PREPRINT MCS-P485-1294, MCS DIVISION, ARGONNE NATIONAL LABORATORY

On the Role of the Objective Function in Barrier Methods*

Florian Jarre†and Stephen Wright‡

December 7, 1994

Abstract

To simplify the analysis of interior-point methods, one commonly formulates the problem so that the objective function is linear, by introducing a single extra variable if necessary. Here we show that a linear objective function makes the Newton direction for a barrier function a useful search direction if the current iterate is sufficiently close to the central path. Hence, there are two advantages to using a linear objective and staying close to the central path. First, the Newton direction (which coincides with the affine scaling direction on the central path) gives a very accurate approximation to the direction to the minimum. Second, a long step along the Newton direction is possible without violating the inequality constraints.

1 Introduction

We consider logarithmic barrier methods applied to the nonlinear programming problem

$$\min f(x) \qquad \text{subject to} \qquad c(x) \geq 0, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are smooth (three times continuously differentiable) functions. We are particularly interested in the case of linear f, that is,

$$f(x) = g^T x. \tag{2}$$

*This research was supported by Obermann Fellowships in the Center for Advanced Studies at the University of Iowa and by the Office of Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38.

†Institute of Statistical Mathematics, Tokyo, Japan (Oct. 1994 – March 1995) and Institut für Angewandte Mathematik und Statistik, Universität Würzburg, am Hubland, 97074 Würzburg, Germany jarre@vax.rz.uni-wuerzburg.d400.de

‡Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439. wright@mcs.anl.gov

The logarithmic barrier function for (1) is

$$P(x;\mu) = f(x) - \mu \sum_{i=1}^{m} \ln c_i(x). \tag{3}$$

We denote by $x(\mu)$ a minimizer of $P(.;\mu)$ for $\mu > 0$ and assume that $x(\mu)$ exists for all sufficiently small μ. Methods based on (3) approximate $x(\mu)$ for a sequence of small, decreasing values of $\mu > 0$. Under certain conditions (see Fiacco and McCormick [1]), we have $\lim_{\mu \downarrow 0} x(\mu) = x^*$. After $x(\mu)$ has been approximated for a particular value of $\mu > 0$, an obvious way to proceed is to decrease μ to some new value μ_+ and then take a Newton step for minimizing the barrier function $P(x;\mu_+)$. It has recently been observed (see, for example, Wright [3]) that the resulting direction is generally a poor one, and some specialized direction generation and line search strategies have been proposed to remedy the situation. Below, we illustrate the bad behavior of these steps by means of a simple example. Then, in the main part of this article, we show that these difficulties do not arise when the objective function f is linear and the current iterate is close to the central path defined by $\{x(\mu) \,|\, \mu > 0\}$.

- **Example:** Consider the problem

$$\min\ x_1 + \tfrac{1}{2}(x_2 - 1)^2, \ \text{ subject to } \ (x_1, x_2) \geq 0, \tag{4}$$

for which the solution is obviously $x^* = (0,1)^T$. Let us assume that the barrier parameter μ is small, and write

$$P(x;\mu) = x_1 + \tfrac{1}{2}(x_2 - 1)^2 - \mu \ln x_1 - \mu \ln x_2.$$

The derivatives of $P(\cdot;\mu)$ are

$$\nabla_x P(x;\mu) = \begin{bmatrix} 1 - \mu/x_1 \\ (x_2 - 1) - \mu/x_2 \end{bmatrix}, \qquad \nabla_{xx} P(x;\mu) = \begin{bmatrix} \mu/x_1^2 & 0 \\ 0 & 1 + \mu/x_2^2 \end{bmatrix}.$$

By equating $\nabla_x P(x;\mu)$ to zero, we find that $x(\mu) \approx (\mu, 1+\mu)$. Suppose the barrier parameter is decreased to $\mu_+ \in (0,\mu)$. Then

$$\nabla_x P(x(\mu);\mu_+) \approx \begin{bmatrix} (\mu - \mu_+)/\mu \\ (\mu - \mu_+) \end{bmatrix}, \qquad \nabla_{xx} P(x;\mu_+) \approx \begin{bmatrix} \mu_+/\mu^2 & 0 \\ 0 & 1 \end{bmatrix}.$$

