AN INTERIOR-POINT PIECEWISE LINEAR PENALTY METHOD FOR NONLINEAR PROGRAMMING

LIFENG CHEN* AND DONALD GOLDFARB†

July 17, 2007

Abstract. We present an interior-point penalty method for nonlinear programming (NLP), where the merit function consists of a piecewise linear penalty function (PLPF) and an ℓ_2-penalty function. The PLPF is defined by a set of penalty parameters that correspond to break points of the PLPF and are updated at every iteration. The ℓ_2-penalty function, like traditional penalty functions for NLP, is defined by a single penalty parameter. At every iteration the step direction is computed from a regularized Newton system of the first-order equations of the barrier problem proposed in [4]. Iterates are updated using line search. In particular, a trial point is accepted if it provides a sufficient reduction in either the PLPF or the ℓ_2-penalty function. We show that the proposed method has the same strong global convergence properties as those established in [4]. Moreover, our method enjoys fast local convergence. Specifically, for each fixed small barrier parameter μ, iterates in a small neighborhood (roughly within $o(\mu)$) of the minimizer of the barrier problem converge Q-quadratically to the minimizer. The overall convergence rate of the iterates to the solution of the nonlinear program is Q-superlinear.

Key words. nonlinear programming, interior-point method, penalty method, piecewise linear penalty function, line search, global convergence, superlinear convergence

AMS subject classifications. 49M37, 65F05, 65K05, 90C30

1. Introduction.

In this paper we consider the nonlinear programming problem:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & c(x) = 0, \\ & x \geq 0, \end{array} \tag{1.1}$$

where x is a vector of dimension n, and functions $f : \Re^n \to \Re$ and $c : \Re^n \to \Re^m$ are real valued and twice continuously differentiable.

In [4] an interior-point ℓ_2-penalty method is proposed for problem (1.1) that can be seen as a regularized Newton method taking advantage of the special properties of ℓ_2-merit functions. Under mild assumptions, this method enjoys strong global convergence properties; namely, it either converges to a Karush-Kuhn-Tucker (KKT) point or it identifies a stationary point of the infeasibility measure. In this paper we present a variant of this method in which a piecewise linear penalty function (PLPF) is used in the line search to accept trial points. This alternative has exhibited encouraging numerical performance compared to state-of-the-art interior-point codes in preliminary tests [6]. Therefore, its theoretical convergence and robustness deserves careful study. Here we show that the proposed approach has the same global convergence properties as those established in [4]. Moreover, by introducing second-order correction steps, we show that the approach also enjoys fast local convergence without incurring much additional cost.

1.1. Interior-point ℓ_2-penalty methods.

We first briefly review the method in [4]. By adding the barrier term $-\mu \sum_{i=1}^n \ln x_i$ to the objective function $f(x)$, we obtain the barrier subproblem

$$\begin{array}{ll} \min & \varphi_\mu(x) = f(x) - \mu \sum_{i=1}^n \ln x_i \\ \text{s.t.} & c(x) = 0, \\ & x > 0, \end{array} \tag{1.2}$$

*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.

where $\mu > 0$ is the barrier parameter. The first-order optimality conditions for problem (1.2) give rise to a nonlinear system of equations in $(x, \lambda, y) \in \Re^{2n+m}$

$$\mathcal{R}_\mu(x, \lambda, y) = \begin{bmatrix} \nabla_x \mathcal{L}(x, \lambda, y) \\ X\lambda - \mu e \\ c(x) \end{bmatrix} = 0, \tag{1.3}$$

where $X = \text{diag}(x)$, $\text{diag}(\cdot)$ denotes the diagonal matrix of a vector and

$$\mathcal{L}(x, u, v) = f(x) - x^\top \lambda + c(x)^\top y \tag{1.4}$$

is the Lagrangian function associated with problem (1.1).

The method in [4] starts with a point strictly satisfying the nonnegativity constraints, i.e., $x > 0$, and uses a perturbed Newton method with a line search strategy to find an approximate solution $(x(\mu), \lambda(\mu), y(\mu))$ of the nonlinear system (1.3). The merit function used in the line search is the ℓ_2-penalty function:

$$\Phi_{\mu,\gamma}(x) = \varphi_\mu(x) + \gamma \|c(x)\|, \tag{1.5}$$

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.

Suppose the current iterate is (x^k, λ^k, y^k) and the current penalty parameter is γ_k. Each step direction $(\triangle\bar{x}^k, \triangle\bar{\lambda}^k, \triangle\bar{y}^k)$ of the perturbed Newton method is obtained by solving a modified Newton system of (1.3):

$$\mathcal{M}_k \begin{bmatrix} \triangle\bar{x}^k \\ \triangle\bar{y}^k \end{bmatrix} = - \begin{bmatrix} \nabla_x \mathcal{L}(x^k, \mu X_k^{-1} e, y^k) \\ c(x^k) - \frac{\|c(x^k)\|}{\gamma_k} y^k \end{bmatrix}, \tag{1.6}$$

and

$$\triangle\bar{\lambda}^k = \mu X_k^{-1} e - X_k^{-1} \Lambda_k \triangle\bar{x}^k - \lambda^k, \tag{1.7}$$

where e is a vector of all ones, $X_k = \text{diag}(x^k)$, $\Lambda_k = \text{diag}(\lambda^k)$, $\mathcal{H}_k$ is the Hessian of the Lagrangian $\nabla^2_{xx}\mathcal{L}(x^k, \lambda^k, y^k)$,

$$\mathcal{M}_k = \begin{bmatrix} \mathcal{H}_k + X_k^{-1}\Lambda_k & \nabla c(x^k) \\ \nabla c(x^k)^\top & -\frac{\|c(x^k)\|}{\gamma_k} I \end{bmatrix}, \tag{1.8}$$

and I is the identity. To ensure (1.6) solvable, the inertia of $\mathcal{M}_k$ is checked, and if necessary a suitable multiple of the identity is added to $\mathcal{H}_k$ so that $\mathcal{M}_k$ has inertia $(n, m, 0)$ (i.e., n positive eigenvalues, m negative eigenvalues and no zero eigenvalues). The primal iterate is updated by setting $x^{k+1} = x^k + t\triangle\bar{x}^k$, where the step size $t \in (0, 1]$ is determined by a backtracking line search strategy that ensures that x^{k+1} is strictly positive and a sufficient reduction is made in the merit function $\Phi_{\mu,\gamma_k}(x)$. A fraction-to-the-boundary rule is employed to obtain the dual iterates λ^{k+1} so that they are positive. The equality constraint multipliers are updated using a unit step size, i.e., $y^{k+1} = y^k + \triangle\bar{y}^k$. The penalty parameter γ_k is updated according to a rule that guarantees proper convergence of the iterates.

1.2. The PLPF. PLPFs provide an alternative to traditional penalty methods, which rely on one single penalty parameter combining optimality and infeasibility, and was first used by Gomes [10] in the context of trust-region sequential quadratic programming methods for NLP. In terms of interior-point methods, for some fixed barrier parameter μ, the PLPF $\mathcal{P}^k_\mu(\rho)$ at iteration k is a function of a penalty parameter ρ and is defined by all previous iterates in the form

$$\mathcal{P}^k_\mu(\rho) = \min\{\varphi_\mu(x^i) + \rho\|c(x^i)\|_p,\ i = 0, 1, ..., k\}, \quad \forall \rho \geq 0, \tag{1.9}$$

where $\|\cdot\|_p$ denotes the p-norm for some $p \geq 1$. Although we use the PLPF in the context of an interior-point method, the PLPF has nothing to do with the fact that $\varphi_\mu(x)$ is a log barrier function and can be applied in non-interior-point contexts.

In both trust region methods and line search methods, PLPFs aim at providing a criterion for accepting trial points. In particular, a new point x is considered acceptable as the next iterate x^{k+1} if there is a $\bar{\rho} \geq 0$ such that $\varphi_\mu(x^{k+1}) + \bar{\rho}\|c(x^{k+1})\|_p$ is sufficiently below $\mathcal{P}_\mu^k(\bar{\rho})$. This implies that at least for some values of ρ, the function $\varphi_\mu(x) + \rho\|c(x)\|_p$ improves the current PLPF and hence we can update the PLPF for the next iteration by setting

$$\mathcal{P}_\mu^{k+1}(\rho) = \min\{\mathcal{P}_\mu^k(\rho), \varphi_\mu(x^{k+1}) + \rho\|c(x^{k+1})\|_p\}, \quad \forall \rho \geq 0. \tag{1.10}$$

Clearly, we can remove those indices that are redundant in defining $\mathcal{P}_\mu^k(\rho)$, and rewrite it as

$$\mathcal{P}_\mu^k(\rho) = \min\{\varphi_\mu^{k,i} + \rho\theta^{k,i},\ i = 1, ..., N_k\}, \quad \forall \rho \geq 0, \tag{1.11}$$

where N_k is the number of linear pieces of $\mathcal{P}_\mu^k(\rho)$, and for every $i \in \{1, ..., N_k\}$, $\varphi_\mu^{k,i}$ and $\theta^{k,i}$ are, respectively, the values of the barrier function $\varphi_\mu(x)$ and the infeasibility $\|c(x)\|_p$ at a corresponding iterate x^j, $0 \leq j \leq k$. Without loss of generality, we can order and relabel those indices so that

$$\varphi_\mu^{k,1} < \varphi_\mu^{k,2} < ... < \varphi_\mu^{k,N_k} \text{ and } \theta^{k,1} > \theta^{k,2} > ... > \theta^{k,N_k}. \tag{1.12}$$

Thus, the PLPF $\mathcal{P}_\mu^k(\rho)$ can be written in the form

$$\mathcal{P}_\mu^k(\rho) = \begin{cases} \varphi_\mu^{k,1} + \rho\theta^{k,1}, & \rho^{k,1} \leq \rho \leq \rho^{k,2}, \\ \vdots & \vdots \\ \varphi_\mu^{k,N_k-1} + \rho\theta^{k,N_k-1}, & \rho^{k,N_k-1} \leq \rho \leq \rho^{k,N_k}, \\ \varphi_\mu^{k,N_k} + \rho\theta^{k,N_k}, & \rho^{k,N_k} \leq \rho, \end{cases} \tag{1.13}$$

where $\rho^{k,i}$, $i = 1, ..., N_k$, are the break points of $\mathcal{P}_\mu^k(\rho)$. Clearly, we have

$$\rho^{k,1} = 0 \text{ and } \rho^{k,i} = \frac{\varphi_\mu^{k,i} - \varphi_\mu^{k,i-1}}{\theta^{k,i-1} - \theta^{k,i}},\ i = 2, ..., N_k. \tag{1.14}$$

Therefore, the PLPF $\mathcal{P}_\mu^k(\rho)$ and its break points are completely determined by the point set $\mathcal{A}_\mu^k = \{(\varphi_\mu^{k,i}, \theta^{k,i}), i = 1, ..., N_k\}$.

The PLPF method differs from traditional penalty methods in that no particular penalty parameter needs to be chosen prior to accepting trial points. This provides more flexibility, especially when it is difficult to choose a suitable penalty parameter, and in our preliminary numerical tests, it leads to better numerical performance than a direct penalty method [6]. In a certain sense, these two approaches function in a complementary manner. Given a penalty parameter, a traditional penalty method tries to find a point that provides a sufficient reduction in the merit function $\Phi_{\mu,\gamma}(x)$, while given a trial point the PLPF method tries to find a penalty parameter that yields a sufficient reduction in the PLPF $\mathcal{P}_\mu^k(\rho)$. In this paper we take advantage of the strength of both of these approaches by suitably combining them.

The PLPF method also has its root in the filter method of Fletcher and Leyffer [7], in which the filter for accepting trial points is also defined by historical data. In particular, a filter $\mathcal{F}_\mu^k$ is a set of iterates that are not dominated by any existing iterate in terms of optimality and feasibility, i.e.,

$$\mathcal{F}_\mu^k = \{(\varphi_\mu(x^i), \|c(x^i)\|_p), i = 0, 1, ..., k | \varphi_\mu(x^j) \geq \varphi_\mu(x^i) \text{ or } \theta^j \geq \theta^i,\ \forall j = 0, 1, ..., k\}.$$

A new point x is considered acceptable to the current filter $\mathcal{F}_\mu^k$ only if it decreases either the barrier function value or the infeasibility of at least one point in $\mathcal{F}_\mu^k$; i.e., x is not dominated by any point in $\mathcal{F}_\mu^k$. It then can be added to the filter, resulting in a new filter consisting of it and possibly a subset of the old filter points.

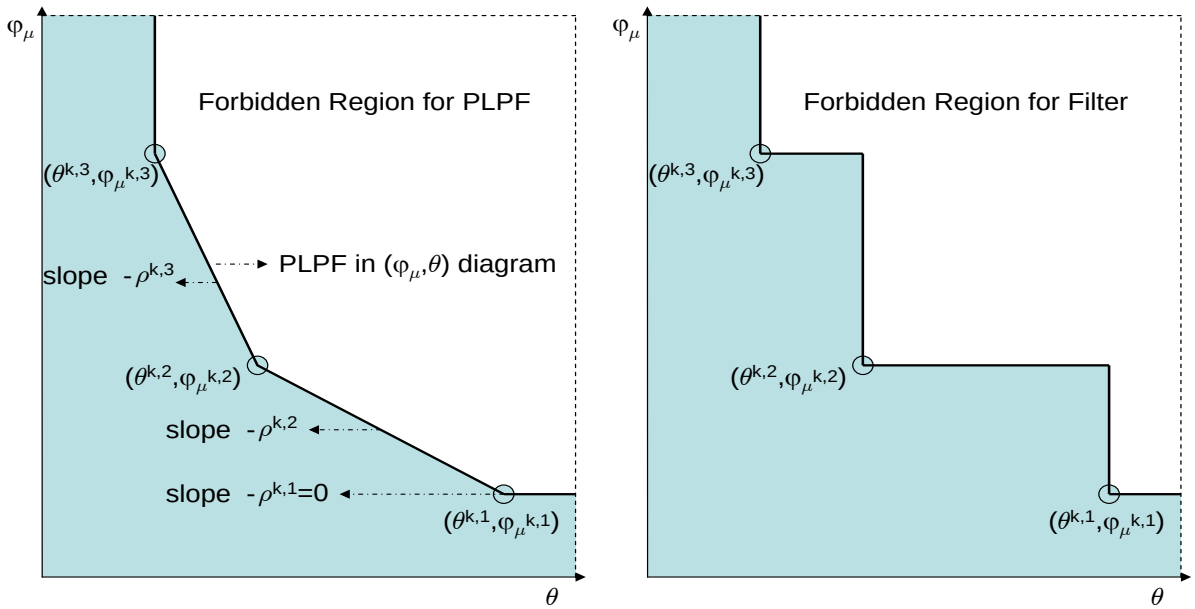

FIG. 1.1. *Comparison of Forbidden Regions for PLPF and Filter*

The criterion for accepting trial points imposed by the PLPF, which requires certain combination of the infeasibility and the optimality to be improved, is more restrictive than that by the filter method, which requires either the infeasibility or the optimality to be improved. This restrictiveness is inherent to traditional penalty methods that take care of both infeasibility and optimality simultaneously and actually leads to favorable performance in our numerical experience [4].

To see the relation between the two methods, first notice that every point in $\mathcal{A}_\mu^k$ is not dominated by any other points in $\mathcal{A}_\mu^k$. Moreover, if there is a $\bar{\rho} \geq 0$ such that the current trial point x provides a reduction in $\mathcal{P}_\mu^k(\bar{\rho})$, x is not dominated by any points in $\mathcal{A}_\mu^k$. Hence, x could also be acceptable using the filter criterion. However, the reverse is not true in general. Fig. 1.1 illustrates the forbidden regions (the unshaded areas) of points $(\varphi_\mu(x), \|c(x)\|_p)$ that must be rejected by PLPF methods and filter methods, respectively. To simplify comparison, we assume in Fig. 1.1 that the current set $\mathcal{A}_\mu^k$ is the same as the current filter $\mathcal{F}_\mu^k$ (both of them have three elements). Note that the forbidden region of the PLPF method is the smallest convex set that contains the forbidden region of the filter method.

1.3. Structure of the paper. This paper is organized as follows. In the next section, we present new algorithmic features of our method that distinguishes it from the methods in [4], including step direction computations, line search strategies, penalty parameter update strategies and second-order correction steps. We then describe a modified Newton method that solves the barrier problem (1.2) for a fixed barrier parameter μ and show that it is well defined. Its global convergence is established in Section 3. The overall algorithm for solving problem (1.1) and its global convergence is presented in Section 4. Fast local convergence including quadratic convergence for each fixed barrier parameter μ and overall superlinear convergence is proved in Section 5. In the last section, we give some concluding remarks.

2. Solving barrier problem (1.2).

2.1. Computing step directions. In addition to (1.6), our method solves at each iteration another perturbed linear system in $(\triangle\tilde{x}^k, \triangle\tilde{\lambda}^k, \triangle\tilde{y}^k)$:

$$\mathcal{M}_k \begin{bmatrix} \triangle\tilde{x}^k \\ \triangle\tilde{y}^k \end{bmatrix} = - \begin{bmatrix} \nabla_x \mathcal{L}(x^k, \mu X_k^{-1} e, y^k) \\ c(x^k) \end{bmatrix}, \tag{2.1}$$

and

$$\triangle\tilde{\lambda}^k = \mu X_k^{-1} e - X_k^{-1} \Lambda_k \triangle\tilde{x}^k - \lambda^k. \tag{2.2}$$

It will be shown that when the iterates are close to the solution of problem (1.2), (2.1) provides to a better approximation to the pure Newton system of (1.3) than (1.6). This plays a critical role in establishing our fast local convergence results. Solving (2.1) instead of the pure Newton system has the advantage of avoiding an extra matrix factorization as (1.6) and (2.1) have the same coefficient matrix. Moreover, our numerical experience indicates that if the Jacobian of the active constraints is singular of nearly singular, solving (2.1) often appears to be more stable than solving the pure Newton system due to the perturbed diagonal elements. This helps overcome numerical difficulties in some irregular problems. Unfortunately, however, the direction $\triangle\bar{x}^k$ is not necessary a descent direction for the merit function $\Phi_{\mu,\gamma_k}(x)$. Let

$$\bar{\mathcal{H}}_k = \begin{cases} \mathcal{H}_k + X_k^{-1}\Lambda_k, & \text{if } \|c(x^k)\| = 0, \\ \mathcal{H}_k + X_k^{-1}\Lambda_k + \frac{\gamma_k}{\|c(x^k)\|} \nabla c(x^k) \nabla c(x^k)^\top, & \text{if } \|c(x^k)\| > 0. \end{cases} \tag{2.3}$$

To guarantee global convergence, the following conditions are checked to determine if the solution of (2.1) is acceptable as a search direction,

$$\begin{cases} \text{(i)} & \left\|\triangle\tilde{x}^k - \triangle\bar{x}^k\right\| \le \left\|\triangle\tilde{x}^k\right\|^{\vartheta}, \\ \text{(ii)} & \|\triangle\tilde{y}^k\| \le \delta^{\max}, \\ \text{(iii)} & \Phi'_{\mu,\gamma_k}(x^k; \triangle\tilde{x}^k) \le -\zeta(\triangle\tilde{x}^k)^\top \bar{\mathcal{H}}^k \triangle\tilde{x}^k, \end{cases} \tag{2.4}$$

where $\vartheta, \zeta \in (0,1)$ and $\delta^{\max} > 0$. If condition (2.4) holds, we use $\triangle\tilde{x}^k$ as the search direction and set $(\triangle x^k, \triangle\lambda^k, \triangle y^k) = (\triangle\tilde{x}^k, \triangle\tilde{\lambda}^k, \triangle\tilde{y}^k)$; otherwise, we set $(\triangle x^k, \triangle\lambda^k, \triangle y^k) = (\triangle\bar{x}^k, \triangle\bar{\lambda}^k, \triangle\bar{y}^k)$.

Condition (i) of (2.4) requires that $\triangle\tilde{x}^k$ does not differ too much from $\triangle\bar{x}^k$ since along the latter direction, global convergence can be guaranteed. Condition (ii) of (2.4) ensures the boundedness of the multiplier step directions. Condition (iii) of (2.4) guarantees that $\triangle\tilde{x}^k$ is a descent direction for the merit function $\Phi_{\mu,\gamma_k}(x)$ as we require that for some positive parameter ν,

$$\begin{cases} d^\top \bar{\mathcal{H}}_k d \ge \nu\|d\|^2, \ \forall d \ne 0 \text{ such that } \nabla c(x^k)^\top d = 0, & \text{if } \|c(x^k)\| = 0, \\ d^\top \bar{\mathcal{H}}_k d \ge \nu\|d\|^2, \ \forall d \in \Re^n \backslash \{0\}, & \text{if } \|c(x^k)\| > 0. \end{cases} \tag{2.5}$$

Condition (2.5) can be achieved by modifying $\mathcal{H}_k$, e.g., by adding multiples of the identity to $\mathcal{H}_k$. Note that there exists some $\nu > 0$ such that (2.5) holds with the exact Hessian, i.e., $\mathcal{H}_k = \nabla^2_{xx}\mathcal{L}(x^k, \lambda^k, y^k)$, in a neighborhood of a local minimizer of problem (1.2) that satisfies the second-order sufficiency conditions. We also note that (2.5) guarantees the solvability of the linear systems (1.6) and (2.1) provided the Mangasarian-Fromovitz constraint qualification (MFCQ) holds at x^k, see [4].

