example 2. Tables 5.7 and 5.8 indicate that Algorithm I converges superlinearly (see Theorem 4.8), while Algorithm II converges linearly. Again, we see from Table 5.8 that $\|g(\cdot)\|_\infty$ decreases very slowly and for most iterations $\|\mathcal{R}_0(\cdot)\|_\infty = \|g(\cdot)\|_\infty$.

Table 5.1
Example 1: Algorithm I for $\mu = 0.001$

k	$\varphi_\mu(\cdot)$	$\|\mathcal{R}_\mu(\cdot)\|_\infty$	$\|g(\cdot)\|_\infty$
11	1.17E+00	2.09E-01	3.55E-02
12	1.02E+00	2.34E-02	2.34E-02
13	1.01E+00	3.47E-04	3.47E-04
14	1.01E+00	3.69E-08	1.31E-09
15	1.01E+00	4.81E-15	1.33E-15

In this paper we have presented a modified interior-point ℓ_2-penalty method for nonlinear programming based on methods that we originally proposed in [3]. We have shown that the method possesses fast local convergence properties in addition to the strong global convergence properties established in [3]. Specifically, under the standard nondegenerate assumptions, we have shown that for each fixed barrier parameter μ, if the iterates generated by the proposed method are within a neighborhood of radius $o(\mu)$ of the solution to the barrier problem, they converge to the solution quadratically. The overall convergence rate of the iterates to the solution of the nonlinear program is superlinear. These properties are a consequence of two local correction steps that incur only modest cost compared with the methods in [3].

Acknowledgments. We would like to thank Andreas Wächter for his help in implementing our algorithms within Ipopt.

REFERENCES

[1] R. H. Byrd, G. Liu and J. Nocedal, *On the local behavior of an interior point method for nonlinear programming.* In: D. F. Griffiths, D. J. Higham (eds), Numerical Analysis 1997, pp. 37-56. Addison-Wesley Longman, Reading, MA, USA, 1997.

TABLE 5.2
Example 1: Algorithm II for $\mu = 0.001$

k	$\varphi_\mu(\cdot)$	$\|\mathcal{R}_\mu(\cdot)\|_\infty$	$\|g(\cdot)\|_\infty$
11	1.07E+00	3.66E+00	1.28E-01
12	1.04E+00	6.62E-02	5.34E-03
13	1.01E+00	1.12E-03	1.12E-03
14	1.01E+00	5.56E-05	5.56E-05
15	1.01E+00	2.77E-06	2.77E-06
16	1.01E+00	1.38E-07	1.38E-07
17	1.01E+00	6.88E-09	6.88E-09
18	1.01E+00	3.43E-10	3.43E-10
19	1.01E+00	1.71E-11	1.71E-11
20	1.01E+00	8.53E-13	8.53E-13
21	1.01E+00	4.25E-14	4.25E-14
22	1.01E+00	2.25E-14	2.25E-14
23	1.01E+00	1.19E-14	1.19E-14
24	1.01E+00	7.77E-16	7.77E-16

TABLE 5.3
Example 2: Algorithm I for $\mu = 0.001$

k	$\varphi_\mu(\cdot)$	$\|\mathcal{R}_\mu(\cdot)\|_\infty$	$\|g(\cdot)\|_\infty$
1	1.39E+01	6.67E-01	2.47E-02
2	1.39E+01	2.29E-02	1.91E-03
3	1.39E+01	6.04E-05	5.66E-06
4	1.39E+01	9.57E-11	6.99E-11
5	1.39E+01	4.43E-17	0.00E-00

[2] R. H. BYRD, M. MARAZZI AND J. NOCEDAL, *On the convergence of Newton iterations to non-stationary points*, Math. Programming, 99 (2004), pp. 127-148.

[3] L. CHEN AND D. GOLDFARB, *Interior-point ℓ_2-penalty methods for nonlinear programming with strong global convergence properties*, Math. Programming, 108 (2006), pp. 1-36.