The Newton step starting at $x(\mu)$ for finding $x(\mu_+)$ is therefore

$$p = -\left[\nabla_{xx} P(x(\mu);\mu_+)\right]^{-1} \nabla_x P(x(\mu);\mu_+) \approx -\begin{bmatrix} (\mu/\mu_+)(\mu - \mu_+) \\ (\mu - \mu_+) \end{bmatrix},$$

while the actual difference between the two points is

$$x(\mu_+) - x(\mu) \approx -\begin{bmatrix} \mu - \mu_+ \\ \mu - \mu_+ \end{bmatrix}.$$

The Newton step provides an excellent estimate of the second component of this difference, but it overestimates the first component by a factor of μ/μ_+. If, as is usual, we conduct a line search along the Newton direction p, the line search parameter α can be at most μ_+/μ because of the requirement $x \geq 0$. This small value of α slows convergence in the second component; the error in this component is decreased only by a factor of about $(1 - \mu_+/\mu)$.

We use the following notation in the rest of the paper. For related positive quantities α and β, we say $\beta = O(\alpha)$ if there is a constant M such that $\beta \leq M\alpha$ for all α sufficiently small. We say $\beta = o(\alpha)$ if $\beta/\alpha \to 0$ as $\alpha \to 0$, and $\beta = \Theta(\alpha)$ if $\beta = O(\alpha)$ and $\alpha = O(\beta)$.

2 Assumptions and Basic Facts

Suppose that for some value of μ we have found an approximation to the minimizer $x(\mu)$ of $P(x;\mu)$. We quantify the inexactness in this approximation by defining a vector w by

$$w = \frac{1}{\mu}\nabla_x P(x;\mu),$$

so that

$$\mu w = \nabla f(x) - \sum_{i=1}^{m} \frac{\mu}{c_i(x)} \nabla c_i(x). \tag{5}$$

We use the notation $x_w(\mu)$ to denote the approximate minimizer, to emphasize its relationship to w. We note in passing that $x_w(\mu)$ is the exact minimizer of the modified log barrier function

$$P_w(x;\mu) = f(x) - \mu \sum_{i=1}^{m} \ln c_i(x) - \mu w^T x \tag{6}$$

and that $x(\mu) = x_0(\mu)$.

As the next section indicates, our results depend on an assumption that $\mu\|w\|$ is sufficiently small. This implies that $x_w(\mu)$ is not too far from the central path traced by the exact minimizers of $P(x;\mu)$ for $\mu > 0$.

Before proceeding, we define terminology and optimality conditions and specify some additional assumptions. The Lagrangian function for (1) is

$$\mathcal{L}(x,\lambda) = f(x) - \lambda^T c(x), \tag{7}$$

where λ is the vector of Lagrange multipliers. A point x^* satisfies the first-order conditions for optimality if $c(x^*) \geq 0$ and there is a vector $\lambda^* \geq 0$ such that $(\lambda^*)^T c(x^*) = 0$ and $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$, that is,

$$\nabla f(x^*) = A(x^*)\lambda^*, \tag{8}$$

where

$$A(x) = [\nabla c_1(x), \nabla c_2(x), \ldots, \nabla c_m(x)]$$

is the transpose of the Jacobian of c. The active constraints are the components of c for which $c_i(x^*) = 0$. Without loss of generality we assume these to be the first p components of c, and define

$$\bar{A}(x) = [\nabla c_1(x), \nabla c_2(x), \ldots, \nabla c_p(x)], \qquad \bar{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_p]. \tag{9}$$

We assume nondegeneracy, that is, $\bar{A}(x^*)$ has full column rank p, and strict complementarity, that is, $\lambda^* + c(x^*) > 0$. A consequence of nondegeneracy is that $p \leq n$. We assume further that x^* satisfies the second-order sufficient conditions for optimality, namely,

$$y^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) y > 0 \quad \text{for all } y \text{ with } \bar{A}(x^*)^T y = 0,\ y \neq 0. \tag{10}$$