2.2. Computing step sizes. To compute a step size α_x^k for x^k, we first compute an upper bound $\bar{\alpha}_x^k$ for α_x^k that ensures that all trial points stay sufficiently away from the boundary,

$$\bar{\alpha}_x^k = \max\{\alpha \in (0,1] | x^k + \alpha\triangle x^k \ge (1-\tau_\mu)x^k\}, \tag{2.6}$$

where $\tau_\mu \in (0,1)$.

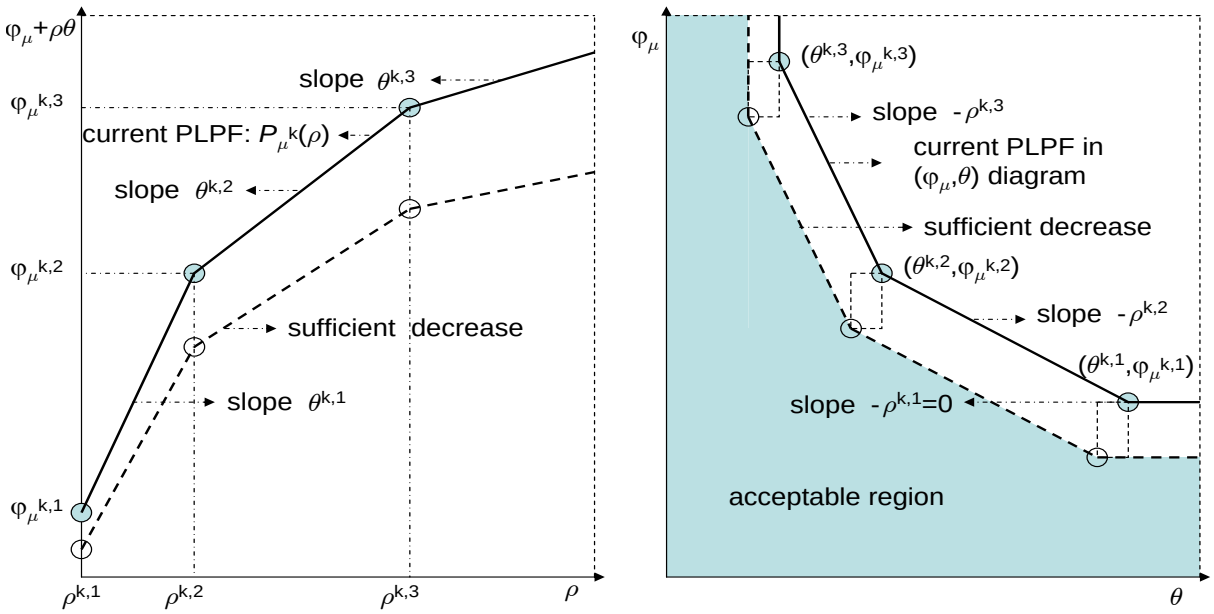

FIG. 2.1. *Sufficient decrease conditions for the PLPF*

Now given a step size $\alpha_x^k \in (0, \bar{\alpha}_x^k]$, we use two criteria to determine if the trial point $x^k + \alpha_x^k \triangle x^k$ is acceptable as the next iterate. Let $\mathcal{P}_\mu^k(\rho)$ be the current PLPF and $\mathcal{A}_\mu^k$ be the point set for defining the break points of $\mathcal{P}_\mu^k(\rho)$ (see Section 1.2), which starts from $\{(\varphi_\mu(x^0), \|c(x^0)\|_p)\}$ at the first iteration. We first check whether this trial point provides a sufficient reduction in the PLPF. In particular, we check if there is a $\rho^{k,i}$, $i = 1, ..., N_k$, such that the following condition holds:

$$\begin{aligned} &\varphi_\mu(x^k + \alpha_x^k \triangle x^k) + \rho^{k,i}\|c(x^k + \alpha_x^k \triangle x^k)\|_p - \mathcal{P}_\mu^k(\rho^{k,i}) \\ &\leq -\sigma \max\{\alpha_x^k, \alpha_x^{\min}\}(\omega_k + \rho^{k,i}\|c(x^k)\|_p), \end{aligned} \tag{2.7}$$

where

$$\omega_k = \max\{(\triangle x^k)^\top (\mathcal{H}_k + X_k^{-1}\Lambda_k)\triangle x^k, \kappa\|\triangle x^k\|^2\}, \tag{2.8}$$

$\alpha_x^{\min} \in (0, 1)$, $\sigma \in (0, \frac{1}{2})$ and $\kappa > 0$. If such a $\rho^{k,i}$ is found, we accept $x^k + \alpha_x^k \triangle x^k$. Otherwise, we continue to check

$$\|c(x^k + \alpha_x^k \triangle x^k)\|_p - \theta^{k,N_k} \leq -\sigma \max\{\alpha_x^{\min}, \alpha_x^k\}\|c(x^k)\|_p. \tag{2.9}$$

If (2.9) holds, we accept $x^k + \alpha_x^k \triangle x^k$. The sufficient decrease conditions (2.7) and (2.9) are illustrated in Fig. 2.1, where we assume that the current PLPF has three break points. In the first illustration of Fig. 2.1, the dashed piecewise linear curve corresponds to the largest value of the trial penalty function allowed by the PLPF criteria as a function of ρ; i.e., for any trial point x, if the semi-infinite line $\{\varphi_\mu(x) + \rho\|c(x)\|_p, \rho \geq 0\}$ intersects with this dashed curve, x is acceptable by the PLPF criteria. The second illustration of Fig. 2.1 provides an alternative view in terms of a (φ_μ, θ) diagram. Here, for any trial point x, if the point $(\varphi_\mu(x), \|c(x)\|_p)$ lies in the shaded area, x is acceptable by the PLPF criteria.

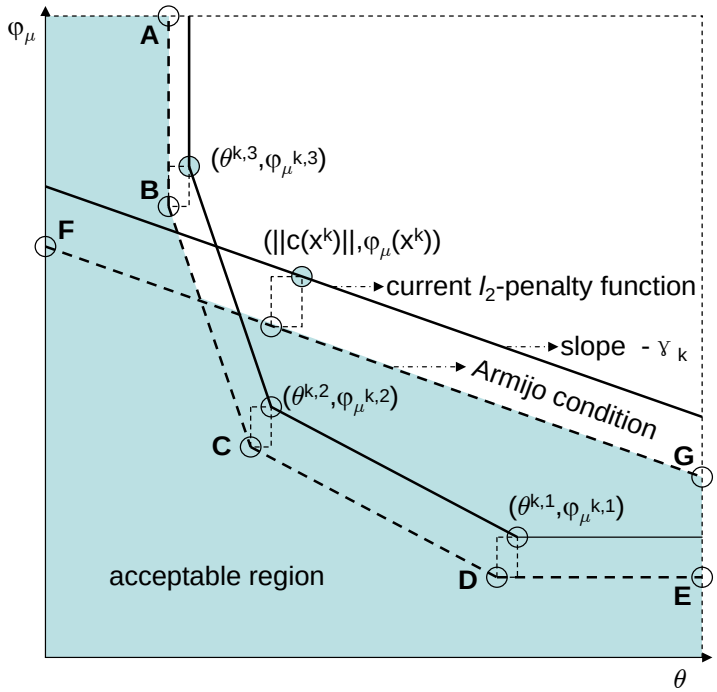

FIG. 2.2. *Acceptable region by combining the PLPF and the ℓ_2-penalty function*

If both (2.7) and (2.9) fail to hold, we conclude that the current trial point does not provide a sufficient reduction in $\mathcal{P}_\mu^k(\rho)$ for any value of $\rho \geq 0$. In this case, unlike filter methods or the PLPF method [10] that either try a new point or switch to a restoration algorithm that focus solely on decreasing infeasibility, our method continues to check an Armijo condition with respect to the ℓ_2-penalty function $\Phi_{\mu,\gamma_k}(x)$,

$$\Phi_{\mu,\gamma_k}(x^k + \alpha_x^k \triangle x^k) - \Phi_{\mu,\gamma_k}(x^k) \leq \sigma \alpha_x^k \Phi'_{\mu,\gamma_k}(x^k; \triangle x^k). \tag{2.10}$$

If (2.10) holds, we accept $x^k + \alpha_x^k \triangle x^k$. Otherwise, we reduce α_x^k by a factor $\beta \in (0,1)$, i.e., set $\alpha_x^k \leftarrow \beta \alpha_x^k$, and repeat the above process until a trial point is accepted. We will show that the ℓ_2-penalty technique [4] guarantees this process always terminates successfully. Given a trial step size, the set of points that are acceptable by our line search criteria is illustrated in Fig. 2.2, where we assume that the PLPF uses the Euclidean norm, i.e., $p = 2$. Any point that lies below the line segment **F-G** is acceptable by the Armijo condition (2.10) and any point that lies to the left and below the piecewise linear curve **A-B-C-D-E** is acceptable by the PLPF condition (2.7) or (2.9). The shaded area contains points that are acceptable by our line search criteria; i.e., by either the Armijo condition or the PLPF condition.

A problem of using the PLPF in a line search algorithm is that the trial points could always decrease the constraint violation but not the barrier function. Consequently, the iterates may converge to a feasible point but not an optimal one. As a remedy, in our algorithm we will not use the PLPF to accept trial points if the current constraint violation is small and if the trial barrier function does not provide a sufficient reduction for any $\varphi_\mu^{k,i}$ $(i = 1, ..., N_k)$. In particular, we check

$$\|c(x^k)\|_p \leq \theta^{\min} \text{ and } \varphi_\mu(x^k + \alpha_x^k \triangle x^k) > \varphi_\mu^{k,N_k} - \sigma \alpha_x^{\min} \omega_k, \tag{2.11}$$

where $\theta^{\min}$ is a small positive parameter. If (2.11) holds, we only employ the ℓ_2-penalty function $\Phi_{\mu,\gamma_k}(\cdot)$ and the Armijo condition (2.10) to accept trial points.

To update the dual iterate λ^k, we use a fraction-to-the-boundary rule to ensure the positivity of λ^{k+1}. Specifically, given x^{k+1}, we first compute a step size α_λ^k

$$\alpha_\lambda^k = \max\left\{\alpha \in (0,1] \,\middle|\, \lambda^k + \alpha\triangle\lambda^k \geq \min\left\{(1-\tau_\mu)\lambda^k, \frac{\mu}{\varrho}X_{k+1}^{-1}e\right\}\right\}, \tag{2.12}$$

where $X_{k+1} = \mathrm{diag}(x^{k+1})$ and $\varrho > 1$. We then compute the next dual iterate λ^{k+1} by

$$\lambda^{k+1} = \min\left\{\lambda^k + \alpha_\lambda^k\triangle\lambda^k, \varrho\mu X_{k+1}^{-1}e\right\}. \tag{2.13}$$

In (2.12) and (2.13), we use the terms $\frac{\mu}{\varrho}X_{k+1}^{-1}e$ and $\varrho\mu X_{k+1}^{-1}e$ to prevent λ^{k+1} from deviating too far from $\mu X_{k+1}^{-1}e$, which is what it must equal in an optimal solution.

The step size α_y^k for y^k is always set to one in the original paper [4] and is chosen using heuristics in the implementation paper [6]. Here, to prove our convergence results, we only require α_y^k to be set to one whenever a solution is approached. To simplify our analysis, we choose $\alpha_y^k = 1$ if $\|\triangle x^k\| \leq \delta^{\min}$ for some threshold $\delta^{\min} > 0$; otherwise, we choose $\alpha_y^k \in (0,1]$.

2.3. Second-order corrections. Penalty methods can suffer from the Maratos effect. Here, a full modified Newton step may increase both the barrier function and the constraint violation even when the iterates are very close to a solution of (1.2). In this case, both the PLPF condition and the Armijo condition in our line search procedure reject this full step and only accept a small fraction of the step, resulting in poor local behavior. As a remedy, second-order corrections aim at improving feasibility by applying an additional Newton-type step for the constraints at the point $x^k + \triangle x^k$. In our algorithm, a second-order correction step is given by

$$\mathcal{M}_k \begin{bmatrix} \triangle\hat{x}^k \\ \triangle\hat{y}^k \end{bmatrix} = -\begin{bmatrix} \nabla_x\mathcal{L}(x^k, \mu X_k^{-1}e, y^k) \\ c(x^k + \bar{\alpha}_x^k\triangle x^k) - \bar{\alpha}_x^k\nabla c(x^k)^\top\triangle x^k \end{bmatrix}. \tag{2.14}$$

Here, we use the step size $\bar{\alpha}_x^k$ instead of the often used unit step size to avoid additional evaluation of the constraint functions. This will not interfere with our fast local convergence results as $\bar{\alpha}_x^k$ eventually becomes one (see Section 5). Moreover, it will be shown that for a full second-order correction step $\triangle\hat{x}^k$ we have $c(x^k + \triangle\hat{x}^k) = o(\|c(x^k)\|)$ if x^k is close to a solution of (1.2). Therefore, the point $x^k + \triangle\hat{x}^k$ is more likely to be acceptable than the original trial point.

In our line search, second-order corrections are used if the first trial point $x^k + \bar{\alpha}_x^k\triangle x^k$ is rejected. In this case, we apply the fraction-to-the-boundary rule

$$\hat{\alpha}_x^k = \max\{\alpha \in (0,1] | x^k + \alpha\triangle\hat{x}^k \geq (1-\tau)x^k\} \tag{2.15}$$

and check if the trial point $x^k + \hat{\alpha}_x^k\triangle\hat{x}^k$ provides a sufficient reduction in the current PLPF or the current ℓ_2-penalty function using the procedure described in Section 2.2. In particular, this is done by checking the following conditions analogous to (2.7), (2.9) and (2.10), respectively,

$$\begin{aligned} &\varphi_\mu(x^k + \hat{\alpha}_x^k\triangle\hat{x}^k) + \rho^{k,i}\|c(x^k + \hat{\alpha}_x^k\triangle\hat{x}^k)\|_p - \mathcal{P}_\mu^k(\rho^{k,i}) \\ &\leq -\sigma\max\{\hat{\alpha}_x^k, \alpha_x^{\min}\}(\omega_k + \rho^{k,i}\|c(x^k)\|_p), \end{aligned} \tag{2.16}$$

$$\|c(x^k + \hat{\alpha}_x^k\triangle\hat{x}^k)\|_p - \theta^{k,N_k} \leq -\sigma\max\{\hat{\alpha}_x^k, \alpha_x^{\min}\}\|c(x^k)\|_p, \tag{2.17}$$

$$\Phi_{\mu,\gamma_k}(x^k + \hat{\alpha}_x^k\triangle\hat{x}^k) - \Phi_{\mu,\gamma_k}(x^k) \leq \sigma\hat{\alpha}_x^k\Phi'_{\mu,\gamma_k}(x^k; \triangle x^k). \tag{2.18}$$

Note that the original direction $\triangle x^k$, which defines ω_k in (2.8), is still used in the right hand sides of (2.16) and (2.18). Similarly to (2.11), if the following condition holds,

$$\|c(x^k)\|_p \leq \theta^{\min} \text{ and } \varphi_\mu(x^k + \hat{\alpha}_x^k\triangle\hat{x}^k) > \varphi_\mu^{k,N_k} - \sigma\alpha_x^{\min}\omega_k. \tag{2.19}$$

we do not use the PLPF criterion to accept the trial point $x^k + \hat{\alpha}_x^k \triangle \hat{x}^k$.

There are a number of ways to choose the quantities in the second block of the right hand side of (2.14) that, in the limit, have the same order of magnitude as those that we use here. It is also possible to perform second-order corrections several times, e.g., by replacing the direction $\triangle x^k$ in (2.14) by a previous second-order correction step $\triangle \hat{x}^k$. Here, for simplicity, we assume that our method performs at most one second-order correction step per iteration. This suffices for our convergence proof.

2.4. Updating the PLPF. In our algorithm, whenever a trial point $\bar{x}$ is accepted as the next iterate by the PLPF conditions (2.7) or (2.9), we update the PLPF for the next iteration; otherwise, we do not change it. To update the PLPF, there are several cases to analyze. The analysis is straightforward using the fact that the PLPF is a concave and monotone increasing function.

Case 1. The function $\varphi_\mu(\bar{x}) + \rho\|c(\bar{x})\|_p$ has two intersection points ρ_1 and ρ_2 ($0 \le \rho_1 < \rho_2$) with the current PLPF $\mathcal{P}_\mu^k(\rho)$. Then $N_k \ge 2$. Define $\rho^{k,0} = -\infty$ and ρ^{k,N_k+1}. Suppose for some indices l_1 and l_2 with $1 \le l_1 < l_2 \le N_k$, we have

$$-\infty = \rho^{k,0} < ... < \rho^{k,l_1} < \rho_1 \le \rho^{k,l_1+1} < ...\rho^{k,l_2} \le \rho_2 < \rho^{k,l_2+1} ... < \rho^{k,N_k+1} = \infty.$$

From (1.11) defining $\mathcal{P}_\mu^k(\rho)$, the set $\mathcal{A}_\mu^{k+1}$ is then given by

$$\mathcal{A}_\mu^{k+1} = \{(\varphi_\mu^{k,1}, \theta^{k,1}), ..., (\varphi_\mu^{k,l_1}, \theta^{k,l_1}), (\varphi_\mu(\bar{x}), \|c(\bar{x})\|_p), (\varphi_\mu^{k,l_2}, \theta^{k,l_2}), ..., (\varphi_\mu^{k,N_k}, \theta^{k,N_k})\}.$$

Case 2. The function $\varphi_\mu(\bar{x}) + \rho\|c(\bar{x})\|_p$ intersects $\mathcal{P}_\mu^k(\rho)$ at one point $\rho > 0$. Then $\varphi_\mu(\bar{x}) \ne \varphi_\mu^{k,1}$. If $\varphi_\mu(\bar{x}) < \varphi_\mu^{k,1}$ and for some index l ($1 \le l \le N_k$),

$$\rho^{k,1} < ... < \rho^{k,l} \le \rho < \rho^{k,l+1} < ... < \rho^{k,N_k},$$

we have

$$\mathcal{A}_\mu^{k+1} = \{(\varphi_\mu(\bar{x}), \|c(\bar{x})\|_p), (\varphi_\mu^{k,l}, \theta^{k,l}), ..., (\varphi_\mu^{k,N_k}, \theta^{k,N_k})\}.$$

If $\varphi_\mu(\bar{x}) > \varphi_\mu^{k,1}$ and

$$\rho^{k,1} < ... < \rho^{k,l} < \rho \le \rho^{k,l+1} < ... < \rho^{k,N_k},$$

we have

$$\mathcal{A}_\mu^{k+1} = \{(\varphi_\mu^{k,1}, \theta^{k,1}), ..., (\varphi_\mu^{k,l}, \theta^{k,l}), (\varphi_\mu(\bar{x}), \|c(\bar{x})\|_p)\}.$$

Case 3. The function $\varphi_\mu(\bar{x})+\rho\|c(\bar{x})\|_p$ intersects $\mathcal{P}_\mu^k(\rho)$ only at $\rho = 0$. Then we have either $\|c(\bar{x})\|_p > \theta^{k,1}$ or $\|c(\bar{x})\|_p < \theta^{k,N_k}$. Otherwise, there are intersection points other than zero. If $\|c(\bar{x})\|_p < \theta^{k,N_k}$, we have $\mathcal{A}_\mu^{k+1} = \{(\varphi_\mu(\bar{x}), \|c(\bar{x})\|_p)\}$. If $\|c(\bar{x})\|_p > \theta^{k,1}$, it follows that $\varphi_\mu(\bar{x}) + \rho\|c(\bar{x})\|_p \ge \mathcal{P}_\mu^k(\rho)$ for all $\rho \ge 0$, implying $\bar{x}$ does not provide a sufficient reduction in $\mathcal{P}_\mu^k(\rho)$. This could happen if $\bar{x}$ is accepted by satisfying (2.7) for $i = 1$ and if $\triangle x^k = 0$.[1] In this case, we set $\mathcal{A}_\mu^{k+1} = \mathcal{A}_\mu^k$.

Case 4. The function $\varphi_\mu(\bar{x}) + \rho\|c(\bar{x})\|_p$ has no intersection point with $\mathcal{P}_\mu^k(\rho)$. Then $\varphi_\mu(\bar{x}) + \rho\|c(\bar{x})\|_p < \mathcal{P}_\mu^k(\rho)$ for all $\rho \ge 0$. Otherwise, $\bar{x}$ could not have been accepted. Therefore, $\mathcal{A}_\mu^{k+1} = \{(\varphi_\mu(\bar{x}), \|c(\bar{x})\|_p)\}$.

Case 5. The function $\varphi_\mu(\bar{x}) + \rho\|c(\bar{x})\|_p$ coincides with $\mathcal{P}_\mu^k(\rho)$ on one of its linear pieces. Then there is a l ($1 \le l \le N_k$) such that $\varphi_\mu^{k,l} = \varphi_\mu(\bar{x})$ and $\theta^{k,l} = \|c(\bar{x})\|_p$, and thus, $\varphi_\mu(\bar{x}) + \rho\|c(\bar{x})\|_p \ge \mathcal{P}_\mu^k(\rho)$ for all $\rho \ge 0$. This could happen if $\triangle x^k = 0$ or $\|c(x^k)\|_p = 0$ according to (2.7) and (2.9). In this case, we set $\mathcal{A}_\mu^{k+1} = \mathcal{A}_\mu^k$.