[4] F. FACCHINEI AND S. LUCIDI, *Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minization problems,* J. Optim. Theory Appl., 85 (1995), pp. 265-289.

[5] R. FLETCHER AND S. LEYFFER, *Nonlinear Programming without a penalty function*, Math. Programming, 91 (2002), pp. 239-269.

[6] A. FORSGREN, *Inertia-controlling factorizations for optimization algorithms*, Appl. Numer. Math., 43 (2002), pp. 91-107.

[7] A. FORSGREN, P. E. GILL AND M. H. WRIGHT, *Interior methods for nonlinear optimization*, SIAM Rev., 44 (2002), pp. 525-597.

[8] J. GAUVIN, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Math. Programming, 12 (1977), pp. 136-138.

[9] W. HOCK AND K. SCHITTKOWSKI, *Tests examples for nonlinear programming codes,* volume 187 of *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, 1981.

[10] C. KANZOW AND H. QI, *A QP-free constrained Newton-type method for variational inequality problems,* Math. Programming, 85 (1999), pp. 81-106.

[11] S. M. ROBINSON, *Strongly regular generalized equations,* Math. Oper. Res., 5 (1980), pp. 43-62.

[12] A. WÄCHTER AND L. T. BIEGLER, *Failure of global convergence for a class of interior point methods for nonlinear programming*, Math. Programming, 88 (2000), pp. 565-574.

[13] ———, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Math. Programming, 106 (2006), pp. 25-57.

[14] S. J. WRIGHT, *Effects of finite-precision arithmetic on interior-point methods for nonlinear programming*, SIAM J. Optim., 12 (2001), 36-78.

[15] ———, *On the convergence of the Newton/log-barrier method*, Math. Programming, 90 (2001), pp. 71-100.

Table 5.4
Example 2: Algorithm II for $\mu = 0.001$

k	$\varphi_\mu(\cdot)$	$\|\mathcal{R}_\mu(\cdot)\|_\infty$	$\|g(\cdot)\|_\infty$
1	1.39E+01	6.67E-01	2.47E-02
2	1.38E+01	2.29E-02	8.04E-03
3	1.39E+01	2.05E-03	2.05E-03
4	1.39E+01	5.20E-04	5.20E-04
5	1.39E+01	1.32E-04	1.32E-04
6	1.39E+01	3.34E-05	3.34E-05
7	1.39E+01	8.48E-06	8.48E-06
8	1.39E+01	2.15E-06	2.15E-06
9	1.39E+01	5.45E-07	5.45E-07
10	1.39E+01	1.38E-07	1.38E-07
11	1.39E+01	3.51E-08	3.51E-08
12	1.39E+01	8.89E-09	8.89E-09
13	1.39E+01	2.26E-09	2.26E-09
14	1.39E+01	5.72E-10	5.72E-10
15	1.39E+01	1.45E-10	1.45E-10
16	1.39E+01	3.68E-11	3.68E-11
17	1.39E+01	9.33E-12	9.33E-12

Table 5.5
Example 1: Algorithm I

k	$f(\cdot)$	μ	$\|\mathcal{R}_0(\cdot)\|_\infty$	$\|\mathcal{R}_\mu(\cdot)\|_\infty$	$\|g(\cdot)\|_\infty$
12	1.07E+00	1.00E-01	1.34E-00	1.34E-00	1.23E-01
13	1.26E+00	1.00E-01	3.10E-01	3.10E-01	2.55E-02
14	1.06E+00	2.00E-02	4.22E-02	4.22E-02	4.22E-02
15	1.00E+00	2.83E-03	5.19E-03	4.11E-03	2.81E-03
16	1.00E+00	1.50E-04	1.62E-04	2.10E-05	2.83E-06
17	1.00E+00	1.84E-06	1.89E-06	7.34E-08	2.22E-08
18	1.00E+00	2.51E-09	2.51E-09	1.00E-11	3.16E-12
19	1.00E+00	1.25E-13	1.25E-13	2.22E-16	2.22E-16