The derivatives of the barrier function (3) are

$$\nabla_x P(x;\mu) = \nabla f(x) - \sum_{i=1}^m \frac{\mu}{c_i(x)} \nabla c_i(x), \tag{11a}$$

$$\nabla_{xx} P(x;\mu) = \nabla^2 f(x) + \mu \sum_{i=1}^m \left[\frac{1}{c_i^2(x)} \nabla c_i(x) \nabla c_i(x)^T - \frac{1}{c_i(x)} \nabla^2 c_i(x) \right]. \tag{11b}$$

Given any strictly feasible approximate solution $x_w(\mu)$, we can define Lagrange multiplier estimates $\lambda_w(\mu)$ by

$$\lambda_w(\mu) = \mu C(x_w(\mu))^{-1} e = \left[\frac{\mu}{c_1(x_w(\mu))}, \ldots, \frac{\mu}{c_m(x_w(\mu))} \right]^T, \tag{12}$$

where $C(x) = \operatorname{diag}(c(x))$. Combining (12) with (5), we have

$$\nabla f(x_w(\mu)) = A(x_w(\mu)) \lambda_w(\mu) + \mu w. \tag{13}$$

Naturally, we use the notation $\lambda(\mu)$ for the Lagrange multipliers associated with the exact minimizer $x(\mu)$.

Let us return to the case of linear f. Setting $(x, \lambda) = (x_w(\mu), \lambda_w(\mu))$ and choosing a new value $\mu_+ \in (0, \mu)$, we have from (2), (11), and (12) that

$$\begin{aligned} \nabla_x P(x;\mu_+) &= g - (\mu_+/\mu) A(x) \lambda = (1 - \mu_+/\mu) g + \mu_+ w, \\ \nabla_{xx} P(x;\mu_+) &= (\mu_+/\mu^2) \left[A(x) \Lambda^2 A(x)^T - \mu \sum_{i=1}^m \lambda_i \nabla^2 c_i(x) \right], \end{aligned}$$

where $\Lambda = \operatorname{diag}(\lambda)$. For the Newton direction p, we have

$$\begin{aligned} p &= -\nabla_{xx} P(x;\mu_+)^{-1} \nabla_x P(x;\mu_+) \\ &= -\mu \left[A(x) \Lambda^2 A(x)^T - \mu \sum_{i=1}^m \lambda_i \nabla^2 c_i(x) \right]^{-1} [(\mu/\mu_+ - 1) g + \mu w]. \end{aligned} \tag{14}$$

3 Limiting Behavior of the Newton Directions

We now compare the Newton direction (14) with an estimate of the differences between $x_w(\mu)$, $x(\mu_+)$, and the solution x^*. Consider the vector function

$$F(x,\lambda) = \begin{bmatrix} \nabla_x \mathcal{L}(x,\lambda) \\ \Lambda c(x) \end{bmatrix} = \begin{bmatrix} g - A(x)\lambda \\ \Lambda c(x) \end{bmatrix}, \tag{15}$$

which maps $\mathbb{R}^n \times \mathbb{R}^m$ to itself, and note that $F(x^*,\lambda^*) = 0$. The Jacobian of F is

$$\nabla F(x,\lambda) = \begin{bmatrix} \nabla_{xx}\mathcal{L}(x,\lambda) & -A(x) \\ \Lambda A(x)^T & C(x) \end{bmatrix} = \begin{bmatrix} -\sum_{i=1}^m \lambda_i \nabla^2 c_i(x) & -A(x) \\ \Lambda A(x)^T & C(x) \end{bmatrix}. \tag{16}$$