Having obtained $\mathcal{A}_\mu^{k+1}$, we can compute the corresponding break points ρ^{k+1} by (1.14) and define $\mathcal{P}_\mu^{k+1}$ by (1.13) with the index k replaced by $k+1$.

[1] Note that even if $\triangle x^k = 0$, x^k may not be a KKT point of problem (1.2) and hence the algorithm does not terminate. In this case, the first two conditions in (1.3) are satisfied by $(x^k, \lambda^k + \triangle\lambda^k, y^k + \triangle y^k)$, while x^k is not feasible to (1.2). If in addition $\|c(x^k)\| = 0$, x^k is a KKT point of (1.2).

2.5. Updating γ_k. To update the ℓ_2-penalty parameter γ_k, we use a slightly different rule from that used in [4]. In particular, we check the following conditions at each iteration,

$$\left\{\begin{array}{ll} \text{(i)} & \|\triangle\bar{x}^k\| \le \pi_k, \\ \text{(ii)} & \kappa_1\mu e \le X_k(\lambda^k + \triangle\bar{\lambda}^k) \le \kappa_2\mu e, \\ \text{(iii)} & \gamma_k < \eta\|y^k + \triangle\bar{y}^k\|, \end{array}\right. \tag{2.20}$$

where $\pi_k > 0$, $0 < \kappa_1 < 1 < \kappa_2$ and $\eta > 1$. If all conditions (i)-(iii) hold, we increase the ℓ_2-penalty parameter. The motivation behind these conditions is similar to that explained in [4]. Conditions (i) and (ii) imply that the iterates are probably converging to a local minimizer of the merit function $\Phi_{\mu,\gamma_k}(x)$, while condition (iii) indicates that the current penalty parameter is not large enough so that this local minimizer may be infeasible to problem (1.2). To see this, we will show that violation of condition (iii) eventually guarantees the feasibility of the limit point of the iterates.

2.6. Inner algorithm. In this section we present our regularized Newton algorithm for solving problem (1.2) for a fixed barrier parameter μ and show that it is well defined. Our algorithm starts from an interior iterate $(x^0, \lambda^0) > 0$ and maintains the strict positivity of (x^k, λ^k) for every k. Besides the cost for evaluating the problem functions and their derivatives, the main computational effort comes from modifying $\mathcal{H}_k$, if necessary, and factorizing the matrix $\mathcal{M}_k$ at each iteration.

Algorithm 2.1. Inner Algorithm for Solving Problem (1.2).
Step 0. Initialization.
Parameters: $\mu > 0$, $\epsilon_\mu > 0$, $\tau_\mu \in (0,1)$, $\vartheta \in (0,1)$, $\eta > 1$, $\sigma \in (0, \frac{1}{2})$,
$\kappa > 0$, $\beta \in (0,1)$, $\alpha_x^{\min} \in (0,1)$, $\theta^{\min} > 0$, $\varrho > 1$, $\delta^{\min} > 0$, $\delta^{\max}$,
$0 < \kappa_1 < 1 < \kappa_2$, $\zeta \in \left(0, 1 - \frac{1}{\eta}\right)$, $\chi > 1$.
Data: $(x^0, \lambda^0, y^0) \in \Re^{2n+m}$ with $x^0 > 0$ and $\lambda^0 > 0$, $\gamma_0 > 0$, $\pi_0 > 0$,
$\mathcal{A}_\mu^0 \leftarrow \{(\varphi_\mu(x^0), \|c(x^0)\|_p)\}$, $N_0 \leftarrow 1$, $\rho^{0,1} \leftarrow 0$.
Set $k \leftarrow 0$.
Step 1. Check convergence.
If $\|\mathcal{R}_\mu(x^k, \lambda^k, y^k)\| \le \epsilon_\mu$, STOP with SUCCESS.
Step 2. Compute search directions.
Step 2.1. Set $\mathcal{H}_k \leftarrow \nabla^2_{xx}\mathcal{L}(x^k, \lambda^k, y^k)$. Modify $\mathcal{H}_k$, if necessary, so that condition (2.5) holds.
Step 2.2. If $\mathcal{M}_k$ is singular, STOP with MFCQ FAILURE.
Step 2.3. Compute $(\triangle\bar{x}^k, \triangle\bar{\lambda}^k, \triangle\bar{y}^k)$, the solution of (1.6) and (1.7).
Step 2.4. Compute $(\triangle\tilde{x}^k, \triangle\tilde{\lambda}^k, \triangle\tilde{y}^k)$, the solution of (2.1) and (2.2).
Step 2.5. If conditions (2.4) hold, set $(\triangle x^k, \triangle\lambda^k, \triangle y^k) \leftarrow (\triangle\tilde{x}^k, \triangle\tilde{\lambda}^k, \triangle\tilde{y}^k)$;
otherwise, set $(\triangle x^k, \triangle\lambda^k, \triangle y^k) \leftarrow (\triangle\bar{x}^k, \triangle\bar{\lambda}^k, \triangle\bar{y}^k)$.
Step 2.6. If $x^k + \triangle\tilde{x}^k > 0$, $\lambda^k + \triangle\tilde{\lambda}^k > 0$ and $\|\mathcal{R}_\mu(x^k + \triangle\tilde{x}^k, \lambda^k + \triangle\tilde{\lambda}^k, y^k + \triangle\tilde{y}^k)\| \le \epsilon_\mu$,
set $(x^{k+1}, \lambda^{k+1}, y^{k+1}) \leftarrow (x^k + \triangle\tilde{x}^k, \lambda^k + \triangle\tilde{\lambda}^k, y^k + \triangle\tilde{y}^k)$, $k \leftarrow k+1$ and go to Step 1.
Step 3. Line Search.
Step 3.1. Compute $\bar{\alpha}_x^k$ by (2.6). Set $\alpha_x^k \leftarrow \bar{\alpha}_x^k$ and $p_k \leftarrow 0$.
Step 3.2. Check the PLPF conditions.
If (2.11) holds, go to Step 3.3; otherwise, continue.
If (2.7) holds for some $i = 1, ..., N_k$ or (2.9) holds, set $p_k \leftarrow 1$,
$x^{k+1} \leftarrow x^k + \alpha_x^k \triangle x^k$ and go to Step 4; otherwise, continue.
Step 3.3. Check the Armijo condition.
If (2.10) holds, set $x^{k+1} \leftarrow x^k + \alpha_x^k \triangle x^k$ and go to Step 4; otherwise, continue.
Step 3.4. Second-order corrections.
Step 3.4.1. If $\alpha_x^k \ne \bar{\alpha}_x^k$ and $\triangle x^k = \triangle\bar{x}^k$, go to Step 3.5; otherwise, continue.

Step 3.4.2. Compute $(\triangle\hat{x}^k, \triangle\hat{y}^k)$ from (2.14) and $\hat{\alpha}_x^k$ from (2.15).

Step 3.4.3. If (2.19) holds, go to Step 3.4.4; otherwise, continue.

If (2.16) holds for some $i = 1, ..., N_k$ or (2.17) holds, set $p_k \leftarrow 1$, $x^{k+1} \leftarrow x^k + \hat{\alpha}_x^k \triangle\hat{x}^k$ and go to Step 4; otherwise, continue.

Step 3.4.4. If (2.18) holds, set $x^{k+1} \leftarrow x^k + \hat{\alpha}_x^k \triangle\hat{x}^k$ and go to Step 4; otherwise, continue.

Step 3.5. Set $\alpha_x^k \leftarrow \beta\alpha_x^k$ and go to Step 3.2.

Step 4. Update.

Step 4.1. Update the iterate.

Compute α_λ^k from (2.12) and λ^{k+1} from (2.13).

If $\|\triangle x^k\| \le \delta^{\min}$, set $\alpha_y^k \leftarrow 1$; otherwise, choose a $\alpha_y^k \in (0, 1]$.

Set $y^{k+1} \leftarrow y^k + \alpha_y^k \triangle y^k$.

Step 4.2. Update the PLPF.

If $p_k = 0$, set $\mathcal{A}_\mu^{k+1} \leftarrow \mathcal{A}_\mu^k$ and $\rho^{k+1} \leftarrow \rho^k$; otherwise, obtain $\mathcal{A}_\mu^{k+1}$ according to the cases analyzed in Section 2.4 and compute ρ^{k+1} by (1.14).

Step 4.3. Update the ℓ_2-penalty parameter.

If (2.20) holds, choose a γ_{k+1} such that $\gamma_{k+1} \ge \chi\gamma_k$ and set $\pi_{k+1} \leftarrow \frac{\pi_k}{\chi}$; otherwise, set $\gamma_{k+1} \leftarrow \gamma_k$ and $\pi_{k+1} \leftarrow \pi_k$.

Step 4.4. Set $k \leftarrow k + 1$ and go to Step 1.

Remark 2.1. Algorithm 2.1 terminates at Step 2.2 if $\mathcal{M}_k$ is singular. Since condition (2.5) holds, we must have that $\|c(x^k)\| = 0$ and that the equality constraint gradients $\{\nabla c_i(x^k), i = 1, ..., m\}$ are linearly dependent according to Lemma 3.1 in [4]. This implies that x^k is a Fritz-John point of problem (1.1) that does not satisfy the MFCQ. If Algorithm 2.1 does not terminate at Step 2.2, the step directions are readily available from (1.6) or (2.1).

Remark 2.2. It follows from Lemma 3.2 in [4] that $\triangle\bar{x}^k$ is a descent direction of the merit function $\Phi_{\mu,\gamma_k}(x)$ provided condition (2.5) holds. In particular, we have

$$\Phi'_{\mu,\gamma_k}(x^k; \triangle\bar{x}^k) = -(\triangle\bar{x}^k)^\top \bar{\mathcal{H}}_k \triangle\bar{x}^k. \tag{2.21}$$

This together with condition (iii) in (2.4) give that the step direction $\triangle x^k$ defined by Step 2.5 of Algorithm 2.1 is a descent direction of $\Phi_{\mu,\gamma_k}(x)$. Since f and c are twice continuously differentiable, it is well known there is a small enough $\alpha \in (0, 1]$ such that for all $\alpha_x^k \in (0, \alpha]$, the Armijo condition (2.10) holds. Therefore, our line search procedure either finds a point that improves the current PLPF or eventually identifies a point that decreases the ℓ_2-penalty function. In Step 3 of Algorithm 2.1, we set the parameter p_k to 1 or 0 to indicate that the next iterate is accepted by the PLPF conditions (Step 3.2 and Step 3.4.3) or the Armijo conditions (Step 3.3 and Step 3.4.4), respectively. In the former case, we update the PLPF for the next iteration in Step 4.2.

Remark 2.3. In Step 4.1 of Algorithm 2.1, for purpose of allowing more algorithmic options, we do not specify how the step size α_y^k is chosen when $\|\triangle x^k\| > \delta^{\min}$. Also for this purpose, we do not give an explicit formula for increasing γ_k in Step 4.3 when (2.20) holds. Instead, we simply require γ_k to be sufficiently increased by imposing the condition $\gamma_{k+1} \ge \chi\gamma_k$.

By Remarks 2.1 and 2.2, at any iteration k Algorithm 2.1 either terminates at Step 1 with an approximate solution of problem (1.2) or at Step 2.2 with a Fritz-John point of problem (1.1) or readily computes the step directions in Step 2 and generates the next iterate in Steps 3 and 4. Moreover, a starting point satisfying $(x^0, \lambda^0) > 0$ is trivially available. Therefore we have proved:

PROPOSITION 2.1. *Algorithm 2.1 is well defined.*

3. Global convergence of Algorithm 2.1. Our proof of the global convergence of Algorithm 2.1 is closely related to our proof of the global convergence of the ℓ_2-penalty method in [4]. However, the later proof does not carry over to Algorithm 2.1 because of the way that it combines the use of the PLPF with that of the ℓ_2-penalty. Hence, the convergence proof given below contains some new features. The following assumptions used in [4] are also needed here.

Assumption A.

A1. *The primal iterate sequence $\{x^k\}$ lies in a bounded set.*

A2. *The modified Hessian sequence $\mathcal{H}_k$ is bounded.*

Let us assume for the present that Algorithm 2.1 generates an infinite sequence of iterates, i.e., it neither stops at Step 1 even if the termination criterion there is met, nor stops at Step 2.2 with a point failing to satisfy the MFCQ.

Using a different rule for updating the ℓ_2-penalty parameter, it is proved in [4] that under Assumption A, if γ_k tends to infinity, there is an accumulation point of the iterates that is either a Fritz-John point of problem (1.1) or a Fritz-John point (also a KKT point in this case) of the feasibility problem: $\min_{x\geq 0}\ \|c(x)\|^2$. The next results shows that this property holds as well for the new rule (2.20).

THEOREM 3.1. *Suppose Assumption A holds. If γ_k is increased infinitely many times, then there exists a limit point of $\{x^k\}$ generated by Algorithm 2.1 that is either a Fritz-John point of problem (1.1) at which the MFCQ fails to hold or a KKT point of the problem* $\min_{x\geq 0}\ \|c(x)\|^2$.

Proof. Since γ_k is increased infinitely many times, it follows from Step 4.3 of Algorithm 2.1 that there exists an infinite index set $\mathcal{K}$ such that $\gamma_{k+1} \geq \chi\gamma_k$ and $\pi_{k+1} = \frac{\pi_k}{\chi}$ for all $k \in \mathcal{K}$. This implies that $\{\gamma_k\} \to \infty$ and $\{\pi_k\} \to 0$ as $\chi > 1$. The conditions that trigger the increase of γ_k must be satisfied. Hence, we know from (2.20) that $\{\|\triangle\bar{x}^k\|\}_{\mathcal{K}} \to 0$ and $\{\|(\lambda^k + \triangle\bar{\lambda}^k, y^k + \triangle\bar{y}^k)\|\}_{\mathcal{K}} \to \infty$. By assumption A1, there exists an infinite set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{x^k\}_{\bar{\mathcal{K}}} \to \bar{x}$ and

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

with $\|(\bar{\lambda}, \bar{y})\| = 1$ and $\bar{\lambda} \geq 0$ by condition (ii) of (2.20). Since $x^k > 0$ for all k, $\bar{x} \geq 0$. There are two cases.

Case 1. $\|c(\bar{x})\| = 0$. From the first equation in (1.6) and (1.7), we have

$$\mathcal{H}_k\triangle\bar{x}^k - (\lambda^k + \triangle\bar{\lambda}^k) + \nabla c(x^k)(y^k + \triangle\bar{y}^k) + \nabla f(x^k) = 0. \tag{3.1}$$

Dividing (3.1) and condition (ii) of (2.20) by $\|(\lambda^k + \triangle\bar{\lambda}^k, y^k + \triangle\bar{y}^k)\|$ and letting $k \in \bar{\mathcal{K}} \to \infty$ yields that $\nabla c(\bar{x})\bar{y} = \bar{\lambda}$ and $\bar{\lambda}^\top\bar{x} = 0$, respectively. Since $\bar{x} \geq 0$, $\bar{\lambda} \geq 0$ and $(\bar{\lambda}, \bar{y}) \neq 0$, it follows that $\bar{x}$ is a Fritz-John point of problem (1.1) failing to satisfy the MFCQ, see Definitions 2.3 and 2.4 in [4].

Case 2. $\|c(\bar{x})\| > 0$. Then $\|c(x^k)\| > 0$ for all $k \in \bar{\mathcal{K}}$ large enough. We have from the second equation of (1.6) that

$$\|\nabla c(x^k)^\top\triangle\bar{x}^k\| = \|c(x^k)\|\left\|\frac{y^k + \triangle\bar{y}^k}{\gamma_k} - \frac{c(x^k)}{\|c(x^k)\|}\right\|.$$

Since $\{\|\triangle\bar{x}^k\|\}_{\bar{\mathcal{K}}} \to 0$ and $\|c(\bar{x})\| > 0$, it follows that $\left\{\frac{y^k+\triangle\bar{y}^k}{\gamma_k}\right\}_{\bar{\mathcal{K}}} \to \frac{c(\bar{x})}{\|c(\bar{x})\|}$ and thus $\left\{\frac{\|y^k+\triangle\bar{y}^k\|}{\gamma_k}\right\}_{\bar{\mathcal{K}}} \to 1$. Hence, dividing (3.1) and condition (ii) of (2.20) by $\|(\lambda^k + \triangle\bar{\lambda}^k, y^k + \triangle\bar{y}^k)\|$ and letting $k \in \bar{\mathcal{K}} \to \infty$ yields that $\bar{\lambda}^\top\bar{x} = 0$ and $\frac{\|\bar{y}\|}{\|c(\bar{x})\|}c(\bar{x})\nabla c(\bar{x}) = \bar{\lambda}$. This implies that $\|\bar{y}\| \neq 0$ as $(\bar{\lambda}, \bar{y}) \neq 0$. Since $\bar{x} \geq 0$ and $\bar{\lambda} \geq 0$, it follows that $\bar{x}$ is a KKT point of the feasibility problem $\min_{x\geq 0}\ \|c(x)\|^2$. □

LEMMA 3.2. *Suppose Assumption A holds. If $\gamma_k = \bar{\gamma}$ for all large k, then $\{x^k\}$ and $\{\lambda^k\}$ are componentwise bounded away from zero and $\{\lambda^k\}$ is bounded above.*

Proof. Suppose first that after a finite number of iterations, the iterates are always generated by the Armijo conditions (2.10) or (2.18), i.e., $p_k = 0$ for all k large enough according Step 3 of Algorithm 2.1. Since $\gamma_k = \bar{\gamma}$ eventually, it follows that $\Phi_{\mu,\bar{\gamma}}(x^{k+1}) \leq \Phi_{\mu,\bar{\gamma}}(x^k)$ for all k large enough. Therefore, the sequence

$\{\varphi_\mu(x^k)\}$ is bounded above. Since $\{f(x^k)\}$ is bounded by Assumption A1, the sequence $\{-\mu\sum_{i=1}^n \ln x_i^k\}$ is bounded above. This implies that $\{x^k\}$ is componentwise bounded away from zero.

We now consider the case that there are an infinite number of indices k such that $p_k = 1$. Suppose the sequence $\{\varphi_\mu^{k,N_k}\}$ is not bounded above. Then there exists an infinite index set $\mathcal{K}$ such that for all $k \in \mathcal{K}$, x^{k+1} is accepted by the PLPF criteria and $\varphi_\mu(x^{k+1}) = \varphi_\mu^{k+1,N_{k+1}} > \varphi_\mu^{k,N_k}$ and $\|c(x^{k+1})\|_p = \theta^{k+1,N_{k+1}}$.

First consider the case that for all $k \in \mathcal{K}$, the iterates x^{k+1} are generated at Step 3.2 of Algorithm 2.1 with $x^{k+1} = x^k + \alpha_x^k \triangle x^k$. Let

$$\hat{\omega}_k = \sigma \max\{\alpha_x^k, \alpha_k^{\min}\}\omega^k \text{ and } \xi_k = \sigma \max\{\alpha_x^k, \alpha_k^{\min}\}\|c(x^k)\|_p, \tag{3.2}$$

where ω^k is defined by (2.8). Let $\bar{\mathcal{P}}_\mu^k(\rho) = \mathcal{P}_\mu^k(\rho) - \hat{\omega}_k - \rho\xi_k$ for $\rho \geq 0$. It follows trivially that $\bar{\mathcal{P}}_\mu^k(\rho)$ is a piecewise linear function as

$$\bar{\mathcal{P}}_\mu^k(\rho) = \min\{\varphi_\mu^{k,i} - \hat{\omega}_k + \rho(\theta^{k,i} - \xi_k),\ i = 1, ..., N_k\},\ \forall \rho \geq 0. \tag{3.3}$$

Moreover, it is easy to verify that $\bar{\mathcal{P}}_\mu^k(\rho)$ has the same break points as those of $\mathcal{P}_\mu^k(\rho)$. In particular, we have

$$\bar{\mathcal{P}}_\mu^k(\rho) = \begin{cases} \varphi_\mu^{k,1} - \hat{\omega}_k + \rho(\theta^{k,1} - \xi_k), & \rho^{k,1} \leq \rho \leq \rho^{k,2}, \\ \vdots & \vdots \\ \varphi_\mu^{k,N_k-1} - \hat{\omega}_k + \rho(\theta^{k,N_k-1} - \xi_k), & \rho^{k,N_k-1} \leq \rho \leq \rho^{k,N_k}, \\ \varphi_\mu^{k,N_k} - \hat{\omega}_k + \rho(\theta^{k,N_k} - \xi_k), & \rho^{k,N_k} \leq \rho, \end{cases} \tag{3.4}$$

Note that $\bar{\mathcal{P}}_\mu^k(\rho)$ is not necessarily an increasing function since the slope $\theta^{k,N_i} - \xi_k$ could be negative for some $i = 1, ..., N_k$.