Table 5.6
Example 1: Algorithm II

k	$f(\cdot)$	μ	$\|\mathcal{R}_0(\cdot)\|_\infty$	$\|\mathcal{R}_\mu(\cdot)\|_\infty$	$\|g(\cdot)\|_\infty$
10	9.49E-01	1.00E-01	4.69E-00	4.69E-00	1.18E-01
11	1.03E+00	1.00E-01	3.55E-01	3.55E-01	3.67E-03
12	1.01E+00	2.00E-02	2.43E-02	1.05E-02	1.59E-04
13	1.00E+00	1.50E-04	6.83E-04	5.32E-04	1.19E-04
14	1.00E+00	1.84E-06	5.97E-06	5.97E-06	5.97E-06
15	1.00E+00	2.51E-09	2.98E-07	2.98E-07	2.98E-07
16	1.00E+00	2.51E-09	1.49E-08	1.49E-08	1.49E-08
17	1.00E+00	1.25E-13	7.46E-10	7.46E-10	7.46E-10
18	1.00E+00	1.25E-13	3.73E-11	3.73E-11	3.73E-11

Table 5.7
Example 2: Algorithm I

k	$f(\cdot)$	μ	$\|\mathcal{R}_0(\cdot)\|_\infty$	$\|\mathcal{R}_\mu(\cdot)\|_\infty$	$\|g(\cdot)\|_\infty$
1	1.39E+01	1.00E-01	6.10E-01	6.10E-01	2.47E-02
2	1.39E+01	2.00E-02	3.72E-02	1.72E-02	1.69E-03
3	1.39E+01	2.83E-03	2.96E-03	1.36E-04	1.29E-05
4	1.39E+01	1.50E-04	1.51E-04	9.20E-07	3.27E-08
5	1.39E+01	1.84E-06	1.85E-06	2.79E-09	7.76E-11
6	1.39E+01	2.51E-09	2.51E-09	4.26E-13	1.15E-14
7	1.39E+01	1.25E-13	1.25E-13	1.38E-15	0.00E-00

Table 5.8
Example 2: Algorithm II

k	$f(\cdot)$	μ	$\|\mathcal{R}_0(\cdot)\|_\infty$	$\|\mathcal{R}_\mu(\cdot)\|_\infty$	$\|g(\cdot)\|_\infty$
1	1.39E+01	1.00E-01	6.10E-01	6.10E-01	2.47E-02
2	1.38E+01	2.00E-02	3.66E-02	1.66E-02	7.94E-03
3	1.39E+01	2.83E-03	3.00E-03	2.03E-03	2.03E-03
4	1.39E+01	1.50E-04	5.15E-04	5.15E-04	5.15E-04
5	1.39E+01	1.84E-06	1.31E-04	1.31E-04	1.31E-04
6	1.39E+01	1.84E-06	3.31E-05	3.31E-05	3.31E-05
7	1.39E+01	1.84E-06	8.40E-06	8.40E-06	8.40E-06
8	1.39E+01	2.51E-09	2.13E-06	2.13E-06	2.13E-06
9	1.39E+01	2.51E-09	5.40E-07	5.40E-07	5.40E-07
10	1.39E+01	2.51E-09	1.37E-07	1.37E-07	1.37E-07
11	1.39E+01	2.51E-09	3.47E-08	3.47E-08	3.47E-08
12	1.39E+01	2.51E-09	8.80E-09	8.80E-09	8.80E-09
13	1.39E+01	1.25E-13	2.23E-09	2.23E-09	2.23E-09
14	1.39E+01	1.25E-13	5.66E-10	5.66E-10	5.66E-10
15	1.39E+01	1.25E-13	1.43E-10	1.43E-10	1.43E-10
16	1.39E+01	1.25E-13	3.64E-11	3.64E-11	3.64E-11