Nonsingularity of $\nabla F(x^*,\lambda^*)$ follows from a standard argument, which we now sketch. Let y and z be such that

$$\nabla F(x^*,\lambda^*) \begin{bmatrix} y \\ z \end{bmatrix} = 0. \tag{17}$$

Because $c_i(x^*) > 0$ and $\lambda_i^* = 0$ for $i = p+1, p+2, \ldots, m$, we have by considering the last $m-p$ rows of (17) that $z_i = 0$, for $i = p+1, p+2, \ldots, m$. Because $c_i(x^*) = 0$ and $\lambda_i^* > 0$ for $i = 1, 2, \ldots, p$, we have again from (17) and (9) that $\bar{A}(x^*)^T y = 0$. Hence,

$$y^T \nabla_{xx}\mathcal{L}(x^*,\lambda^*) y = y^T A(x^*) z = y^T \bar{A}(x^*) \begin{bmatrix} z_1 \\ \vdots \\ z_p \end{bmatrix} = 0. \tag{18}$$

Because of second-order sufficiency, (10) and (18) are consistent only if $y = 0$. Full rank of $\bar{A}(x^*)$ then implies that $z_i = 0$, $i = 1, \ldots, p$. Hence $z = 0$, and our claim is proved.

The following result is a consequence of the implicit function theorem (see, for example, Lang [2, p. 131]).

Theorem 3.1 *Let the vector pair $(x(z,\sigma), \lambda(z,\sigma))$ be defined implicitly as the solution of the nonlinear system*

$$F(x,\lambda) = \begin{bmatrix} z \\ \sigma e \end{bmatrix}, \tag{19}$$

for given (z,σ). Then there are positive constants $\epsilon > 0$ and $M > 0$ such that the following statements hold.

(i) $(x(z,\sigma), \lambda(z,\sigma))$ is a C^2 function of (z,σ) in the neighborhood defined by

$$\mathcal{N}_\epsilon = \{(z,\sigma) \mid \|z\| + |\sigma| \le \epsilon\}.$$

(ii) For $\sigma > 0$ and $(z,\sigma) \in \mathcal{N}_\epsilon$, we have $\lambda_i(z,\sigma) > 0$ and $c_i(z,\sigma) > 0$ for $i = 1, 2, \ldots, m$.

(iii) For (z_1, σ_1) and (z_2, σ_2) in $\mathcal{N}_\epsilon$, we have

$$\begin{aligned} \begin{bmatrix} x_1 \\ \lambda_1 \end{bmatrix} - \begin{bmatrix} x_2 \\ \lambda_2 \end{bmatrix} &= \nabla F(x_1, \lambda_1)^{-1} \begin{bmatrix} z_1 - z_2 \\ (\sigma_1 - \sigma_2)e \end{bmatrix} + r_1, & (20a) \\ &= \nabla F(x_2, \lambda_2)^{-1} \begin{bmatrix} z_1 - z_2 \\ (\sigma_1 - \sigma_2)e \end{bmatrix} + r_2, & (20b) \end{aligned}$$

where $(x_1, \lambda_1) = (x(z_1, \sigma_1), \lambda(z_1, \sigma_1))$ and $(x_2, \lambda_2) = (x(z_2, \sigma_2), \lambda(z_2, \sigma_2))$, and

$$\begin{aligned} \|r_1\| &\leq M(\|z_1 - z_2\| + |\sigma_1 - \sigma_2|)^2, & (21a) \\ \|r_2\| &\leq M(\|z_1 - z_2\| + |\sigma_1 - \sigma_2|)^2. & (21b) \end{aligned}$$

Proof. Since $\nabla F(x^*, \lambda^*)$ is nonsingular and $F(\cdot, \cdot)$ is C^2 in a neighborhood of (x^*, λ^*), we have from the implicit function theorem that there is $\epsilon_1 > 0$ such that $(x(z, \sigma), \lambda(z, \sigma))$ is a C^2 function of (z, σ) with $(x(0, 0), \lambda(0, 0)) = (x^*, \lambda^*)$ whenever $(z, \sigma) \in \mathcal{N}_{\epsilon_1}$.