We now prove that (2.9) holds for all $k \in \mathcal{K}$. Assume to the contrary that for some $k \in \mathcal{K}$,

$$\|c(x^{k+1})\|_p - \theta^{k,N_k} > -\xi_k. \tag{3.5}$$

Multiplying both sides of (3.5) by ρ and adding it to the inequality $\varphi_\mu(x^{k+1}) - \varphi_\mu^{k,N_k} > 0$ yields that

$$\varphi_\mu(x^{k+1}) + \rho\|c(x^{k+1})\|_p - \varphi_\mu^{k,N_k} - \rho\theta^{k,N_k} > -\rho\xi_k \geq -\hat{\omega}_k - \rho\xi_k,\ \forall \rho \geq 0. \tag{3.6}$$

Letting $\rho = \rho^{k,i}$ $(i = 1, ..., N_k)$ in (3.6), we have from (3.6), (3.3) and (3.4) that

$$\begin{aligned} & \varphi_\mu(x^{k+1}) + \rho^{k,i}\|c(x^{k+1})\|_p \\ > \ & \varphi_\mu^{k,N_k} - \hat{\omega}_k + \rho^{k,i}(\theta^{k,N_k} - \xi_k) \\ \geq \ & \bar{\mathcal{P}}_\mu^k(\rho^{k,i}) = \varphi_\mu^{k,N_i} - \hat{\omega}_k + \rho^{k,i}(\theta^{k,N_i} - \xi_k). \end{aligned} \tag{3.7}$$

Hence, (2.7) does not hold for any $i = 1, ..., N_k$. This together with the violation of (2.9) implies that x^{k+1} is not acceptable by the PLPF criteria. We get a contradiction. Therefore, (2.9) holds for all $k \in \mathcal{K}$ and thus

$$\theta^{k+1,N_{k+1}} = \|c(x^{k+1})\|_p \leq \theta^{k,N_k} - \sigma\alpha_x^{\min}\|c(x^k)\|_p,\ \forall k \in \mathcal{K}. \tag{3.8}$$

Note that (3.8) can be proved similarly if for all $k \in \mathcal{K}$, the iterates x^{k+1} are generated by the second-order step of Algorithm 2.1, i.e., $x^{k+1} = x^k + \hat{\alpha}_x^k \triangle \hat{x}^k$. In this case we just need to replace $\triangle x^k$ and α_x^k in (3.2) by $\triangle \hat{x}^k$ and $\hat{\alpha}_x^k$, respectively.

We now prove that $\{\theta^{k,N_k}\}$ is a non-increasing sequence. To see this, suppose that a new point $\{(\varphi_\mu(x^{k+1}), \|c(x^{k+1})\|_p)\}$ is added to $\mathcal{A}_\mu^{k+1}$. If $\|c(x^{k+1})\|_p \leq \theta^{k,N_k}$, we have $\theta^{k+1,N_{k+1}} = \|c(x^{k+1})\|_p \leq \theta^{k,N_k}$. If $\|c(x^{k+1})\|_p > \theta^{k,N_k}$, we have for all sufficiently large $\rho \geq 0$,

$$\varphi^{k,N_k} + \rho\theta^{k,N_k} = \min\{\mathcal{P}_\mu^k(\rho), \varphi_\mu(x^{k+1}) + \rho\|c(x^{k+1})\|_p\}.$$

Hence, it follows from (1.10) that $\theta^{k+1,N_{k+1}} = \theta^{k,N_k}$.

Since the sequence $\{\theta^{k,N_k}\}$ is non-increasing, we have from (3.8) that $\{\|c(x^k)\|_p\} \to 0$ as $k \in \mathcal{K} \to \infty$. Hence, it follows that $\|c(x^k)\|_p \le \theta^{\min}$ for all $k \in \mathcal{K}$ large enough. This implies that the first condition in (2.11) or (2.19) holds for all $k \in \mathcal{K}$ large enough. By Step 3.2 of Algorithm 2.1, if (2.11) or (2.19) holds, the PLPF criteria are not used to accept trial points. Therefore, the second condition in (2.11) or (2.19) is violated and thus $\varphi_\mu(x^{k+1}) < \varphi_\mu^{k,N_k}$ for all $k \in \mathcal{K}$ large enough. However, this contradicts our assumption that $\varphi_\mu(x^{k+1}) > \varphi_\mu^{k,N_k}$ for all $k \in \mathcal{K}$. Therefore, we have proved that the sequence $\{\varphi_\mu^{k,N_k}\}$ is bounded above.

Let $\varphi_\mu^{\sup} = \sup_k\{\varphi_\mu^{k,N_k}\}$ and $\theta^{\sup} = \sup_k\{\|c(x^k)\|\}$ as $\{\|c(x^k)\|\}$ is bounded by Assumption A1. For every x^k that is generated by the PLPF criteria, we have $\varphi_\mu(x^k) \le \varphi_\mu^{\sup}$. Since $\gamma_k = \bar\gamma$ eventually, for every x^k that is generated by the Armijo conditions, if k is large enough, we have

$$\Phi_{\mu,\bar\gamma}(x^k) \le \Phi_{\mu,\bar\gamma}(x^{k-1}) \le ... \le \Phi_{\mu,\bar\gamma}(x^{k-l}) \le \varphi_\mu^{\sup} + \bar\gamma\theta^{\sup},$$

where x^{k-l} is the first iterate before x^k that is accepted by the PLPF criteria. Therefore, we conclude that $\{\Phi_{\mu,\gamma_k}(x^k)\}$ and thus $\{\varphi_\mu(x^k)\}$ are bounded. Using the same argument as the one at the beginning of the proof, we obtain that $\{x^k\}$ is componentwise bounded away from zero.

Since $\{x^k\}$ is also bounded, we obtain from (2.13) that $\{\lambda^k\}$ is bounded above and componentwise bounded away from zero. □

LEMMA 3.3. *Suppose Assumption A holds and $\gamma_k = \bar\gamma$ for all k large enough. If the sequence $\{\|\mathcal{M}_k^{-1}\|\}_\mathcal{K}$ tends to infinity, where $\mathcal{K}$ is an infinite index set, then any accumulation point of x^k is a Fritz-John point of problem (1.1) at which the MFCQ fails to hold.*

Proof. Since $\{\|\mathcal{M}_k^{-1}\|\}_\mathcal{K} \to \infty$, any limit matrix of $\{\mathcal{M}_k\}_\mathcal{K}$ is singular. The rest of the proof is identical to the proof of Lemma 3.8 in [4]. □

LEMMA 3.4. *Suppose Assumption A holds and $\gamma_k = \bar\gamma$ for all k large enough. If the sequence $\{\|\mathcal{M}_k^{-1}\|\}_\mathcal{K}$ is bounded and $\{\|\triangle x^k\|\}_\mathcal{K} \to 0$, where $\mathcal{K}$ is an infinite index set, then $\{\|c(x^k)\|\}_\mathcal{K} \to 0$, the sequence $\{(x^{k+1}, \lambda^{k+1}, y^{k+1})\}_\mathcal{K}$ is bounded and any accumulation point of it satisfies the first-order optimality conditions (1.3).*

Proof. We first prove that $\{\|c(x^k)\|\}_\mathcal{K} \to 0$. By Step 2.5 of Algorithm 2.1, if (2.4) holds, $\triangle x^k = \triangle\bar x^k$; otherwise, $\triangle x^k = \triangle\tilde x^k$ and $\|\triangle\bar x^k\| \le \|\triangle x^k\| + \|\triangle x^k\|^\vartheta$ by condition (i) of (2.4). Since $\{\|\triangle x^k\|\}_\mathcal{K} \to 0$, it follows that $\{\|\triangle\bar x^k\|\}_\mathcal{K} \to 0$. Since $\gamma_k = \bar\gamma$, we have from Step 4.3 of Algorithm 2.1 that $\pi_k = \bar\pi$ for some $\bar\pi > 0$ and for all k large enough. This implies that condition (i) of (2.20) holds for all $k \in \mathcal{K}$ large enough. Since $\{\lambda^k\}$ is bounded by Lemma 3.2, we have from (1.7) that

$$\{\|X_k(\lambda^k + \triangle\bar\lambda^k) - \mu e\|\}_\mathcal{K} \to 0 \text{ as } \{\|\triangle\bar x^k\|\}_\mathcal{K} \to 0.$$

Therefore, condition (ii) of (2.20) holds for all $k \in \mathcal{K}$ large enough. However, the satisfaction of conditions (i) and (ii) of (2.20) implies that the last condition (i.e., condition (iii) of (2.20)) that may trigger the increase of γ_k must be violated for all $k \in \mathcal{K}$ large enough. Hence, we obtain from the second equation of (1.6) that for all $k \in \mathcal{K}$ large enough,

$$\|\nabla c(x^k)^\top \triangle\bar x^k\| \ge (1 - \frac{1}{\eta})\|c(x^k)\|.$$

This implies that $\{\|c(x^k)\|\}_\mathcal{K} \to 0$ as $\eta > 1$ and $\{\|\nabla c(x^k)\|\}$ is bounded.

Now we prove that the sequence $\{(\lambda^k + \triangle\lambda^k, y^k + \triangle y^k)\}_\mathcal{K}$ is bounded. To obtain a contradiction, suppose there exists an infinite index set $\bar\mathcal{K} \subseteq \mathcal{K}$ such that

$$\{x^k\}_{\bar\mathcal{K}} \to x^*, \{\|(\lambda^k + \triangle\lambda^k, y^k + \triangle y^k)\|\}_{\bar\mathcal{K}} \to \infty \text{ and } \left\{\frac{y^k + \triangle y^k}{\max\{1, \|y^k + \triangle y^k\|\}}\right\}_{\bar\mathcal{K}} \to y^*.$$

We have from (1.7) and (2.2) that

$$\lambda^k + \triangle\lambda^k = \mu X_k^{-1} e - X_k^{-1}\Lambda_k \triangle x^k. \tag{3.9}$$

By Lemma 3.2 the sequence $\{\lambda^k\}$ is bounded and $x^* > 0$. Therefore, (3.9) implies that $\{\lambda^k + \triangle\lambda^k\}_{\bar{\mathcal{K}}}$ is bounded as $\{\triangle x^k\}_{\bar{\mathcal{K}}} \to 0$. Thus, we have $\{\|y^k + \triangle y^k\|\}_{\bar{\mathcal{K}}} \to \infty$. From the first equations of (1.6) and (2.1) we have

$$(\mathcal{H}_k + X_k^{-1}\Lambda_k)\triangle x^k + \nabla c(x^k)(y + \triangle y^k) = -\nabla f(x^k) + \mu X_k^{-1} e. \tag{3.10}$$

Since $\{\mathcal{H}_k\}$ is bounded by Assumption A2, dividing both sides of (3.10) by $\max\{1, \|y + \triangle y^k\|\}$ and letting $k \in \bar{\mathcal{K}} \to \infty$ yields that $\nabla c(x^*)y^* = 0$. Since $\|y^*\| = 1$, $y^* \neq 0$ and the Jacobian $\nabla c(x^*)$ is rank-deficient. Moreover, since $\{\|c(x^k)\|\}_{\mathcal{K}} \to 0$, we have $c(x^*) = 0$. Hence, we conclude that the limit matrix of $\{\mathcal{M}_k\}_{\bar{\mathcal{K}}}$ is singular. This contradicts our assumption that $\{\|\mathcal{M}_k^{-1}\|\}_{\mathcal{K}}$ is bounded.

Now we prove that any accumulation point of the sequence $\{(x^{k+1}, \lambda^{k+1}, y^{k+1})\}_{\mathcal{K}}$ satisfies (1.3). Suppose there is an infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{x^k\} \to x^*$, $\{\lambda^k + \triangle\lambda^k\} \to \lambda^*$ and $\{y^k + \triangle y^k\} \to y^*$ as $k \in \bar{\mathcal{K}} \to \infty$. By (3.9) and (3.10) we have

$$\mathcal{H}_k \triangle x^k - (\lambda^k + \triangle\lambda^k) + \nabla c(x^k)^\top (y^k + \triangle y^k) = -\nabla f(x^k). \tag{3.11}$$

Letting $k \in \bar{\mathcal{K}} \to \infty$ in (3.9) and (3.11) yields that $\lambda^* = \mu X_*^{-1} e$, $X_* = \mathrm{diag}(x^*)$, and $\nabla_x \mathcal{L}(x^*, \lambda^*, y^*) = 0$. This together with the fact that $c(x^*) = 0$ imply that (x^*, λ^*, y^*) is a solution of (1.3).

We now show that $\{x^{k+1}, \lambda^{k+1}, y^{k+1}\}_{\bar{\mathcal{K}}} \to (x^*, \lambda^*, y^*)$. At any iteration k, there are two cases for updating the iterate: $x^{k+1} = x^k + \alpha_x^k \triangle x^k$ or $x^{k+1} = x^k + \hat{\alpha}_x^k \triangle \hat{x}^k$. If there is an infinite index set $\hat{\mathcal{K}} \subseteq \bar{\mathcal{K}}$ such that $x^{k+1} = x^k + \alpha_x^k \triangle x^k$ for all $k \in \hat{\mathcal{K}}$, we clearly have $\lim_{k\in\hat{\mathcal{K}}\to\infty} x^{k+1} = \lim_{k\in\hat{\mathcal{K}}\to\infty} x^k = x^*$. In the second case, i.e., if a second-order correction step is used, we have $\triangle x^k = \triangle \tilde{x}^k$ by Step 3.4.1 of Algorithm 2.1. From (2.1) and (2.14) we have

$$\begin{bmatrix} \triangle \hat{x}^k - \triangle \tilde{x}^k \\ \triangle \hat{y}^k - \triangle \tilde{y}^k \end{bmatrix} = -\mathcal{M}_k^{-1} \begin{bmatrix} 0 \\ c(x^k + \bar{\alpha}_x^k \triangle \tilde{x}^k) - \bar{\alpha}_x^k \nabla c(x^k)^\top \triangle \tilde{x}^k - c(x^k) \end{bmatrix}. \tag{3.12}$$

If there is an infinite index set $\hat{\mathcal{K}} \subseteq \bar{\mathcal{K}}$ such that $x^{k+1} = x^k + \hat{\alpha}_x^k \triangle \hat{x}^k$ for all $k \in \hat{\mathcal{K}}$, the right hand side of the above equations and hence the sequence $\{\triangle \hat{x}^k\}$ tend to zero as $\{\|\mathcal{M}_k^{-1}\|\}_{\hat{\mathcal{K}}}$ is bounded and the sequences $\{\triangle \tilde{x}^k (= \triangle x^k)\}_{\hat{\mathcal{K}}}$, $\{c(x^k)\}_{\hat{\mathcal{K}}}$ and $\{c(x^k + \bar{\alpha}_x^k \triangle \tilde{x}^k)\}_{\hat{\mathcal{K}}}$ tend to zero. Therefore, it follows that $\{x^{k+1}\}_{\hat{\mathcal{K}}} \to x^*$. Combining the above two cases, we obtain that $\{x^{k+1}\}_{\bar{\mathcal{K}}} \to x^*$.

Since $\lambda^* = \mu X_*^{-1} e$, we have for all $k \in \bar{\mathcal{K}}$ large enough that

$$\varrho \mu X_{k+1}^{-1} e \geq \lambda^k + \triangle\lambda^k \geq \frac{\mu}{\varrho} X_{k+1}^{-1} e$$

as $\varrho > 1$. This together with (2.12) and (2.13) imply that $\alpha_\lambda^k = 1$ and $\lambda^{k+1} = \lambda^k + \triangle\lambda^k$ for all $k \in \bar{\mathcal{K}}$ large enough. Hence, $\{\lambda^{k+1}\}_{\bar{\mathcal{K}}} \to \lambda^*$. Since $\{\|\triangle x^k\|\}_{\bar{\mathcal{K}}} \to 0$, we have from Step 4.1 of Algorithm 2.1 that $\alpha_y^k = 1$, i.e., $y^{k+1} = y^k + \triangle y^k$, for all $k \in \bar{\mathcal{K}}$ large enough. Hence, $\{y^{k+1}\}_{\bar{\mathcal{K}}} \to y^*$.

Since $\bar{\mathcal{K}}$ could be any chosen infinite subset of $\mathcal{K}$, the lemma follows. □

LEMMA 3.5. *Suppose Assumption A holds and $\gamma_k = \bar{\gamma}$ for all k large enough. If there exists an infinite index set $\mathcal{K}$ such that $p_k = 1$ for all $k \in \mathcal{K}$, then*

$$\lim_{k\in\mathcal{K}\to\infty} \min\{\|c(x^k)\|_p, \|\triangle x^k\|\} = 0.$$

Proof. For the purpose of obtaining a contradiction, suppose there exists an infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that

$$\|c(x^k)\|_p \geq \varepsilon \text{ and } \omega_k \geq \varepsilon \tag{3.13}$$

for all $k \in \bar{\mathcal{K}}$ and for some $\varepsilon > 0$. By Assumption A1, the sequences $\{f(x^k)\}$ and $\{\|c(x^k)\|_p\}$ are bounded. Moreover, the sequence $\{\varphi_\mu(x^k)\}$ is bounded as $\{x^k\}$ is componentwise bounded away from zero by Lemma 3.2. Therefore, we can define

$$\overline{\varphi} = \sup_k\{\varphi_\mu(x^k)\},\ \underline{\varphi} = \inf_k\{\varphi_\mu(x^k)\} \text{ and } \overline{\theta} = \sup_k\{\|c(x^k)\|_p\}.$$

Let $\mathcal{Q}$ be the following rectangle that contains all the pairs $(\varphi_\mu(x^k), \|c(x^k)\|_p)$:

$$\mathcal{Q} = \left[\underline{\varphi}, \overline{\varphi} + \sigma\alpha_x^{\min}\varepsilon\right] \times \left[0, \overline{\theta} + \sigma\alpha_x^{\min}\varepsilon\right].$$

For every k, define sets $\mathcal{Q}^k$ and $\hat{\mathcal{Q}}^k$ as

$$\begin{aligned}
\mathcal{Q}^k &= \left\{(\varphi,\theta) \in \mathcal{Q} | \varphi + \rho^{k,i}\theta > \varphi^{k,i} + \rho^{k,i}\theta^{k,i}, i = 1, ..., N_k,\ \theta > \theta^{k,N_k}\right\}, \\
\hat{\mathcal{Q}}^k &= \left\{(\varphi,\theta) \in \mathcal{Q} | \varphi + \rho^{k,i}\theta > \hat{\varphi}^{k,i} + \rho^{k,i}\hat{\theta}^{k,i}, i = 1, ..., N_k,\ \theta > \hat{\theta}^{k,N_k}\right\},
\end{aligned}$$

where

$$\hat{\varphi}_\mu^{k,i} = \varphi_\mu^{k,i} - \sigma\alpha_x^{\min}\varepsilon \text{ and } \hat{\theta}^{k,i} = \theta^{k,i} - \sigma\alpha^{\min}\varepsilon,\ i = 1, ..., N_k.$$

Since $p_k = 1$ (i.e., x^{k+1} is accepted according to the PLPF criteria) for all $k \in \bar{\mathcal{K}}$, we know from (2.7) (or (2.16)) and (2.9) (or (2.17)) that

$$(\varphi_\mu(x^{k+1}), \|c(x^{k+1})\|_p) \in \mathcal{Q}\backslash\hat{\mathcal{Q}}^k,\ \forall k \in \bar{\mathcal{K}}.$$

This together with the definitions of $\hat{\varphi}_\mu^{k,i}$ and $\hat{\theta}^{k,i}$ $(i = 1, ..., N_k)$ imply that

$$(\varphi_\mu(x^{k+1}) + \sigma\alpha^{\min}\varepsilon, \|c(x^{k+1})\|_p + \sigma\alpha^{\min}\varepsilon) \in \mathcal{Q}\backslash\mathcal{Q}^k,\ \forall k \in \bar{\mathcal{K}}.$$

Since the pair $(\varphi_\mu(x^{k+1}) + \sigma\alpha^{\min}\varepsilon, \|c(x^{k+1})\|_p + \sigma\alpha^{\min}\varepsilon)$ is dominated by all other pairs in the square

$$\tilde{\mathcal{Q}}^k = \left[\varphi_\mu(x^{k+1}), \varphi_\mu(x^{k+1}) + \sigma\alpha^{\min}\varepsilon\right] \times \left[\|c(x^{k+1})\|_p, \|c(x^{k+1})\|_p + \sigma\alpha^{\min}\varepsilon\right],$$

it follows that

$$\tilde{\mathcal{Q}}^k \subseteq \mathcal{Q}\backslash\mathcal{Q}^k,\ \forall k \in \bar{\mathcal{K}}. \tag{3.14}$$

On the other hand, according to our procedure for updating the PLPF, the fact that $p_k = 1$ also implies $(\varphi_\mu(x^{k+1}), \|c(x^{k+1})\|_p) \in \mathcal{A}_\mu^{k+1}$ for all $k \in \bar{\mathcal{K}}$. Since $(\varphi_\mu(x^{k+1}), \|c(x^{k+1})\|_p)$ dominates all other pairs in $\tilde{\mathcal{Q}}^k$, it follows that

$$\tilde{\mathcal{Q}}^k \subseteq \mathrm{cl}(\mathcal{Q}^{k+1}),\ \forall k \in \bar{\mathcal{K}}. \tag{3.15}$$

Notice that $\mathcal{Q}^k$ never shrinks. In particular, for any $\rho \geq 0$, we have from (1.10) that $\mathcal{P}_\mu^k(\rho) \geq \mathcal{P}_\mu^{k+1}(\rho)$ for all k. For any pair $(\varphi,\theta) \in \mathcal{Q}^k$, we have $\theta > \theta^{k,N_k}$ and the relation $\varphi + \rho\theta > \mathcal{P}_\mu^k(\rho)$ holds at all break points $\rho = \rho^{k,i}$ $(i = 1, ..., N_k)$. Hence, using the piecewise linear concavity of $\mathcal{P}_\mu^k(\rho)$, we obtain that

$$\varphi + \rho\theta > \mathcal{P}_\mu^k(\rho) \geq \mathcal{P}_\mu^{k+1}(\rho),\ \forall \rho \geq 0.$$

This implies $(\varphi,\theta) \in \mathcal{Q}^{k+1}$ and thus $\mathcal{Q}^k \subseteq \mathcal{Q}^{k+1}$ for all k. Therefore, we have from (3.15) that

$$\mathrm{cl}(\mathcal{Q}^k) \cup \tilde{\mathcal{Q}}^k \subseteq \mathrm{cl}(\mathcal{Q}^{k+1}),\ \forall k \in \bar{\mathcal{K}}. \tag{3.16}$$

Since the area of $\tilde{\mathcal{Q}}^k$ is $(\sigma\alpha_x^{\min}\varepsilon)^2$ for every $k \in \bar{\mathcal{K}}$, we obtain from (3.14) and (3.16) that the area of $\mathrm{cl}(\mathcal{Q}^k)$ tends to infinity as k goes to infinity. However, this contradicts the fact that $\mathrm{cl}(\mathcal{Q}^k) \subseteq \mathcal{Q}$ for all k. □

LEMMA 3.6. *Suppose Assumption A holds and $\gamma_k = \bar{\gamma}$ for all k large enough. Let $\mathcal{K}$ be the set of iterations at which the PLPF criteria are used to accept iterates, i.e., $\mathcal{K} = \{k | p_k = 1\}$. If $\mathcal{K}$ is an infinite set and the sequence $\{\|\mathcal{M}_k^{-1}\|\}_{\mathcal{K}}$ is bounded, then the sequence $\{(x^{k+1}, \lambda^{k+1}, y^{k+1})\}_{\mathcal{K}}$ is bounded and any accumulation point of it satisfies the first-order optimality conditions (1.3).*

Proof. Suppose there exists an infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\|c(x^k)\|_p > \theta^{\min}$ for all $k \in \bar{\mathcal{K}}$, where $\theta^{\min} > 0$ is the constant used in (2.11) and (2.19). By Lemma 3.5, we have $\{\|\triangle x^k\|\}_{\bar{\mathcal{K}}} \to 0$. This together with Lemma 5.44 imply $\{\|c(x^k)\|\}_{\bar{\mathcal{K}}} \to 0$. We obtain a contradiction. Hence, it follows that $\|c(x^k)\|_p \le \theta^{\min}$ for all $k \in \mathcal{K}$ large enough.

By Step 3.2 (or Step 3.4.3 if a second-order correction step is used), a trial point needs to violate (2.11) (or (2.19)) in order to be accepted by the PLPF criteria. Since the first condition in (2.11) (or 2.19) always holds when $k \in \mathcal{K}$ is large enough, the second condition in them must be eventually violated. Hence, we have for all $k \in \mathcal{K}$ large enough that

$$\varphi_\mu(x^{k+1}) \le \varphi_\mu^{k,N_k} - \sigma \alpha_x^{\min} \omega_k, \tag{3.17}$$

and thus $\varphi_\mu^{k+1,N_{k+1}} = \varphi_\mu(x^{k+1})$ as the pair $(\varphi_\mu^{k+1,N_{k+1}}, \theta^{k+1,N_{k+1}})$ is added to $\mathcal{A}_\mu^{k+1}$ at Step 4.2 of Algorithm 2.1. Since $\varphi^{k+1,N_{k+1}} = \varphi^{k,N_k}$ if $p_k = 0$, the sequence $\{\varphi^{k,N_k}\}$ is non-increasing when k is large enough. Since $\{\varphi^{k,N_k}\}$ is bounded below, we have from (3.17) that $\{\omega_k\}_{\mathcal{K}} \to 0$ and hence $\{\|\triangle x^k\|\}_{\mathcal{K}} \to 0$. The lemma now follows from Lemma 5.44. ∎

LEMMA 3.7. *Suppose Assumption A holds and $\gamma_k = \bar{\gamma}$ and $p_k = 0$ for all k large enough. If the sequence $\{\|\mathcal{M}_k^{-1}\|\}$ is bounded, then $\{\|\triangle x^k\|\} \to 0$.*

Proof. From (1.6) we have

$$\mathcal{M}_k \begin{bmatrix} \triangle \bar{x}^k \\ y^k + \triangle \bar{y}^k \end{bmatrix} = - \begin{bmatrix} \nabla f(x^k) - \mu X_k^{-1} e \\ c(x^k) \end{bmatrix}. \tag{3.18}$$

Since $\{\nabla f(x^k)\}$ and $\{c(x^k)\}$ are bounded by Assumption A1 and $\{X_k^{-1}\}$ is bounded by Lemma 3.2, the right hand side of (3.18) is bounded. Hence, the boundedness of $\{\|\mathcal{M}_k^{-1}\|\}$ implies the boundedness of $\{\|\triangle \bar{x}^k\|\}$ and $\{\|y^k + \triangle \bar{y}^k\|\}$.

To prove the boundedness of $\{\|\triangle x^k\|\}$, assume to the contrary that there exists an infinite index set $\mathcal{K}$ such that $\{\|\triangle x^k\|\}_{\mathcal{K}}$ tends to infinity. Since $\{\|\triangle \bar{x}^k\|\}_{\mathcal{K}}$ is bounded, it follows that $\triangle x^k = \triangle \tilde{x}^k$, i.e., (2.4) holds, for all $k \in \mathcal{K}$ large enough. Hence, $\{\|\triangle \tilde{x}^k\|\}_{\mathcal{K}} \to \infty$. Since $\vartheta \in (0,1)$, dividing both sides of condition (i) of (2.4) by $\|\triangle \tilde{x}^k\|$ and letting $k \in \mathcal{K} \to \infty$ yields that

$$1 = \lim_{k \in \mathcal{K}} \frac{\|\triangle \tilde{x}^k - \triangle \bar{x}^k\|}{\|\triangle \tilde{x}^k\|} \le \lim_{k \in \mathcal{K}} \|\triangle \tilde{x}^k\|^{\vartheta - 1} = 0.$$

This yields a contradiction.

By Step 3.4.1 of Algorithm 2.1, a second-order correction step is computed only if $\triangle x^k = \triangle \tilde{x}^k$. Given the boundedness of $\{\|\triangle x^k\|\}$, we know from (3.12) that the sequence of second-order correction steps $\{\|\triangle \hat{x}^k\|\}$ is bounded.

We now prove $\{\|\triangle x^k\|\} \to 0$. Suppose there exists an infinite index set $\mathcal{K}$ such that $\{x^k\}_{\mathcal{K}} \to x^*$ and $\{\triangle x^k\}_{\mathcal{K}} \to \triangle x^* \ne 0$. Clearly, $x^* > 0$ by Lemma 3.2. If $\|c(x^*)\| > 0$, the ℓ_2-penalty function $\Phi_{\mu,\bar{\gamma}}(x)$ is continuously differentiable on a small neighborhood of x^* and hence

$$\{\Phi'_{\mu,\bar{\gamma}}(x^k; \triangle x^k)\}_{\mathcal{K}} \to \Phi'_{\mu,\bar{\gamma}}(x^*; \triangle x^*). \tag{3.19}$$

To prove (3.19) when $\|c(x^*)\| = 0$, there are two cases.

Case 1. Consider any infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\triangle x^k = \triangle \bar{x}^k$ for all $k \in \bar{\mathcal{K}}$. We have from the second equation of (1.6) that

$$\nabla c(x^k)^\top \triangle \bar{x}^k - \frac{\|c(x^k)\|}{\bar{\gamma}} (y^k + \triangle \bar{y}^k) = -c(x^k). \tag{3.20}$$

Since $\{y^k + \triangle\bar{y}^k\}_{\bar{\mathcal{K}}}$ is bounded and $c(x^*) = 0$, (3.20) implies $\nabla c(x^*)^\top \triangle x^* = 0$ and hence

$$\begin{aligned}
&\Phi'_{\mu,\bar{\gamma}}(x^*; \triangle x^*) \\
&= \nabla\varphi_\mu(x^*)^\top \triangle x^* + \lim_{t\downarrow 0} \frac{\|c(x^* + t\triangle x^*)\| - \|c(x^*)\|}{t/\bar{\gamma}} \\
&= \nabla\varphi_\mu(x^*)^\top \triangle x^* + \lim_{t\downarrow 0} \frac{\|c(x^*) + t\nabla c(x^*)^\top \triangle x^*)\| - \|c(x^*)\|}{t/\bar{\gamma}} \\
&= \nabla\varphi_\mu(x^*)^\top \triangle x^*.
\end{aligned} \tag{3.21}$$

By Lemma 3.2 in [4], we have

$$\Phi'_{\mu,\bar{\gamma}}(x^k; \triangle\bar{x}^k) = \nabla\varphi_\mu(x^k)^\top \triangle\bar{x}^k + c(x^k)^\top (y^k + \triangle\bar{y}^k) - \bar{\gamma}\|c(x^k)\|. \tag{3.22}$$

This together with (3.21) imply that

$$\{\Phi'_{\mu,\bar{\gamma}}(x^k; \triangle x^k)\}_{\bar{\mathcal{K}}} \to \nabla\varphi_\mu(x^*)^\top \triangle x^* = \Phi'_{\mu,\bar{\gamma}}(x^*; \triangle x^*). \tag{3.23}$$

Case 2. Consider any infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\triangle x^k = \triangle\tilde{x}^k$, i.e., (2.4) holds, for all $k \in \bar{\mathcal{K}}$. We have the second equation of (2.1) that

$$\nabla c(x^k)^\top \triangle\tilde{x}^k - \frac{\|c(x^k)\|}{\bar{\gamma}} \triangle\tilde{y}^k = -c(x^k). \tag{3.24}$$

Since $\{\triangle\tilde{y}^k\}_{\bar{\mathcal{K}}}$ is bounded by condition (ii) of (2.4) and $\|c(x^*)\| = 0$, (3.24) implies $\nabla c(x^*)^\top \triangle x^* = 0$. Hence (3.21) holds. Also by (3.24), if $\|c(x^k)\| > 0$, we have

$$\begin{aligned}
&\Phi'_{\mu,\bar{\gamma}}(x^k; \triangle\tilde{x}^k) \\
&= \nabla\varphi_\mu(x^k)^\top \triangle\tilde{x}^k + \tfrac{\bar{\gamma}}{\|c(x^k)\|} c(x^k)^\top \nabla c(x^k)^\top \triangle\tilde{x}^k \\
&= \nabla\varphi_\mu(x^k)^\top \triangle\tilde{x}^k + \tfrac{\bar{\gamma}}{\|c(x^k)\|} c(x^k)^\top \left(\tfrac{\|c(x^k)\|}{\bar{\gamma}} \triangle\tilde{y}^k - c(x^k)\right) \\
&= \nabla\varphi_\mu(x^k)^\top \triangle\tilde{x}^k + c(x^k)^\top \triangle\tilde{y}^k - \bar{\gamma}\|c(x^k)\|.
\end{aligned} \tag{3.25}$$

If $\|c(x^k)\| = 0$, we have from (3.24) that $\nabla c(x^k)^\top \triangle\tilde{x}^k = 0$, and thus, similarly to (3.21), we have $\Phi'_{\mu,\bar{\gamma}}(x^k; \triangle\tilde{x}^k) = \nabla\varphi_\mu(x^k)^\top \triangle\tilde{x}^k$. This together with (3.25) and (3.21) imply (3.23) as $\{\triangle\tilde{y}^k\}_{\bar{\mathcal{K}}}$ is bounded and $\|c(x^*)\| = 0$.

Combining Cases 1 and 2, we have proved (3.19).

Since $x^* > 0$, we know from (2.6) that there exists a $\bar{\alpha} \in (0, 1]$ such that $\bar{\alpha}_x^k \geq \bar{\alpha}$ for all $k \in \mathcal{K}$. By (2.21), condition (iii) of (2.4) and (2.5), we have

$$\Phi'_{\mu,\bar{\gamma}}(x^k; \triangle x^k) \leq -\zeta\nu\|\triangle x^k\|^2. \tag{3.26}$$

Hence, we have from (3.19) that

$$\Phi'_{\mu,\bar{\gamma}}(x^*; \triangle x^*) \leq -\zeta\nu\|\triangle x^*\|^2.$$

Since $\triangle x^* \neq 0$, there exists a $\tilde{\alpha} \in (0, \bar{\alpha}]$ such that for all $\alpha \in (0, \tilde{\alpha}]$,

$$\Phi_{\mu,\bar{\gamma}}(x^* + \alpha\triangle x^*) - \Phi_{\mu,\bar{\gamma}}(x^*) \leq 2\sigma\alpha\Phi'_{\mu,\bar{\gamma}}(x^*; \triangle x^*) \tag{3.27}$$

as $\sigma \in (0, \frac{1}{2})$.

Suppose there exists an infinite index set $\bar{\mathcal{K}} \subseteq \mathcal{K}$ such that $\{\alpha_x^k\}_{\bar{\mathcal{K}}} \to 0$. Consider an infinite index set $\hat{\mathcal{K}} \subseteq \bar{\mathcal{K}}$ such that $\{\bar{\alpha}_x^k\}_{\hat{\mathcal{K}}} \to \hat{\alpha}$ for some $\hat{\alpha} \in [\bar{\alpha}, 1]$. Let $j_* = \min\{j | \hat{\alpha}\beta^j \leq \tilde{\alpha}\}$. Since (3.27) is valid for $\alpha = \hat{\alpha}\beta^{j_*}$ and $\Phi_{\mu,\bar{\gamma}}(x)$ is continuous, we know from (3.19) and (3.27) that for $\alpha_*^k = \bar{\alpha}_x^k\beta^{j_*}$ and $k \in \hat{\mathcal{K}}$ large enough,

$$\Phi_{\mu,\bar{\gamma}}(x^k + \alpha_*^k\triangle x^k) - \Phi_{\mu,\bar{\gamma}}(x^k) \leq \sigma\alpha_*^k\Phi'_{\mu,\bar{\gamma}}(x^k; \triangle x^k).$$

Therefore, we have $\alpha_x^k \geq \alpha_*^k$ for all $k \in \hat{\mathcal{K}}$ large enough according to the backtracking line search procedure of Algorithm 2.1. However, this contradicts our assumption that $\{\alpha_x^k\}_{\bar{\mathcal{K}}} \to 0$ as $\{\alpha_*^k\}_{\hat{\mathcal{K}}} \to \hat{\alpha}\beta^{j_*}$.

Therefore, there exists a $\alpha_* > 0$ such that $\alpha_x^k \geq \alpha_*$ for all iterations $k \in \bar{\mathcal{K}}$ at which second-order correction steps are not used to update the iterates. If there is an infinite index set $\hat{\mathcal{K}} \subseteq \bar{\mathcal{K}}$ such that $\triangle\hat{x}^k$ is used to produce trial points, due to the boundedness of $\{\triangle\hat{x}^k\}_{\hat{\mathcal{K}}}$ and the fact that $x^* > 0$, we know from (2.15) that $\hat{\alpha}_x^k \geq \hat{\alpha}_*$ for some $\hat{\alpha}_* > 0$ and all $k \in \hat{\mathcal{K}}$ large enough. Since $p_k = 0$ and $\gamma_k = \bar{\gamma}$ for all large k, the iterates x^k are eventually accepted by the Armijo condition (2.10) or (2.18) and the sequence $\{\Phi_{\mu,\gamma_k}(x^k)\}$ is eventually non-increasing. Consequently, we have from (2.10), (2.18) and (3.26) that for all $k \in \bar{\mathcal{K}}$ large enough

$$\Phi_{\mu,\bar{\gamma}}(x^{k+1}) \leq \Phi_{\mu,\bar{\gamma}}(x^k) - \sigma\zeta\nu \min\{\alpha_*, \hat{\alpha}_*\}\|\triangle x^k\|^2.$$

This implies $\triangle x^* = 0$ as $\{\Phi_{\mu,\bar{\gamma}}(x^k)\}$ is bounded below by Assumption A1. We obtain a contradiction. □

The next result follows immediately from Lemmas 5.44 and 3.7.

LEMMA 3.8. *Suppose Assumption A holds and $\gamma_k = \bar{\gamma}$ and $p_k = 0$ for all k large enough. Then the sequence $\{x^k, \lambda^k, y^k\}$ is bounded and any accumulation point of it satisfies the first-order optimality conditions (1.3).*

Combining the results of Theorem 3.1, Lemma 3.6 and Lemma 3.8, we are now in a position to state the global convergence properties of Algorithm 2.1 when an infinite sequence of iterates is generated.

THEOREM 3.9. *Suppose Assumption A holds and Algorithm 2.1 generates an infinite sequence of iterates. Then*

(i) if the penalty parameter $\{\gamma_k\}$ tends to infinity, there exists an accumulation point of $\{x^k\}$ that is either a Fritz-John point of problem (1.1) at which the MFCQ fails to hold or a KKT point of the feasibility problem $\min_{x\geq 0} \|c(x)\|^2$ that is infeasible to problem (1.1);

(ii) if $\{\gamma_k\}$ is bounded and $\{\|\mathcal{M}_k^{-1}\|\}$ is unbounded, there is an accumulation point of the sequence $\{x^k\}$ that is a Fritz-John point of problem (1.1) at which the MFCQ fails to hold;

(iii) if $\{\gamma_k\}$ and $\{\|\mathcal{M}_k^{-1}\|\}$ are bounded, there is an accumulation point of the sequence $\{x^k, \lambda^k, y^k\}$ that satisfies the first-order optimality conditions (1.3).

4. Solving problem (1.1). Our overall interior-point algorithm for solving problem (1.1) successively solves the barrier problem (1.2) for a decreasing sequence $\{\mu\}$ by applying Algorithm 2.1. To prove our fast local convergence results, we require the sequence $\{\mu\}$ to converge to zero at a superlinear rate. The tolerance ϵ_μ in Step 1 of Algorithm 2.1, which determines the accuracy in the solution of the barrier problems, is also decreased from one barrier problem to the next and needs to converge to zero.

We are ready to state our overall algorithm, in which the index j denotes an outer iteration, while k denotes the last inner iteration of Algorithm 2.1.

ALGORITHM 4.1. OUTER ALGORITHM FOR SOLVING PROBLEM (1.1)

Step 0. Initialization.

Parameters: $\mu_0 > 0$, $\epsilon_{\mu_0} > 0$, $\tau_{\mu_0} \in (0, 1)$, $\epsilon_{\text{tol}} > 0$.

Data: $(x^0, \lambda^0, y^0) \in \Re^{2n+m}$ with $x^0 > 0$ and $\lambda^0 > 0$, $\gamma_0 > 0$, $\pi_0 > 0$.

Set $j \leftarrow 0$.

Step 1. Check convergence.

If $\|\mathcal{R}_0(x^j, \lambda^j, y^j)\| \leq \epsilon_{\text{tol}}$, STOP with SUCCESS.

Step 2. Solve problem (1.2) for μ_j.

Apply Algorithm 2.1, starting from $(x^j, \lambda^j, y^j, \gamma_j, \pi_j)$ with parameters

$(\mu_j, \epsilon_{\mu_j}, \tau_{\mu_j})$, to find an approximate solution $(x^{j,k}, \lambda^{j,k}, y^{j,k})$ of problem (1.2), which satisfies $\|\mathcal{R}_{\mu_j}(x^{j,k}, \lambda^{j,k}, y^{j,k})\| \le \epsilon_{\mu_j}$.

Step 3. Update.

Decrease μ_j and ϵ_{μ_j} to μ_{j+1} and $\epsilon_{\mu_{j+1}}$, respectively. Choose $\tau_{\mu_{j+1}} \in (\tau_{\mu_j}, 1)$.
Set $x^{j+1} \leftarrow x^{j,k}$, $\lambda^{j+1} \leftarrow \lambda^{j,k}$, $y^{j+1} \leftarrow y^{j,k}$, $\gamma_{j+1} \leftarrow \gamma_{j,k}$, $\pi_{j+1} \leftarrow \pi_{j,k}$.
Set $j \leftarrow j+1$ and go to Step 1.

The next result, which is an analog of Theorem 3.13 in [4], gives the global convergence of Algorithm 4.1.

THEOREM 4.1. *Suppose Assumption A1 holds with the same bound for every μ_j, A2 holds for each μ_j and Algorithm 4.1 is applied by ignoring its termination criterion. Suppose Algorithm 4.1 successfully terminates for each μ_j and $\{\mu_j\} \to 0$ and $\{\epsilon_{\mu_j}\} \to 0$. If the sequence $\{x^j, \lambda^j, y^j\}_{\mathcal{J}}$ is bounded, where $\mathcal{J}$ is an infinite index set, any accumulation point of $\{x^j, \lambda^j, y^j\}_{\mathcal{J}}$ satisfies the first-order optimality conditions (1.3) with $\mu = 0$. If $\{\|(\lambda^j, y^j)\|\}_{\mathcal{J}} \to \infty$, there exists an accumulation point of $\{x^j\}_{\mathcal{J}}$ that is Fritz-John point of problem (1.1) failing to satisfy the MFCQ.*

By Theorems 3.9 and 4.1, Algorithm 2.1 always terminates successfully unless the iterates (including the inner iterates) converge to an infeasible KKT point of the feasibility problem $\min_{x\ge 0} \|c(x)\|^2$, indicating that problem (1.1) is locally infeasible, or a Fritz-John point of problem (1.1) that fails to satisfy the MFCQ, indicating that locally there may be no feasible KKT points.

5. Fast local convergence. By Theorem 4.1, if Algorithm 4.1 successfully terminates for every μ_j and if there is an infinite index set $\mathcal{J}$ such that the sequence $\{x^j, \lambda^j, y^j\}_{\mathcal{J}}$ is bounded, there is an accumulation point (x^*, λ^*, y^*) of $\{x^j, \lambda^j, y^j\}_{\mathcal{J}}$ that satisfies the first-order optimality conditions for problem (1.1), i.e., x^* is a KKT point of (1.1) and (λ^*, y^*) is an associated multiplier vector. In this section we show that our interior point method also enjoys fast local convergence provided the following standard nondegeneracy conditions hold at (x^*, λ^*, y^*).