For (ii), we can use the strict complementarity condition to choose $\epsilon_2 \in (0, \epsilon_1]$ such that for $(z, \sigma) \in \mathcal{N}_{\epsilon_2}$ we have

$$\lambda_i(z, \sigma) > 0, \quad i = 1, \ldots, p, \qquad c_i(x(z, \sigma)) > 0, \quad i = p + 1, \ldots, m.$$

The condition $\lambda_i c_i(x) = \sigma > 0$ ensures that the complementary components are also strictly positive; that is,

$$c_i(x(z, \sigma)) > 0, \quad i = 1, \ldots, p, \qquad \lambda_i(z, \sigma) > 0, \quad i = p + 1, \ldots, m.$$

For (iii), we use the result of (i) to choose $\epsilon_3 \in (0, \epsilon_2]$ such that the condition number of $\nabla F(x(z, \sigma), \lambda(z, \sigma))$ is uniformly bounded in the neighborhood $\mathcal{N}_{\epsilon_3}$. By a Taylor series argument and Lipschitz continuity of ∇F in this neighborhood, we have

$$F(x_1, \lambda_1) - F(x_2, \lambda_2) = \nabla F(x_1, \lambda_1) \begin{bmatrix} x_1 - x_2 \\ \lambda_1 - \lambda_2 \end{bmatrix} + O\left(\left\| \begin{bmatrix} x_1 - x_2 \\ \lambda_1 - \lambda_2 \end{bmatrix} \right\|^2 \right).$$

Therefore, substituting for $F(x_1, \lambda_1)$ and $F(x_2, \lambda_2)$ from (19) and multiplying through by the inverse of $\nabla F(x_1, \lambda_1)$, we have

$$\begin{bmatrix} x_1 - x_2 \\ \lambda_1 - \lambda_2 \end{bmatrix} = \nabla F(x_1, \lambda_1)^{-1} \begin{bmatrix} z_1 - z_2 \\ (\sigma_1 - \sigma_2)e \end{bmatrix} + O\left(\|\nabla F(x_1, \lambda_1)^{-1}\| \left\| \begin{bmatrix} x_1 - x_2 \\ \lambda_1 - \lambda_2 \end{bmatrix} \right\|^2 \right). \quad (22)$$

Since (x, λ) is a C^2 function of (z, λ) on $\mathcal{N}_{\epsilon_3}$, it is certainly Lipschitz C^1. Hence we have

$$\begin{bmatrix} x_1 - x_2 \\ \lambda_1 - \lambda_2 \end{bmatrix} = \begin{bmatrix} \frac{dx}{dz}(z_1, \sigma_1) & \frac{dx}{d\sigma}(z_1, \sigma_1) \\ \frac{d\lambda}{dz}(z_1, \sigma_1) & \frac{d\lambda}{d\sigma}(z_1, \sigma_1) \end{bmatrix} \begin{bmatrix} z_1 - z_2 \\ \sigma_1 - \sigma_2 \end{bmatrix} + O\left(\left\| \begin{bmatrix} z_1 - z_2 \\ \sigma_1 - \sigma_2 \end{bmatrix} \right\|^2 \right). \quad (23)$$

The Jacobian term in (23) is bounded, so we can combine (22) and (23) to obtain the result (20a) for r_1 satisfying (21a), for some $M > 0$. The remaining result (20b), (21b) follows identically.

The choice $\epsilon = \epsilon_3 > 0$ is sufficient for all of (i), (ii), and (iii). ∎

It is not difficult to show that $x(0, \mu)$ defined from (19) is an isolated local minimizer of $P(\cdot; \mu)$ for sufficiently small μ. In fact, $x(\mu w, \mu)$ is an isolated local minimizer of the modified log barrier function $P_w(\cdot; \mu)$ provided that w is not too large. (The claim certainly holds when $\|w\| = o(\mu^{-1/2})$, as we assume below in (29).)