Assumption B.

B1. *The Hessian matrices $\nabla^2 f(x)$ and $\nabla^2 c_i(x)$, $i = 1, ..., m$ are locally Lipschitz continuous at x^*.*

B2. *The linear independence constraint qualification (LICQ) holds: the active constraint gradients $\nabla c_i(x)$, $i = 1, ..., m$ and e_i, $i \in \mathcal{B} = \{i = 1, ..., n | x_i^* = 0\}$ are linearly independent, where e_i is the n-dimensional gradient of x_i, i.e., e_i is the i-th column of the $n \times n$ identity matrix.*

B3. *The second-order sufficiency conditions (SOSC) hold: there exists a $\bar{\nu} > 0$ such that*

$$d^\top \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*, y^*) d \ge \bar{\nu} \|d\|^2$$

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

B4. *Strict complementarity holds: $x^* + \lambda^* > 0$.*

Note that under the LICQ, the multiplier vector (λ^*, y^*) at x^* is unique. We first study the local convergence properties of Algorithm 2.1 for a fixed μ.

5.1. Local analysis for a fixed μ. It is well know that under Assumption B, if μ is sufficiently small, the nonlinear system of equations (1.3) has a unique solution $z(\mu) = (x(\mu), \lambda(\mu), y(\mu))$ in a small neighborhood of $z^* = (x^*, \lambda^*, y^*)$ 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 $\|z(\mu) - z^*\| \le \bar{C}\mu$ for all μ small enough. By Theorem 3.1, the sequence of penalty parameters $\{\gamma_k\}$ is bounded unless the iterates converge to a stationary point of the feasibility problem $\min_{x\ge 0} \|c(x)\|^2$ or a Fritz-John point of problem (1.1) at which the MFCQ fails to hold. Therefore, to analyze the rate of convergence of Algorithm 2.1 to a KKT solution of problem (1.2), we can assume that $\gamma_k = \bar{\gamma}$ for all k large enough.

Before we establish the fast local convergence of Algorithm 2.1, we need a number of results that study the local behavior of the step direction $(\triangle\tilde{x}^k, \triangle\tilde{\lambda}^k, \triangle\tilde{y}^k)$ and the second-order correction step $\triangle\hat{x}^k$. To simplify notation, we discard the iteration index k in the following analysis.

Consider the following linear system, which is equivalent to (2.1) and (2.2),

$$\mathcal{W}_{\bar{\gamma}}(x,\lambda,y)\begin{bmatrix}\triangle\tilde{x}\\ \triangle\tilde{\lambda}\\ \triangle\tilde{y}\end{bmatrix}=\begin{bmatrix}-\nabla_x\mathcal{L}(x,\lambda,y)\\ \mu e-\Lambda x\\ -c(x)\end{bmatrix}, \tag{5.1}$$

where

$$\mathcal{W}_{\bar{\gamma}}(x,\lambda,y)=\begin{bmatrix}\mathcal{H} & -I & \nabla c(x)\\ \Lambda & X & 0\\ \nabla c(x)^\top & 0 & -\frac{\|c(x)\|}{\bar{\gamma}}I\end{bmatrix}. \tag{5.2}$$

If in (2.6) the step size $\bar{\alpha}_x=1$, we have from (2.14) that the second-order correction step $\triangle\hat{x}$ can be obtained from

$$\mathcal{W}_{\bar{\gamma}}(x,\lambda,y)\begin{bmatrix}\triangle\hat{x}\\ \triangle\hat{\lambda}\\ \triangle\hat{y}\end{bmatrix}=\begin{bmatrix}-\nabla_x\mathcal{L}(x,\lambda,y)\\ \mu e-\Lambda x\\ -c(x+\triangle\tilde{x})+\nabla c(x)^\top\triangle\tilde{x}\end{bmatrix}. \tag{5.3}$$

For simplicity we use in the remainder of this section the notation

$$z=(x,\lambda,y),\ \triangle\tilde{z}=(\triangle\tilde{x},\triangle\tilde{\lambda},\triangle\tilde{y}),\ \triangle\hat{z}=(\triangle\hat{x},\triangle\hat{\lambda},\triangle\hat{y}).$$

Since the SOSC holds at z^*, condition (2.5) holds with the exact Hessian if z is sufficiently close to z^* and if ν in (2.5) is chosen small enough so that $\nu<\bar{\nu}$, where $\bar{\nu}$ is defined in Assumption B3. Therefore, by Step 2.1 of Algorithm 2.1, we have

$$\mathcal{H}=\nabla^2_{xx}\mathcal{L}(z), \tag{5.4}$$

if μ is sufficiently small and if z is sufficiently close to $z(\mu)$. Moreover, by the nondegeneracy assumptions B2, B3 and B4, the matrix $\mathcal{W}(z)$ is uniformly nonsingular for any z sufficiently close to z^*. Specifically, there exist a constant $W>0$ and a neighborhood $\mathcal{N}(z^*)$ of z^* such that

$$\|\mathcal{W}_r(z)^{-1}\|\le W,\quad \forall z\in\mathcal{N}(z^*). \tag{5.5}$$

The next theorem shows that $z+\triangle\tilde{z}$ and $z+\triangle\hat{z}$ converge quadratically to $z(\mu)$.

THEOREM 5.1. *Suppose Assumption B holds. If $z\in\mathcal{N}(z^*)$, then*
(i) $\|z+\triangle\tilde{z}-z(\mu)\|=\mathcal{O}\left(\|z-z(\mu)\|^2\right)$ and $\|z+\triangle\hat{z}-z(\mu)\|=\mathcal{O}\left(\|z-z(\mu)\|^2\right)$;
(ii)$\|\triangle\tilde{z}\|=\Omega(\|z-z(\mu)\|)$ and $\|\triangle\hat{z}\|=\Omega(\|z-z(\mu)\|)$.

Proof. We have from (5.1) and (1.3) that

$$\mathcal{W}_{\bar{\gamma}}(z)(z+\triangle\tilde{z}-z(\mu))=\begin{bmatrix}\mathcal{H}(x-x(\mu))-\nabla_x\mathcal{L}(x,\lambda(\mu),y(\mu))\\ \Lambda(x-x(\mu))-X\lambda(\mu)+\mu e\\ \frac{\|c(x)\|}{\bar{\gamma}}(y(\mu)-y)-c(x)+\nabla c(x)^\top(x-x(\mu))\end{bmatrix}. \tag{5.6}$$

For the right hand of (5.6), we have from (5.4), Assumption B1 and Taylor's Theorem that

$$\begin{aligned}&\mathcal{H}(x-x(\mu))-\nabla_x\mathcal{L}(x,\lambda(\mu),y(\mu))\\ =\ &(\nabla^2_{xx}\mathcal{L}(z)-\nabla^2_{xx}\mathcal{L}(z(\mu)))(x-x(\mu))-\nabla_x\mathcal{L}(z(\mu))+\mathcal{O}(\|x-x(\mu)\|^2)\\ =\ &\mathcal{O}(\|z-z(\mu)\|^2),\end{aligned} \tag{5.7}$$

$$\begin{aligned}&\Lambda(x-x(\mu))-X\lambda(\mu)+\mu e\\ =\ &\Lambda(x-x(\mu))-\Lambda(\mu)(x-x(\mu))\\ =\ &(\Lambda-\Lambda(\mu))(x-x(\mu))=\mathcal{O}(\|z-z(\mu)\|^2),\end{aligned} \tag{5.8}$$

where $\Lambda(\mu) = \mathrm{diag}(\lambda(\mu))$, and

$$\frac{\|c(x)\|}{\bar{\gamma}}(y(\mu) - y) - c(x) + \nabla c(x)^\top (x - x(\mu)) = \mathcal{O}(\|z - z(\mu)\|^2). \tag{5.9}$$

Since $z \in \mathcal{N}(z^*)$, the quadratic convergence of $z + \triangle\tilde{z}$ to $z(\mu)$ follows from (5.5). Also by (5.5), subtracting (5.1) from (5.3) and using Taylor's theorem yields

$$\|\triangle\tilde{z} - \triangle\hat{z}\| = \mathcal{O}\left(\|\triangle\tilde{x}\|^2\right). \tag{5.10}$$

This implies the quadratic convergence of $z + \triangle\hat{z}$ to $z(\mu)$. Result (ii) follows immediately from result (i). □

The strict complementarity assumption B4 implies that

$$\begin{cases} \lambda_i(\mu) = \Omega(1), & x_i(\mu) = \Omega(\mu), \quad \forall i \in \mathcal{B}, \\ \lambda_i(\mu) = \Omega(\mu), & x_i(\mu) = \Omega(1), \quad \forall i \in \mathcal{I}\backslash\mathcal{B}, \end{cases} \tag{5.11}$$

where $\mathcal{I} = \{1, ..., n\}$. Hence, if $\|z - z(\mu)\| = o(\mu)$, we have

$$\begin{cases} \lambda_i = \Omega(1), & x_i = \Omega(\mu), \quad \forall i \in \mathcal{B}, \\ \lambda_i = \Omega(\mu), & x_i = \Omega(1), \quad \forall i \in \mathcal{I}\backslash\mathcal{B}. \end{cases} \tag{5.12}$$

Lemma 5.2. *Suppose Assumption A holds. If μ is sufficiently small and $\|z - z(\mu)\| = o(\mu)$, then for $\tau_\mu \in (0, 1)$ we have (i) $x + \triangle\tilde{x} \geq (1 - \tau_\mu)x$, (ii) $x + \triangle\hat{x} \geq (1 - \tau_\mu)x$, (iii) $\lambda + \triangle\tilde{\lambda} \geq (1 - \tau_\mu)\lambda$, and (iv)*

$$\lambda + \triangle\tilde{\lambda} \leq \min\{\varrho\mu(\mathrm{diag}(x + \triangle\tilde{x}))^{-1}e, \varrho\mu(\mathrm{diag}(x + \triangle\hat{x}))^{-1}\},$$

where $\varrho > 1$.

Proof. First, we have $z \in \mathcal{N}(z^*)$ since μ is sufficiently small and $\|z - z(\mu)\| = o(\mu)$. It then follows from Theorem 5.1 (ii) that $\|\triangle\tilde{z}\| = o(\mu)$ and $\|\triangle\hat{z}\| = o(\mu)$. Now results (i), (ii) and (iii) follow from (5.12). To prove result (iv), it suffices to show that

$$(\lambda_i + \triangle\tilde{\lambda}_i)(x_i + \triangle\tilde{x}_i) \leq \varrho\mu \text{ and } (\lambda_i + \triangle\tilde{\lambda}_i)(x_i + \triangle\hat{x}_i) \leq \varrho\mu, \ \forall i \in \mathcal{I}. \tag{5.13}$$

From the second equation of (5.1) we have

$$(\lambda_i + \triangle\tilde{\lambda}_i)(x_i + \triangle\tilde{x}_i) = \mu + \triangle\tilde{\lambda}_i\triangle\tilde{x}_i = \mu + o(\mu^2).$$

Hence the first part of (5.13) holds. It follows from (5.10) that

$$(\lambda_i + \triangle\tilde{\lambda}_i)(x_i + \triangle\hat{x}_i) = (\lambda_i + \triangle\tilde{\lambda}_i)(x_i + \triangle\tilde{x}_i) + \mathcal{O}(\|\triangle\tilde{x}\|^2) = \mu + o(\mu^2).$$

Hence the second part of (5.13) holds. □

Lemma 5.3. *Suppose Assumption B holds. Suppose $\bar{\gamma} > 0$ and $\bar{\gamma} \geq \eta\|y(\mu)\|$ for some $\eta > 1$. If μ is sufficiently small and $\|z - z(\mu)\| = o(\mu)$, then*

$$\Phi'_{\mu,\bar{\gamma}}(x; \triangle\tilde{x}) \leq -\zeta\triangle\tilde{x}^\top\bar{\mathcal{H}}\triangle\tilde{x}, \tag{5.14}$$

where $\zeta \in (0, 1 - \frac{1}{\eta})$ is a constant and $\bar{\mathcal{H}}$ is defined by (2.3).

Proof. First note that $\triangle\tilde{x}^\top\bar{\mathcal{H}}\triangle\tilde{x} \geq 0$ by Assumption B3 and (5.4). If $c(x) = 0$, it is easy to verify that $\Phi'_{\mu,\bar{\gamma}}(x; \triangle\tilde{x}) = \nabla\varphi_\mu(x)^\top\triangle\tilde{x}$. See the proof of Lemma 3.1 in [4]. Moreover, if $c(x) = 0$, we have from the third equation of (5.1) that $\nabla c(x)^\top\triangle\tilde{x} = 0$. From the first two equation of (5.1) we have

$$\nabla\varphi_\mu(x) = -(\mathcal{H} + X^{-1}\Lambda)\triangle\tilde{x} + (y + \triangle\tilde{y})\nabla c(x). \tag{5.15}$$

Hence, if $c(x)=0$, we have $\Phi'_{\mu,\bar\gamma}(x;\triangle\tilde x)=-\triangle\tilde x^\top\bar{\mathcal{H}}\triangle\tilde x$ as $\bar{\mathcal{H}}=\mathcal{H}+X^{-1}\Lambda$. Thus, (5.14) holds trivially.

Now suppose $\|c(x)\|>0$. Then it follows from (5.15) and the third equation of (5.1) that

$$\begin{aligned}
&\Phi'_{\mu,\bar\gamma}(x;\triangle\tilde x)=\nabla\Phi_{\mu,r}(x)^\top\triangle\tilde x\\
=\;&\nabla\varphi_\mu(x)^\top\triangle\tilde x+\tfrac{\bar\gamma}{\|c(x)\|}c(x)^\top\nabla c(x)^\top\triangle\tilde x\\
=\;&-\triangle\tilde x^\top(\mathcal{H}+X^{-1}\Lambda)\triangle\tilde x-y^\top\nabla c(x)^\top\triangle\tilde x\\
&-\left(\triangle\tilde y-\tfrac{\bar\gamma}{\|c(x)\|}c(x)\right)^\top\nabla c(x)^\top\triangle\tilde x\\
=\;&-\triangle\tilde x^\top(\mathcal{H}+X^{-1}\Lambda)\triangle\tilde x-y^\top\nabla c(x)^\top\triangle\tilde x\\
&-\tfrac{\bar\gamma}{\|c(x)\|}\triangle\tilde x^\top(\nabla c(x)\nabla c(x)^\top)\triangle\tilde x\\
=\;&-\triangle\tilde x^\top\bar{\mathcal{H}}\triangle\tilde x-y^\top\nabla c(x)^\top\triangle\tilde x.
\end{aligned}\tag{5.16}$$

To prove (5.14), it suffices to show that

$$(1-\zeta)\triangle\tilde x^\top\bar{\mathcal{H}}\triangle\tilde x+y^\top\nabla c(x)^\top\triangle\tilde x\ge 0.$$

Since $\zeta\in(0,1-\frac{1}{\eta})$, there is a constant $\bar\zeta$ such that $\bar\zeta\in(\frac{1}{\eta(1-\zeta)},1)$. Moreover, if μ is small enough, we have $\bar\gamma>\tilde\zeta\eta\|y\|$ for some $\tilde\zeta\in(\frac{1}{\eta\bar\zeta(1-\zeta)},1)$ as $\|z-z(\mu)\|=o(\mu)$, $\bar\gamma>0$ and $\bar\gamma\ge\eta\|y(\mu)\|$. By Assumption B3, for small enough $\|c(x)\|$, we have

$$\mathcal{H}+X^{-1}\Lambda+\frac{\bar\gamma(1-\bar\zeta)}{\|c(x)\|}\nabla c(x)\nabla c(x)^\top\succ 0,\quad\text{if}\quad\|c(x)\|>0.\tag{5.17}$$

Now we can derive from (5.1) that if $\|c(x)\|>0$,

$$\begin{aligned}
&(1-\zeta)\triangle\tilde x^\top\bar{\mathcal{H}}\triangle\tilde x+y^\top\nabla c(x)^\top\triangle\tilde x\\
=\;&(1-\zeta)\triangle\tilde x^\top\left(\mathcal{H}+X^{-1}\Lambda+\tfrac{\bar\gamma(1-\bar\zeta)}{\|c(x)\|}\nabla c(x)\nabla c(x)^\top\right)\triangle\tilde x\\
&+(1-\zeta)\tfrac{\bar\gamma\bar\zeta}{\|c(x)\|}\triangle\tilde x^\top\nabla c(x)\nabla c(x)^\top\triangle\tilde x+y^\top\nabla c(x)^\top\triangle\tilde x\\
\ge\;&(1-\zeta)\tfrac{\bar\gamma\bar\zeta}{\|c(x)\|}\triangle\tilde x^\top\nabla c(x)\nabla c(x)^\top\triangle\tilde x+y^\top\nabla c(x)^\top\triangle\tilde x\\
=\;&(1-\zeta)\tfrac{\bar\gamma\bar\zeta}{\|c(x)\|}\left\|c(x)-\tfrac{\|c(x)\|}{\bar\gamma}\triangle\tilde y\right\|^2-y^\top\left(c(x)-\tfrac{\|c(x)\|}{\bar\gamma}\triangle\tilde y\right)\\
=\;&(1-\zeta)\bar\gamma\bar\zeta\left(\|c(x)\|-\tfrac{2}{\bar\gamma}c(x)^\top\triangle\tilde y+\tfrac{1}{\bar\gamma^2}\|c(x)\|\|\triangle\tilde y\|^2\right)-y^\top c(x)+\tfrac{\|c(x)\|}{\bar\gamma}y^\top\triangle\tilde y\\
=\;&(1-\zeta)\bar\gamma\bar\zeta\|c(x)\|-y^\top c(x)+o\left(\|c(x)\|\right)\\
\ge\;&(\bar\gamma-\tilde\zeta\eta\|y\|+\tilde\zeta\eta\|y\|)\bar\zeta(1-\zeta)\|c(x)\|-\|y\|\|c(x)\|+o(\|c(x)\|)\\
=\;&(\tilde\zeta\eta\bar\zeta(1-\zeta)-1)\|y\|\|c(x)\|+(\bar\gamma-\tilde\zeta\eta\|y\|)\bar\zeta(1-\zeta)\|c(x)\|+o(\|c(x)\|)\\
\ge\;&0,
\end{aligned}$$

where the fourth equality uses the fact that $\|y\|=\mathcal{O}(1)$ and $\|\triangle\tilde y\|=o(1)$, and the last inequality follows since $\tilde\zeta\eta\bar\zeta(1-\zeta)>1$ and $\bar\gamma>\tilde\zeta\eta\|y\|$. Hence, the lemma follows. □

THEOREM 5.4. *Suppose Assumption B holds. Suppose $\sigma\in(0,\frac{1}{2})$, $\bar\gamma>0$ and $\bar\gamma\ge\eta\|y(\mu)\|$ for some $\eta>1$. If μ is sufficiently small and $\|z-z(\mu)\|=o(\mu)$, then*

$$\Phi_{\mu,\bar\gamma}(x+\triangle\hat x)-\Phi_{\mu,\bar\gamma}(x)\le\sigma\Phi'_{\mu,\bar\gamma}(x;\triangle\hat x).\tag{5.18}$$

Proof. By (5.12) and Lemma 5.2, we have $x > 0$ and $x + \triangle\hat{x} > 0$. Then by Taylor's theorem we have

$$\varphi_\mu(x+\triangle\hat{x}) = \varphi_\mu(x) + \nabla\varphi_\mu(x)^\top\triangle\hat{x} + \frac{1}{2}\triangle\hat{x}^\top\nabla^2 f(x+\bar{\xi}\triangle\hat{x})\triangle\hat{x} + \frac{\mu}{2}\triangle\hat{x}^\top(\mathrm{diag}(x+\bar{\xi}\triangle\hat{x}))^{-2}\triangle\hat{x} \tag{5.19}$$

for some $\bar{\xi} \in [0,1]$. Lemma 5.1 implies that $\|\triangle\hat{x}\| = \mathcal{O}(\|z - z(\mu)\|) = o(\mu)$. This together with (5.12) give that $\frac{\|\triangle\hat{x}\|}{x_i} = o(1)$, $i \in \mathcal{I}$. Hence, we have

$$\frac{1}{x_i + \bar{\xi}\triangle\hat{x}_i} = \frac{1}{x_i}\left(1 - \frac{\bar{\xi}\triangle\hat{x}_i/x_i}{1+\bar{\xi}\triangle\hat{x}_i/x_i}\right) = \frac{1}{x_i} + \mathcal{O}\left(\frac{\|\triangle\hat{x}\|}{x_i^2}\right), \forall\ i \in \mathcal{I}. \tag{5.20}$$