Uniform nonsingularity of ∇F in a neighborhood of (x^*, λ^*) implies that the first terms in the right-hand side of (20a) and (20b) dominate the second terms for ϵ sufficiently small; that is,

$$\left\| \begin{bmatrix} x_1 \\ \lambda_1 \end{bmatrix} - \begin{bmatrix} x_2 \\ \lambda_2 \end{bmatrix} \right\| = \Theta\left(\left\| \begin{bmatrix} z_1 - z_2 \\ (\sigma_1 - \sigma_2)e \end{bmatrix} \right\| \right) = \Theta(\|z_1 - z_2\| + |\sigma_1 - \sigma_2|). \tag{24}$$

Let ϵ be the quantity in Theorem 3.1, and suppose that μ and w are such that

$$\mu(\|w\| + 1) \le \epsilon.$$

Then for any $\mu_+ \in [0, \mu]$, the vector pairs $(z, \sigma) = (-\mu w, \mu)$ and $(z, \sigma) = (0, \mu_+)$ both lie in the neighborhood $\mathcal{N}_\epsilon$. It is easy to verify that

$$\begin{aligned} (z, \sigma) = (\mu w, \mu) \quad &\Rightarrow \quad (x, \lambda) = (x_w(\mu), \lambda_w(\mu)), \\ (z, \sigma) = (0, \mu_+) \quad &\Rightarrow \quad (x, \lambda) = (x(\mu_+), \lambda(\mu_+)). \end{aligned}$$

Hence, Theorem 3.1 (iii) implies that

$$\begin{bmatrix} x(\mu_+) \\ \lambda(\mu_+) \end{bmatrix} - \begin{bmatrix} x_w(\mu) \\ \lambda_w(\mu) \end{bmatrix} = \nabla F(x_w(\mu), \lambda_w(\mu))^{-1} \begin{bmatrix} -\mu w \\ -(\mu - \mu_+)e \end{bmatrix} + r, \tag{25}$$

where

$$\|r\| \le M\left((\mu\|w\| + |\mu - \mu_+|)^2 \right) \le M(\mu^2(\|w\| + 1)^2).$$

Returning to (14), using $(x, \lambda) = (x_w(\mu), \lambda_w(\mu))$, and noting that $c(x) > 0$, we can define a vector $\delta\lambda \in \mathbb{R}^m$ implicitly by the equation

$$\Lambda A(x)^T (\mu_+/\mu) p + C(x)\delta\lambda = -(\mu - \mu_+)e. \tag{26}$$

We now show that (14) and (26) together imply that $((\mu_+/\mu)p, \delta\lambda)$ satisfies

$$\nabla F(x, \lambda) \begin{bmatrix} (\mu_+/\mu)p \\ \delta\lambda \end{bmatrix} = \begin{bmatrix} \nabla_{xx}\mathcal{L}(x, \lambda) & -A(x) \\ \Lambda A(x)^T & C(x) \end{bmatrix} \begin{bmatrix} (\mu_+/\mu)p \\ \delta\lambda \end{bmatrix} = \begin{bmatrix} -\mu w \\ -(\mu - \mu_+)e \end{bmatrix}. \tag{27}$$

The second block row of (27) follows immediately from (26). To recover the Newton equations (14), multiply the second block row from the left by $A(x)C(x)^{-1}$, add it to the first block row, and use the identities

$$C(x)^{-1} = \frac{1}{\mu}\Lambda, \qquad A(x)\Lambda e = g - \mu w.$$

This manipulation yields

$$\left[\frac{1}{\mu}A(x)\Lambda^2 A(x)^T - \sum_{i=1}^{m}\lambda_i \nabla^2 c_i(x)\right]\frac{\mu_+}{\mu}p = -\mu w + \frac{\mu_+ - \mu}{\mu}A(x)\lambda = \left(\frac{\mu_+}{\mu} - 1\right)g - \mu_+ w,$$

which is identical to (14).

It follows from (25) and (27) that

$$\left[\begin{array}{c} x(\mu_+) - x_w(\mu) \\ \lambda(\mu_+) - \lambda_w(\mu) \end{array}\right] = \left[\begin{array}{c} (\mu_+/\mu)p \\ \delta\lambda \end{array}\right] + O(\mu^2(\|w\|+1)^2),$$

and consequently

$$(\mu_+/\mu)p = x(\mu_+) - x_w(\mu) + O(\mu^2(\|w\|+1)^2). \tag{28}$$

The result we are most interested in — that $(\mu_+/\mu)p$ and $x(\mu_+) - x_w(\mu)$ are asymptotically the same direction — will follow from (28) if we can show that the $O(\mu^2(\|w\|+1)^2)$ remainder term is eventually dominated by either (hence both) of these two vectors. To show this, we need additional conditions on w and μ_+, namely,

$$\|w\| = o(\mu^{-1/2}) \qquad \text{and} \qquad \mu_+ \le \rho\mu, \quad \text{for some } \rho \in (0,1). \tag{29}$$

The first condition controls the inexactness that we allow in the calculation of the minimizer of $P(x;\mu)$ as $\mu \downarrow 0$. Because of (5), the ratio

$$\nabla_x P(x_w(\mu);\mu)/\mu^{1/2}$$

should approach zero as $\mu \downarrow 0$, where $x_w(\mu)$ is, as usual, the approximate minimizer for this value of μ. The condition on μ_+ in (29) simply means that there is a significant decrease in μ at each step. If this condition fails to hold, so that $\mu_+ \approx \mu$, we expect the distance even between two successive *exact* minimizers $x(\mu)$ and $x(\mu_+)$ to be short; indeed, it might be swallowed up in the remainder term in (28).

Next, we show that (29) implies $\mu^2(\|w\|+1)^2 = o(||(\mu_+/\mu)p||)$. Given (29) and sufficiently small μ, we have from (28) that

$$(\mu_+/\mu)p = x(\mu_+) - x_w(\mu) + o(\mu). \tag{30}$$

Hence, our desired result will hold if we can show that $(\mu_+/\mu)p = \Theta(\mu)$. From (27) and uniform nonsingularity of $\nabla F(x,\lambda)$ near (x^*,λ^*), we have

$$\left[\begin{array}{c} (\mu_+/\mu)p \\ \delta\lambda \end{array}\right] = O(\mu(\|w\|+1)) = o(\mu^{1/2}). \tag{31}$$

Applying Theorem 3.1(iii) or, alternatively, Equation (24), with

$$\begin{aligned} (z_1,\sigma_1) = (\mu w, \mu), &\quad \text{that is,} \quad (x_1,\lambda_1) = (x_w(\mu), \lambda_w(\mu)), \\ (z_2,\sigma_2) = (0,0), &\quad \text{that is,} \quad (x_2,\lambda_2) = (x^*,\lambda^*), \end{aligned}$$

we have

$$\left\| \begin{bmatrix} x_w(\mu) - x^* \\ \lambda_w(\mu) - \lambda^* \end{bmatrix} \right\| = O(\mu(\|w\| + 1)) = o(\mu^{1/2}).$$

Hence, for any active constraint ($i = 1, 2, \ldots, p$), we have

$$c_i(x_w(\mu)) = c_i(x_w(\mu)) - c_i(x^*) = o(\mu^{1/2}). \tag{32}$$

Further, by our strict complementarity assumptions, we can define a constant $\gamma > 0$ such that

$$[\lambda_w(\mu)]_i \leq \lambda_i^* + o(\mu^{1/2}) \leq 2\lambda_i^* \leq \gamma, \qquad i = 1, 2, \ldots, p, \tag{33}$$

for μ sufficiently small.

Consider now the first p components of the equation (26). Defining $\bar{A}(x)$ and $\bar{\lambda}$ as in (9), and $\bar{c}(x)$, $\bar{\Lambda}$, $\delta\bar{\lambda}$ accordingly, we have for $(x, \lambda) = (x_w(\mu), \lambda_w(\mu))$ that

$$\begin{aligned} &\bar{\Lambda}\bar{A}(x)(\mu_+/\mu)p + \bar{C}(x)\delta\bar{\lambda} = -(\mu - \mu_+)e \\ \Rightarrow \quad &\bar{A}(x)^T(\mu_+/\mu)p = -\bar{\Lambda}^{-1}\left[(\mu - \mu_+)e + \bar{C}(x)\delta\bar{\lambda}\right]. \end{aligned} \tag{34}$$

Since $\bar{C}(x)\delta\bar{\lambda} = o(\mu)$ from (32) and (31), we have

$$\bar{\Lambda}^{-1}\left[(\mu - \mu_+)e + \bar{C}(x)\delta\bar{\lambda}\right] \geq \gamma^{-1}[(1-\rho)\mu + o(\mu)]e \geq \gamma^{-1}\frac{1-\rho}{2}\mu e, \tag{35}$$

for sufficiently small μ, so the right-hand side of (34) is $\Theta(\mu)$.