This together with (5.10) give the following estimates,

$$\begin{aligned}\triangle\hat{x}^\top\nabla^2 f(x+\bar{\xi}\triangle\hat{x})\triangle\hat{x} &= \triangle\hat{x}^\top\nabla^2 f(x)\triangle\hat{x} + \mathcal{O}\left(\|\triangle\hat{x}\|^3\right)\\ &= \triangle\tilde{x}^\top\nabla^2 f(x)\triangle\tilde{x} + \mathcal{O}\left(\|\triangle\tilde{x}\|^3\right),\end{aligned} \tag{5.21}$$

$$\begin{aligned}\frac{\triangle\hat{x}_i}{x_i+\bar{\xi}\triangle\hat{x}_i} &= \triangle\hat{x}_i\left(\frac{1}{x_i} + \mathcal{O}\left(\frac{\|\triangle\hat{x}\|}{x_i^2}\right)\right)\\ &= \triangle\tilde{x}_i\left(\frac{1}{x_i} + \mathcal{O}\left(\frac{\|\triangle\tilde{x}\|}{x_i^2}\right)\right) + \mathcal{O}\left(\frac{\|\triangle\tilde{x}\|^2}{x_i}\right), \quad \forall i \in \mathcal{I}.\end{aligned} \tag{5.22}$$

From (5.22), (5.12) and the fact that $\|\triangle\tilde{x}\| = o(\mu)$, we have for $i \in \mathcal{I}$ that

$$\begin{aligned}\left(\frac{\triangle\hat{x}_i}{x_i+\bar{\xi}\triangle\hat{x}_i}\right)^2 &= \left(\frac{\triangle\tilde{x}_i}{x_i}\right)^2 + \mathcal{O}\left(\frac{\triangle\tilde{x}_i^2\|\triangle\tilde{x}\|}{x_i^3}\right) + \mathcal{O}\left(\frac{\triangle\tilde{x}_i^2\|\triangle\tilde{x}\|^2}{x_i^4}\right)\\ &\quad + \mathcal{O}\left(\frac{\triangle\tilde{x}_i\|\triangle\tilde{x}\|^2}{x_i^2}\right) + \mathcal{O}\left(\frac{\triangle\tilde{x}_i\|\triangle\tilde{x}\|^3}{x_i^3}\right) + \mathcal{O}\left(\frac{\|\triangle\tilde{x}\|^4}{x_i^2}\right)\\ &= \left(\frac{\triangle\tilde{x}_i}{x_i}\right)^2 + \triangle\tilde{x}_i^2\left(\mathcal{O}\left(\frac{\|\triangle\tilde{x}\|}{x_i^3}\right) + \mathcal{O}\left(\frac{\|\triangle\tilde{x}\|^2}{x_i^4}\right)\right)\\ &\quad + \mathcal{O}\left(\frac{\|\triangle\tilde{x}\|^3}{x_i^2}\right) + \mathcal{O}\left(\frac{\|\triangle\tilde{x}\|^4}{x_i^3}\right) + \mathcal{O}\left(\frac{\|\triangle\tilde{x}\|^4}{x_i^2}\right)\\ &= \left(\frac{\triangle\tilde{x}_i}{x_i}\right)^2 + \triangle\tilde{x}_i^2\left(\mathcal{O}\left(\frac{\|\triangle\tilde{x}\|}{x_i^3}\right)\right) + \mathcal{O}\left(\frac{\|\triangle\tilde{x}\|^3}{x_i^2}\right).\end{aligned} \tag{5.23}$$

Combining (5.19), (5.21), (5.23), (5.10) and (5.12), we obtain

$$\begin{aligned}\varphi_\mu(x+\triangle\hat{x}) - \varphi_\mu(x) &= \nabla\varphi_\mu(x)^\top\triangle\hat{x} + \frac{1}{2}\triangle\tilde{x}^\top\nabla^2 f(x)\triangle\tilde{x} + \frac{\mu}{2}\textstyle\sum_{i\in\mathcal{I}}\left(\frac{\triangle\tilde{x}_i}{x_i}\right)^2\\ &\quad + \mathcal{O}\left(\mu^{-2}\|\triangle\tilde{x}\|\right)\textstyle\sum_{i\in\mathcal{I}}\triangle\tilde{x}_i^2 + \mathcal{O}\left(\mu^{-1}\|\triangle\tilde{x}\|^3\right).\end{aligned} \tag{5.24}$$

From (5.10) and the third equation of (5.3) we have

$$\begin{aligned}& c(x+\triangle\hat{x})\\ =\ & c(x+\triangle\tilde{x}) + \nabla c(x+\triangle\hat{x})^\top(\triangle\hat{x}-\triangle\tilde{x}) + \mathcal{O}\left(\|\triangle\tilde{x}\|^4\right)\\ =\ & c(x+\triangle\tilde{x}) + \nabla c(x)^\top(\triangle\hat{x}-\triangle\tilde{x}) + \mathcal{O}\left(\|\triangle\tilde{x}\|^3\right)\\ =\ & \tfrac{\|c(x)\|}{\bar{\gamma}}\triangle\hat{y} + \mathcal{O}\left(\|\triangle\tilde{x}\|^3\right).\end{aligned} \tag{5.25}$$

By the third equation of (5.1) we have

$$\|c(x)\| = \left\|\nabla c(x)^\top \triangle\tilde{x} - \frac{\|c(x)\|}{\bar{\gamma}}\triangle\tilde{y}\right\| \le \|\nabla c(x)^\top \triangle\tilde{x}\| + \frac{\|c(x)\|\|\triangle\tilde{y}\|}{\bar{\gamma}}. \tag{5.26}$$

Hence it follows from $\|\triangle\tilde{y}\| = o(1)$ that $\|c(x)\| = \mathcal{O}(\|\triangle\tilde{x}\|)$. This together with (5.1) and (5.10) give

$$\begin{aligned} & c_i(x+\triangle\hat{x}) \\ = \; & \left(c_i(x) + \nabla c_i(x)^\top \triangle\tilde{x}\right) + \nabla c_i(x)^\top\left(\triangle\hat{x} - \triangle\tilde{x}\right) + \tfrac{1}{2}(\triangle\hat{x})^\top \nabla^2 c_i(x)\triangle\hat{x} + \mathcal{O}\left(\|\triangle\hat{x}\|^3\right) \\ = \; & \tfrac{\|c(x)\|}{\bar{\gamma}}\triangle\tilde{y} + \nabla c_i(x)^\top(\triangle\hat{x} - \triangle\tilde{x}) + \tfrac{1}{2}(\triangle\tilde{x})^\top\nabla^2 c_i(x)\triangle\tilde{x} + \mathcal{O}\left(\|\triangle\tilde{x}\|^3\right) \\ = \; & \tfrac{\|c(x)\|}{\bar{\gamma}}\triangle\hat{y} + \nabla c_i(x)^\top(\triangle\hat{x} - \triangle\tilde{x}) + \tfrac{1}{2}(\triangle\tilde{x})^\top\nabla^2 c_i(x)\triangle\tilde{x} + \mathcal{O}\left(\|\triangle\tilde{x}\|^3\right), \quad \forall i = 1, ..., m. \end{aligned} \tag{5.27}$$

Comparing the last equalities of (5.25) and (5.27), we obtain that

$$\nabla c_i(x)^\top(\triangle\tilde{x} - \triangle\hat{x}) = \frac{1}{2}(\triangle\tilde{x})^\top\nabla^2 c_i(x)\triangle\tilde{x} + \mathcal{O}\left(\|\triangle\tilde{x}\|^3\right), \quad \forall i = 1, ..., m. \tag{5.28}$$

From (5.15), the third equation of (5.1), (5.10), (5.12) and (5.28), we have for any constant $\bar{\sigma}$ that

$$\begin{aligned} & \nabla\varphi_\mu(x)^\top\triangle\hat{x} = \nabla\varphi_\mu(x)^\top\triangle\tilde{x} - \nabla\varphi_\mu(x)^\top\left(\triangle\tilde{x} - \triangle\hat{x}\right) \\ = \; & \left(\tfrac{1}{2} - \bar{\sigma}\right)\nabla\varphi_\mu(x)^\top\triangle\tilde{x} - \left(\tfrac{1}{2} + \bar{\sigma}\right)\left(\triangle\tilde{x}^\top(\mathcal{H} + X^{-1}\Lambda) + (y + \triangle\tilde{y})^\top\nabla c(x)^\top\right)\triangle\tilde{x} \\ & + \left(\triangle\tilde{x}^\top(\mathcal{H} + X^{-1}\Lambda) + (y + \triangle\tilde{y})^\top\nabla c(x)^\top\right)\left(\triangle\tilde{x} - \triangle\hat{x}\right) \\ = \; & \left(\tfrac{1}{2} - \bar{\sigma}\right)\nabla\varphi_\mu(x)^\top\triangle\tilde{x} - \left(\tfrac{1}{2} + \bar{\sigma}\right)\left(\triangle\tilde{x}^\top(\mathcal{H} + X^{-1}\Lambda) + (y + \triangle\tilde{y})^\top\nabla c(x)^\top\right)\triangle\tilde{x} \\ & + \tfrac{1}{2}\triangle\tilde{x}^\top\textstyle\sum_{i=1}^m (y_i + \triangle\tilde{y}_i)\nabla^2 c_i(x)\triangle\tilde{x} + \mathcal{O}\left(\mu^{-1}\|\triangle\tilde{x}\|^3\right). \end{aligned} \tag{5.29}$$

If $\|c(x)\| > 0$, we have from the third equation of (5.1) that

$$\begin{aligned} & \left(\tfrac{1}{2} + \bar{\sigma}\right)(y + \triangle\tilde{y})^\top\nabla c(x)^\top\triangle\tilde{x} \\ = \; & \tfrac{1}{2}(y + \triangle\tilde{y})^\top\left(\tfrac{\|c(x)\|}{\bar{\gamma}}\triangle\tilde{y} - c(x)\right) \\ & + \tfrac{\bar{\sigma}\bar{\gamma}}{\|c(x)\|}\left(\nabla c(x)^\top\triangle\tilde{x} + c(x) + \tfrac{\|c(x)\|}{\bar{\gamma}}y\right)^\top\nabla c(x)^\top\triangle\tilde{x} \\ = \; & -\tfrac{1}{2}c(x)^\top(y + \triangle\tilde{y}) + \tfrac{\bar{\sigma}\bar{\gamma}}{\|c(x)\|}\left\|\nabla c(x)^\top\triangle\tilde{x}\right\|^2 \\ & + \tfrac{\bar{\sigma}\bar{\gamma}}{\|c(x)\|}\left(c(x) + \tfrac{\|c(x)\|}{\bar{\gamma}}y\right)^\top\left(\tfrac{\|c(x)\|}{\bar{\gamma}}\triangle\tilde{y} - c(x)\right) + o(\|c(x)\|) \\ = \; & -\tfrac{1}{2}c(x)^\top(y + \triangle\tilde{y}) + \tfrac{\bar{\sigma}\bar{\gamma}}{\|c(x)\|}\left\|\nabla c(x)^\top\triangle\tilde{x}\right\|^2 \\ & -\bar{\sigma}\bar{\gamma}\|c(x)\| - \bar{\sigma}c(x)^\top y + o(\|c(x)\|), \end{aligned} \tag{5.30}$$

where we have used the facts that $\|y\| = \mathcal{O}(1)$, $\|\triangle\tilde{y}\| = o(\mu)$ and $\|c(x)\| = \mathcal{O}(\|\triangle\tilde{x}\|) = o(\mu)$. Since $\|z - z(\mu)\| = o(\mu)$ and strict complementarity holds, if μ is sufficiently small, then $\frac{\lambda_i}{2x_i}$ is sufficiently large for all $i \in \mathcal{B}$, and moreover, if $\|c(x)\| > 0$, $\frac{\bar{\gamma}}{\|c(x)\|}$ is also sufficiently large. Hence, we know from Assumption B3 that

$$\mathcal{H} + \sum_{i\in\mathcal{B}}\frac{\lambda_i}{2x_i}e_i e_i^\top + \frac{\bar{\gamma}}{\|c(x)\|}\nabla c(x)\nabla c(x)^\top \succeq 0, \quad \text{if} \quad \|c(x)\| > 0. \tag{5.31}$$

If $\|c(x)\| = 0$, we have from the third equation of (5.1) that $\nabla c(x)^\top\triangle\tilde{x} = 0$ and hence from Assumption B3 that

$$\triangle\tilde{x}^\top\left(\mathcal{H} + \sum_{i\in\mathcal{B}}\frac{\lambda_i}{2x_i}e_i e_i^\top\right)\triangle\tilde{x} \ge 0, \quad \text{if} \quad \|c(x)\| = 0. \tag{5.32}$$

Moreover, from (5.12) and the facts that $\|z - z(\mu)\| = o(\mu)$ and $\lambda_i = \mathcal{O}(1)$ and $x_i = \Omega(\mu)$ for $i \in \mathcal{B}$, we have

$$\begin{aligned} \lambda_i - \frac{\mu}{x_i} &= \frac{1}{x_i}\left(\lambda_i x_i - \lambda_i(\mu) x_i(\mu)\right) \\ &= \frac{1}{x_i}\left(\lambda_i(x_i - x_i(\mu)) + x_i(\mu)(\lambda_i - \lambda_i(\mu))\right) \\ &= \mathcal{O}\left(\frac{\|z - z(\mu)\|}{\mu}\right) = o(1), \quad \forall i \in \mathcal{B}. \end{aligned} \tag{5.33}$$

Therefore, we obtain for any $\bar{\sigma} > 0$ that

$$\left(\frac{1}{2} + \frac{\bar{\sigma}}{2}\right)\frac{\lambda_i}{x_i} - \frac{1}{2}\left(\frac{\mu}{x_i^2} + \mathcal{O}\left(\mu^{-2}\|\triangle\tilde{x}\|\right)\right) > 0, \quad \forall i \in \mathcal{B}. \tag{5.34}$$

Now let us require that $\bar{\sigma} \in \left(0, \min\left\{\frac{1}{2} - \sigma, \frac{\eta - 1}{2}\right\}\right)$. Combining (5.24), (5.29), (5.30) and (5.12), we obtain for the case $\|c(x)\| > 0$ that

$$\begin{aligned} & \varphi_\mu(x + \triangle\hat{x}) - \varphi(x) \\ = \; & \left(\tfrac{1}{2} - \bar{\sigma}\right)\nabla\varphi_\mu(x)^\top\triangle\tilde{x} - \left(\tfrac{1}{2} + \bar{\sigma}\right)\triangle\tilde{x}^\top(\mathcal{H} + X^{-1}\Lambda)\triangle\tilde{x} \\ & + \tfrac{1}{2}c(x)^\top(y + \triangle\tilde{y}) - \tfrac{\bar{\sigma}\bar{\gamma}}{\|c(x)\|}\triangle\tilde{x}^\top\nabla c(x)\nabla c(x)^\top\triangle\tilde{x} + \bar{\sigma}\bar{\gamma}\|c(x)\| + \bar{\sigma}c(x)^\top y \\ & + \tfrac{1}{2}\triangle\tilde{x}^\top\textstyle\sum_{i=1}^m(y_i + \triangle\tilde{y}_i)\nabla^2 c_i(x)\triangle\tilde{x} + \tfrac{1}{2}\triangle\tilde{x}^\top\nabla^2 f(x)\triangle\tilde{x} \\ & + \tfrac{1}{2}\triangle\tilde{x}^\top\textstyle\sum_{i\in\mathcal{B}}\left(\left(\frac{\mu}{x_i^2} + \mathcal{O}\left(\mu^{-2}\|\triangle\tilde{x}\|\right)\right)e_i e_i^\top\right)\triangle\tilde{x} \\ & + o(\|c(x)\|) + \mathcal{O}\left(\mu\|\triangle\tilde{x}\|^2\right) + \mathcal{O}\left(\mu^{-1}\|\triangle\tilde{x}\|^3\right). \end{aligned} \tag{5.35}$$

Decomposing the term $(\frac{1}{2} + \bar{\sigma})(\mathcal{H} + X^{-1}\Lambda)$ in (5.35) as $\bar{\sigma}\left(\mathcal{H} + \frac{1}{2}X^{-1}\Lambda\right) + \frac{1}{2}\mathcal{H} + \left(\frac{1}{2} + \frac{\bar{\sigma}}{2}\right)X^{-1}\Lambda$, and using (5.31), (5.34) and the fact that $\mathcal{H} = \nabla^2 f(x) + \sum_{i=1}^m y_i\nabla^2 c_i(x)$, we obtain

$$\begin{aligned} & \varphi_\mu(x + \triangle\hat{x}) - \varphi(x) \\ = \; & \left(\tfrac{1}{2} - \bar{\sigma}\right)\nabla\varphi_\mu(x)^\top\triangle\tilde{x} + \tfrac{1}{2}c(x)^\top(y + \triangle\tilde{y}) + \bar{\sigma}\bar{\gamma}\|c(x)\| + \bar{\sigma}c(x)^\top y \\ & - \bar{\sigma}\triangle\tilde{x}^\top\left(\mathcal{H} + \textstyle\sum_{i\in\mathcal{B}}\frac{\lambda_i}{2x_i}e_i e_i^\top + \frac{\bar{\gamma}}{\|c(x)\|}\nabla c(x)\nabla c(x)^\top\right)\triangle\tilde{x} \\ & + \tfrac{1}{2}\triangle\tilde{x}^\top\textstyle\sum_{i=1}^m\triangle\tilde{y}_i\nabla^2 c_i(x)\triangle\tilde{x} - \frac{1+2\bar{\sigma}}{2x_i}\sum_{i\in\mathcal{I}\setminus\mathcal{B}}\lambda_i\triangle\tilde{x}_i^2 \\ & - \triangle\tilde{x}^\top\textstyle\sum_{i\in\mathcal{B}}\left(\left(\frac{1}{2} + \frac{\bar{\sigma}}{2}\right)\frac{\lambda_i}{x_i} - \frac{1}{2}\left(\frac{\mu}{x_i^2} + \mathcal{O}\left(\mu^{-2}\|\triangle\tilde{x}\|\right)\right)\right)e_i e_i^\top\triangle\tilde{x} \\ & + o(\|c(x)\|) + \mathcal{O}\left(\mu\|\triangle\tilde{x}\|^2\right) + \mathcal{O}\left(\mu^{-1}\|\triangle\tilde{x}\|^3\right) \\ \leq \; & \left(\tfrac{1}{2} - \bar{\sigma}\right)\nabla\varphi_\mu(x)^\top\triangle\tilde{x} + \tfrac{1}{2}c(x)^\top(y + \triangle\tilde{y}) + \bar{\sigma}\bar{\gamma}\|c(x)\| + \bar{\sigma}c(x)^\top y + o(\|c(x)\|) + o\left(\|\triangle\tilde{x}\|^2\right). \end{aligned} \tag{5.36}$$

Similarly to (3.25), we have from (5.1) that

$$\Phi'_{\mu,\bar{\gamma}}(x; \triangle\tilde{x}) = \nabla\varphi_\mu(x)^\top\triangle\tilde{x} - \bar{\gamma}\|c(x)\| + c(x)^\top\triangle\tilde{y}. \tag{5.37}$$

By Assumption B3, there exists some $\tilde{\nu} > 0$ such that $\triangle\tilde{x}^\top\bar{\mathcal{H}}\triangle\tilde{x} \geq \tilde{\nu}\|\triangle\tilde{x}\|^2$. Therefore, since $\bar{\sigma} < \frac{1}{2} - \sigma$, we have from Lemma 5.3 that

$$\left(\frac{1}{2} - \bar{\sigma}\right)\Phi'_{\mu,\bar{\gamma}}(x; \triangle\tilde{x}) \leq \sigma\Phi'_{\mu,\bar{\gamma}}(x; \triangle\tilde{x}) - \left(\frac{1}{2} - \bar{\sigma} - \sigma\right)\zeta\triangle\tilde{x}^\top\bar{\mathcal{H}}\triangle\tilde{x}. \tag{5.38}$$

Moreover, since $\zeta > 0$, we have

$$-\left(\frac{1}{2} - \bar{\sigma} - \sigma\right)\zeta\triangle\tilde{x}^\top\bar{\mathcal{H}}\triangle\tilde{x} + o(\|\triangle\tilde{x}\|^2) \leq -\left(\frac{1}{2} - \bar{\sigma} - \sigma\right)\zeta\tilde{\nu}\|\triangle\tilde{x}\|^2 + o(\|\triangle\tilde{x}\|^2) \leq 0. \tag{5.39}$$

Since $\|z - z(\mu)\| = o(\mu)$, Lemma 5.1 implies that $\|\triangle\tilde{y}\| = o(\mu)$ and $\|\triangle\hat{y}\| = o(\mu)$. Hence we have $c(x)^\top\triangle\tilde{y} = o(\|c(x)\|)$ and $c(x)^\top\triangle\hat{y} = o(\|c(x)\|)$ and hence from (5.25) that $c(x + \triangle\hat{x}) = o(\|c(x)\|) + \mathcal{O}(\|\triangle\tilde{x}\|^3)$. Since

$\eta > 2\bar{\sigma}+1$, $\bar{\gamma} > 0$ and $\bar{\gamma} \geq \eta\|y(\mu)\|$, if μ is sufficiently small, there is a constant $\tilde{\sigma} \in (0,1)$ such that $\tilde{\sigma}\bar{\gamma} > (2\bar{\sigma}+1)\|y\|$. Combining (5.36), (5.37), (5.38) and (5.39) and adding and subtracting the term $\frac{1}{2}\tilde{\sigma}\bar{\gamma}\|c(x)\|$, we obtain that

$$\begin{aligned}
& \Phi_{\mu,\bar{\gamma}}(x+\triangle\hat{x}) - \Phi_{\mu,\bar{\gamma}}(x) = \varphi_\mu(x+\triangle\hat{x}) - \varphi_\mu(x) + \bar{\gamma}\left(\|c(x+\triangle\hat{x})\| - \|c(x)\|\right) \\
\leq\ & \left(\tfrac{1}{2}-\bar{\sigma}\right)\Phi'_{\mu,\bar{\gamma}}(x;\triangle\tilde{x}) + \left(\bar{\sigma}+\tfrac{1}{2}\right)c(x)^\top y + \bar{\sigma}c(x)^\top\triangle\tilde{y} - \tfrac{\bar{\gamma}}{2}\|c(x)\| + o(\|c(x)\|) + o\left(\|\triangle\tilde{x}\|^2\right) \\
\leq\ & \sigma\Phi'_{\mu,\bar{\gamma}}(x;\triangle\tilde{x}) - \left(\tfrac{1}{2}-\bar{\sigma}-\sigma\right)\zeta\triangle\tilde{x}^\top\bar{\mathcal{H}}\triangle\tilde{x} - \tfrac{1}{2}\left(\tilde{\sigma}\bar{\gamma} - (2\bar{\sigma}+1)\|y\|\right)\|c(x)\| \\
& -\tfrac{1}{2}(1-\tilde{\sigma})\bar{\gamma}\|c(x)\| + o(\|c(x)\|) + o\left(\|\triangle\tilde{x}\|^2\right) \\
\leq\ & \sigma\Phi'_{\mu,\bar{\gamma}}(x;\triangle\tilde{x}).
\end{aligned}$$

The case of $\|c(x)\| = 0$ follows straightforwardly from (5.37) since $\nabla c(x)^\top\triangle\tilde{x} = 0$. □

We are now ready to state the quadratic convergence of Algorithm 2.1.

THEOREM 5.5. *Suppose μ is sufficiently small, $z(\mu)$ is sufficiently close to z^* and Assumption B holds. Suppose Algorithm 2.1 generates an infinite sequence of iterates $\{z^k\}$, where $z^k = (x^k, \lambda^k, y^k)$, and $\gamma_k = \bar{\gamma}$ for all large enough k. Suppose $z(\mu)$ is an accumulation point of $\{z^k\}$. If $\|z^k - z(\mu)\| = o(\mu)$, then $\|z^{k+1} - z(\mu)\| = \mathcal{O}(\|z^k - z(\mu)\|^2)$.*

Proof. First, since μ is sufficiently small and $\|z^k - z(\mu)\| = o(\mu)$, we have from Assumption B3 and Step 2.1 of Algorithm 2.1 that $\mathcal{H}_k = \nabla^2_{xx}\mathcal{L}(z^k)$, i.e., (5.4) holds. Since $\gamma_k = \bar{\gamma}$ eventually, we have from Step 4.3 of Algorithm 2.1 that $\pi_k = \bar{\pi}$ for some $\bar{\pi} > 0$ and all k large enough. From (1.6), (1.7), (2.1), (2.2) and (5.1) we have

$$\mathcal{W}_{\bar{\gamma}}(z^k)(\triangle\bar{z}^k - \triangle\tilde{z}^k) = \mathcal{O}(\|c(x^k)\|), \tag{5.40}$$

where $\triangle\bar{z}^k = (\triangle\bar{x}^k, \triangle\bar{\lambda}^k, \triangle\bar{y}^k)$ and $\triangle\tilde{z}^k = (\triangle\tilde{x}^k, \triangle\tilde{\lambda}^k, \triangle\tilde{y}^k)$. By Assumption B3 and (5.5), we have $\|\mathcal{W}_{\bar{\gamma}}(z^k)^{-1}\| \leq W$. This together with (5.26) and (5.40) imply $\|\triangle\bar{z}^k - \triangle\tilde{z}^k\| = \mathcal{O}(\|c(x^k)\|) = \mathcal{O}(\|\triangle\tilde{x}^k\|)$. Hence, we have from Theorem 5.1 (ii) that conditions (i) and (ii) of (2.20) and condition (ii) of (2.4) hold. Moreover, since $\vartheta \in (0,1)$, condition (i) of (2.4) holds as well. Since $\gamma_k = \bar{\gamma}$ for all k large enough, the satisfaction of conditions (i) and (ii) of (2.20) implies condition (iii) of (2.20) must be violated, i.e., $\bar{\gamma} \geq \eta\|y^k + \triangle\bar{y}^k\|$. Since $z(\mu)$ is an accumulation point of $\{z^k\}$, there is an infinite sequence of iterates that converges to $z(\mu)$. Hence, by taking limit on k, we have $\bar{\gamma} \geq \eta\|y(\mu)\|$. Since we choose $\zeta \in \left(0, 1-\frac{1}{\eta}\right)$ in Step 0 of Algorithm 2.1, the conditions in Lemma 5.3 are all satisfied and thus condition (iii) of (2.4) holds. Hence (2.4) holds and $\triangle z^k = \triangle\tilde{z}^k$ by Step 2.5 of Algorithm 2.1, where $\triangle z^k = (\triangle x^k, \triangle\lambda^k, \triangle y^k)$.

To use the results of this section and complete our proof, we need to show that, as convergence occurs, all step sizes $\bar{\alpha}^k_x$, $\hat{\alpha}^k_x$, α^k_λ and α^k_y eventually equal 1. It follows from Lemma 5.2 and (2.6) that $\bar{\alpha}^k_x = 1$ and hence $\triangle\hat{x}^k$ can be computed from (5.3). This means that all the results regarding the second-order correction step in this section are valid for x^k. From Lemma 5.2 (ii) and (2.15) we have $\hat{\alpha}^k_x = 1$. By the line search procedure of Algorithm 2.1 and Theorem 5.4, if the first trial point $x^k + \bar{\alpha}^k_x\triangle x^k$ (here $\bar{\alpha}^k_x = 1$ and $\triangle x^k = \triangle\tilde{x}^k$) is rejected, the trial point $x^k + \triangle\hat{x}$ is accepted as the Armijo condition (2.18) is satisfied, i.e., $x^{k+1} = x^k + \triangle\hat{x}^k$. Otherwise, $x^{k+1} = x^k + \triangle\tilde{x}^k$. Lemma 5.2 (iii) and (2.12) imply $\alpha^k_\lambda = 1$. This together with Lemma 5.2 (iv) and (2.13) imply $\lambda^{k+1} = \lambda^k + \triangle\tilde{\lambda}^k$. By Step 4.1 of Algorithm 2.1 we have $\alpha^k_y = 1$ and thus $y^{k+1} = y^k + \triangle\tilde{y}^k$. Now the theorem follows from Theorem 5.1 (i). □

5.2. Local analysis for the overall algorithm. Suppose Algorithm 4.1 generates an infinite sequence of iterates. If the penalty parameter γ_j in Algorithm 4.1 tends to infinity, by following the same argument as in the proof of Theorem 3.1, we obtain that there is an accumulation point of $\{x^j\}$ that is either a Fritz-John point of problem (1.1) at which the MFCQ fails to holds or an infeasible stationary point of the feasibility problem $\min_{x\geq 0}\|c(x)\|^2$. To analyze the rate of convergence of Algorithm 4.1 to a KKT point of problem

(1.1) satisfying the LICQ, we assume that $\gamma_j = \bar{\gamma}$ for all large j. For simplicity we discard the iteration index j in the following analysis.

Suppose the current iterate z satisfies the following criteria for terminating the inner algorithm,

$$\|\mathcal{R}_\mu(z)\| \le \epsilon_\mu, x > 0, \lambda > 0. \tag{5.41}$$

We study in this section the outcome of applying Algorithm 4.1 when μ and ϵ_μ are decreased to μ^+ and ϵ_{μ^+}, respectively. At this time, the μ in the right hand sides of (5.1) and (2.1) is replaced by μ^+.

THEOREM 5.6. *Suppose (5.41) and Assumption B hold. If $z \in \mathcal{N}(z^*)$ and $\triangle\tilde{z}$ is computed by solving (5.1) with μ replaced by μ^+, then $\|z + \triangle\tilde{z} - z^*\| = \mathcal{O}(\|z - z^*\|^2) + \mathcal{O}(\mu^+)$.*

Proof. Analogously to (5.6), the following linear system can be derived from (5.1)

$$\mathcal{W}_{\bar{\gamma}}(z)(z + \triangle\tilde{z} - z^*) = \begin{bmatrix} \mathcal{H}(x - x^*) - \nabla_x \mathcal{L}(x, \lambda^*, y^*) \\ \Lambda(x - x^*) - X\lambda^* + \mu^+ e \\ \frac{\|c(x)\|}{\bar{\gamma}}(y^* - y) - c(x) + \nabla c(x)^\top (x - x^*) \end{bmatrix}, \tag{5.42}$$

where $\mathcal{H} = \nabla^2_{xx}\mathcal{L}(z)$ in view of Assumption B3. Similarly to (5.7) and (5.9), we have

$$\begin{array}{l} \mathcal{H}(x - x^*) - \nabla_x \mathcal{L}(x, \lambda^*, y^*) = \mathcal{O}(\|z - z^*\|^2), \\ \frac{\|c(x)\|}{\bar{\gamma}}(y^* - y) - c(x) + \nabla c(x)^\top (x - x^*) = \mathcal{O}(\|z - z^*\|^2). \end{array}$$

Moreover, letting $X_* = \mathrm{diag}(x^*)$, we have

$$\begin{array}{rl} & \Lambda(x - x^*) - X\lambda^* + \mu^+ e = \Lambda(x - x^*) - X\lambda^* + X_*\lambda^* + \mu^+ \\ = & (X - X_*)(\lambda - \lambda^*) + \mu^+ e = \mathcal{O}(\|z - z^*\|^2) + \mathcal{O}(\mu^+). \end{array}$$

The result then follows from (5.5). ∎

THEOREM 5.7. *Suppose (5.41) and Assumption B hold. If z is sufficiently close to z^*, μ is sufficiently small and $\mu^+ \le \mu$, then $\|\mathcal{R}_{\mu^+}(z + \triangle\tilde{z})\| = \mathcal{O}(\epsilon_\mu^2) + \mathcal{O}(\mu^2)$.*

Proof. Let $\nabla\mathcal{R}_\mu(z)$ be the Jacobian of $\mathcal{R}_\mu(z)$. By Assumption B, if μ is small enough, $\nabla\mathcal{R}_0(z(\mu))$ is nonsingular and $\|(\nabla\mathcal{R}_0(z(\mu)))^{-1}\|$ is bounded. Since z is sufficiently close to z^*, by Taylor's theorem, we have

$$\|z - z(\mu)\| = \left\|(\nabla\mathcal{R}_0(z(\mu)))^{-1}\left(\mathcal{R}_\mu(z) + o(\|z - z(\mu)\|)\right)\right\| = \mathcal{O}(\|\mathcal{R}_\mu(z)\|). \tag{5.43}$$

Theorem 5.1 gives that $\|z + \triangle\tilde{z} - z(\mu^+)\| = \mathcal{O}(\|z - z(\mu^+)\|^2)$. 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 (5.43) and (5.41), we obtain

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

∎

THEOREM 5.8. *Suppose (5.41) and Assumption B hold. Suppose $\mu^+ \le \mu$ and $(\mu + \epsilon_\mu)^2 = o(\mu^+)$. If z is sufficiently close to z^* and μ is sufficiently small, then $x + \triangle\tilde{x} > 0$ and $\lambda + \triangle\tilde{\lambda} > 0$.*

Proof. We have from (5.1), (5.5) and (5.41) that

$$\begin{array}{rl} & \|\triangle\tilde{z}\| = \|\mathcal{W}_{\bar{\gamma}}(z)^{-1}\mathcal{R}_{\mu^+}(z)\| \\ = & \left\|\mathcal{W}_{\bar{\gamma}}(z)^{-1}\left(\mathcal{R}_\mu(z) + (\mu^+ - \mu)(0, e, 0)^\top\right)\right\| \\ = & \mathcal{O}(\epsilon_\mu) + \mathcal{O}(\mu), \end{array}$$

where the last equation uses the fact that $\mu^+ \le \mu$. From the second equation of (5.1) with μ replaced by μ^+, we have

$$\begin{aligned} & \operatorname{diag}(\lambda + \triangle\tilde{\lambda})(x + \triangle\tilde{x}) \\ = \; & \Lambda x + X\triangle\tilde{\lambda} + \Lambda\triangle\tilde{x} + \operatorname{diag}(\triangle\tilde{\lambda})\triangle\tilde{x} \\ = \; & \mu^+ e + \mathcal{O}(\|\triangle\tilde{z}\|^2) = \mu^+ e + \mathcal{O}((\epsilon_\mu + \mu)^2). \end{aligned} \tag{5.44}$$

Since $(\mu + \epsilon_\mu)^2 = o(\mu^+)$, we have for $i \in \mathcal{B}$ that $\lambda_i + \triangle\tilde{\lambda}_i > 0$ by strict complementarity and hence that $x_i + \triangle\tilde{x}_i > 0$ by (5.44). Also, for $i \in \mathcal{I}\backslash\mathcal{B}$, $x_i + \triangle\tilde{x}_i > 0$ and thus $\lambda_i + \triangle\tilde{\lambda}_i > 0$ by (5.44). □

We are now ready to prove the superlinear convergence of Algorithm 4.1.

THEOREM 5.9. *Suppose all assumptions in Theorem 4.1 hold and Algorithm 4.1 generates an infinite sequence of iterates $\{z^j\}$, where $z^j = (x^j, \lambda^j, y^j)$. Suppose z^* is an accumulation point of $\{z^j\}$ and Assumption B holds. Suppose ϵ_{μ_j} and μ_j are decreased so that*

$$\begin{cases} (\epsilon_{\mu_j} + \mu_j)^2 = o(\epsilon_{\mu_{j+1}}), \\ \mu_{j+1} = o(\|\mathcal{R}_0(z^j)\|). \end{cases} \tag{5.45}$$

If z^j is sufficiently close to z^*, then z^{j+1} can be obtained by the inner algorithm within one iteration and $\|z^{j+1} - z^*\| = o(\|z^j - z^*\|)$.

Proof. We use the notation $z^{j,k}$ to denote the kth inner iterate at the jth outer iteration. At iteration $j+1$ of Algorithm 4.1, Algorithm 2.1 with the starting point $z^{j+1,0}(= z^j)$ is used to solve the barrier problem (1.2) for the barrier parameter $\mu_{j+1}(\le \mu_j)$. Since z^j satisfies the termination criterion of the inner algorithm for the barrier parameter μ_j, i.e., $\|\mathcal{R}_{\mu_j}(z^j)\| \le \epsilon_{\mu_j}$, and (5.45) holds, it follows that all the conditions in Theorem 5.8 are satisfied. Hence we have from Theorem 5.8 that $x^{j+1,0} + \triangle\tilde{x}^{j+1,0} > 0$ and $\lambda^{j+1,0} + \triangle\tilde{\lambda}^{j+1,0} > 0$. Moreover, we have from Theorem 5.7 and (5.45) that

$$\|\mathcal{R}_{\mu_{j+1}}(z^{j+1,0} + \triangle\tilde{z}^{j+1,0})\| = \mathcal{O}(\epsilon_{\mu_j}^2) + \mathcal{O}(\mu_j^2) = o(\epsilon_{\mu_{j+1}}).$$

Hence, $\|\mathcal{R}_{\mu_{j+1}}(z^{j+1,0} + \triangle\tilde{z}^{j+1,0})\| \le \epsilon_{\mu_{j+1}}$. Consequently, we obtain from Step 2.6 and Step 1 of Algorithm 2.1 that $z^{j+1,1} = z^{j+1,0} + \triangle\tilde{z}^{j+1,0}$ and $\|\mathcal{R}_{\mu_{j+1}}(z^{j+1,1})\| \le \epsilon_{\mu_{j+1}}$. This implies that $z^{j+1} = z^{j+1,1}$ according to Algorithm 4.1. By Theorem 5.6 we have $\|z^{j+1} - z^*\| = \mathcal{O}(\|z^j - z^*\|^2) + \mathcal{O}(\mu_{j+1})$. By the smoothness of $f(x)$ and $c(x)$, we have $\|\mathcal{R}_0(z^j)\| = \|\mathcal{R}_0(z^j) - \mathcal{R}_0(z^*)\| = \mathcal{O}(\|z^j - z^*\|)$. Since $\mu_{j+1} = o(\|\mathcal{R}_0(z^j)\|)$, we obtain that $\|z^{j+1} - z^*\| = o(\|z^j - z^*\|)$. □

Several strategies for updating μ_j and ϵ_{μ_j} in exact penalty methods are discussed in [2]. In particular, two strategies are given that guarantee (5.45) 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 $\alpha \in [0, \sqrt{n})$. The first strategy chooses $\mu_{j+1} = \mu_j^{1+\delta}$ with $\delta \in (0,1)$, while second strategy chooses $\mu_{j+1} = \|\mathcal{R}_0(z^j)\|^{1+\delta}$ for all large j. It can be readily verified that these two strategies can be used in our method as well to generate superlinear convergence.

6. Conclusion. In this paper we have presented a line search interior-point penalty method for nonlinear programming, in which the step direction of each iteration is computed based on the modified Newton method proposed in [4]. The line search step employs two merit functions: the PLPF $\mathcal{P}_\mu(\rho)$ and the ℓ_2-penalty function $\Phi_{\mu,\gamma}(x)$. A trial point is accepted if it yields a sufficient reduction in either one of the two merit functions. Our PLPF method is closely related to the filter method of Fletcher and Leyffer [7]. In particular, they both are defined by previous iterates and try to measure iteration progress according to historical data. However, the PLPF method inherits the feature of penalty methods that requires the progress to be made in a certain combination of the barrier function and the constraint violation. This differs from the nondomination idea of filter methods. By combining the PLPF method and the ℓ_2-penalty method, we are able to guarantee that our line search step always terminates successfully with an acceptable

step size. This enables our method to avoid the use of a restoration phase used by filter methods to reduce infeasibility.

We have shown that our method has the same strong global convergence properties as those established in [4]. Moreover, by introducing a second-order correction step, our method also enjoys fast local convergence properties. Specifically, under the standard nondegeneracy assumptions, we have shown that for each small enough 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. As shown in [5], these fast local convergence properties also hold for a suitably modified version of the interior-point ℓ_2-penalty method proposed in [4]. We have implemented the proposed method within the software package IPOPT [21]. The obtained numerical results show that our method is competitive with both penalty methods and filter methods. These results will be presented in a future paper [6].

REFERENCES

[1] R. H. Byrd, M. E. Hribar and J. Nocedal, *An interior point algorithm for large-scale nonlinear programming*, SIAM J. Optim., 9 (1999), pp. 877-900.

[2] 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.

[3] R. M. Chamberlain, C. Lemarechal, H. C. Pedersen and M. J. D. Powell, *The watchdog technique for forcing convergence in algorithms for constrained optimization*, Math. Programming Study, 16 (1982), pp. 1-17.

[4] 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.

[5] ———, *On the fast local convergence of interior-point ℓ_2-penalty methods for nonlinear programming,* CORC Technical Report, Columbia University, 2007.

[6] L. Chen, D. Goldfarb, A. Wächter and Z. Wen, *On the implementation of an interior-point penalty method for nonlinear programming*, in preparation, Columbia University, 2007.

[7] R. Fletcher and S. Leyffer, *Nonlinear Programming without a penalty function*, Math. Programming, 91 (2002), pp. 239-269.

[8] A. Forsgren, P. E. Gill and M. H. Wright, *Interior methods for nonlinear optimization*, SIAM Rev., 44 (2002), pp. 525-597.

[9] E. M. Gertz and P. E. Gill, *A primal-dual trust region algorithm for nonlinear optimization*, Math. Programming, 100 (2004), pp. 49-94.

[10] F. Gomes, *A sequential quadratic programming algorithm with a piecewise linear merit function*, Optimization online, 2004.

[11] N. I. M. Gould, D. Orban and Ph. L. Toint, *CUTEr (and SifDec), a constrained and unconstrained testing environment, revised*, Technical Report TR/PA/01/04, CERFACS, Toulouse, France, 2001.

[12] ———, *An interior-point l_1-penalty method for nonlinear optimization*, RAL-TR-2003-022, Computational Science and Engineering Department, Rutherford Appleton Laboratory, Chilton, Oxfordshire, England, 2003.

[13] I. Griva, D. F. Shanno and R. J. Vanderbei, *Convergence analysis of a primal-dual interior-point method for nonlinear programming*, Optimization Online (http://www.optimization-online.org/DB_HTML/2004/07/913.html), July, 2004.

[14] S. Leyffer, G. López-Calva and J. Nocedal, *Interior methods for mathematical programs with complementarity constraints*, SIAM J. Optim., 17 (2006), pp. 52-77.

[15] S. Scholtes and M. Stöhr, *Exact penalization of mathematical programs with equilibrium constraints*, SIAM J. Control Optim., 37 (1999), pp. 617-652.

[16] S. Scholtes and H. Scheel, *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity*, Math. Oper. Res., 25 (2000), pp. 1-22.

[17] A. L. Tits, A. Wächter, S. Bakhtiari, T. J. Urban and C. T. Lawrence, *A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties,* SIAM J. Optim., 14 (2003), pp. 173-199.

[18] H. M. Xu and D. Ralph, *Convergence of a penalty method for mathematical programming with complementairty constraints*, J. Optim. Theory Appl., 123 (2004), pp. 365-390.

[19] R. J. Vanderbei and D. F. Shanno, *An interior-point algorithm for nonconvex nonlinear programming*, Comput. Optim. Appl., 13 (1999), pp. 231-252.

[20] 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.

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

[22] H. Yamashita and H. Yabe, *An interior point method with a primal-dual quadratic barrier penalty function for nonlinear optimizaiton*, SIAM J. Optim., 14 (2003), pp. 479-499.