If we had $(\mu_+/\mu)p = o(\mu)$, then the right-hand side of (34) would be $o(\mu)$, which would contradict (35). Hence $(\mu_+/\mu)p = \Theta(\mu)$, and so, from (30), the two directions $x(\mu_+) - x_w(\mu)$ and $(\mu_+/\mu)p$ are asymptotically identical.

We state our main result as a theorem.

Theorem 3.2 *Assume that the objective function $f(x)$ of problem (1) is linear and x is a point near the minimizer $x(\mu)$ of $P(\ .\ ,\mu)$ in the sense that w defined from (5) satisfies (29). If μ_+ satisfies the second part of (29), then the Newton direction p in (14) for the logarithmic barrier function P is asymptotically the exact direction to the minimum $x(\mu_+)$ of $P(\ .\ ,\mu_+)$ in the sense that*

$$x + \frac{\mu_+}{\mu}p - x(\mu_+) = o(x - x(\mu_+)).$$

Even if μ_+ is chosen very small ($\mu_+ \ll \mu$), Theorem 3.2 by itself is not sufficient to guarantee that the Newton direction will bring us close to the optimal solution x^* of (1). We need to check that the line segment defined by

$$x + \alpha(\mu_+/\mu)p, \qquad \alpha \in [0, 1], \tag{36}$$

lies within the feasible region, except perhaps for values of α close to 1. It suffices to check that the constraints that are active at x^* are satisfied. For any $i = 1, \ldots, p$, we have from

(34), the definition $\lambda_i = \mu/c_i(x)$, and the estimate $|c_i(x)\delta\lambda_i| = o(\mu)$ that

$$\begin{aligned} c_i(x) + \alpha\nabla c_i(x)^T(\mu_+/\mu)p &= \frac{\mu}{\lambda_i} - \frac{\alpha}{\lambda_i}[(\mu - \mu_+) + c_i(x)\delta\lambda_i] \\ &= \frac{1}{\lambda_i}[(1-\alpha)\mu + \alpha\mu_+ + o(\mu)]. \end{aligned}$$

Now, using the bound (33), the estimate $\|(\mu_+/\mu)p\| = \Theta(\mu)$, and the smoothness of c_i, we have

$$\begin{aligned} c_i(x + \alpha(\mu_+/\mu)p) &= c_i(x) + \alpha\nabla c_i(x)^T(\mu_+/\mu)p + O(\mu^2) \\ &= \frac{1}{\lambda_i}[(1-\alpha)\mu + \alpha\mu_+ + o(\mu)] + O(\mu^2) \\ &\geq \frac{1}{\gamma}[(1-\alpha)\mu + o(\mu)]. \end{aligned} \tag{37}$$

Hence, if the segment (36) crosses the i-th constraint, we have $c_i(x + \alpha(\mu_+/\mu)p) = 0$ for some $\alpha \in [0,1]$. Because of (37), this is possible only when $1 - \alpha = o(\mu)/\mu$. (In the case of $w = 0$, that is, $x = x(\mu)$ exactly on the central path, this estimate can be strengthened to $1 - \alpha = O(\mu)$.) Hence, near-unit steps can eventually be taken without violating the constraints.

4 Discussion

In this paper we have analyzed the first Newton step that a logarithmic barrier method takes after decrementing the barrier parameter. Assuming linearity of the objective (without which the direction may be no good at all), we have quantified the accuracy of the direction in terms of the Euclidean norm of the gradient $\|w\|$ and the barrier parameter μ. We conjecture that condition $\|w\| = o(\mu^{-1/2})$ can be weakened slightly if we use a more appropriate norm, such as $||w||_{\nabla^2 P(x,\mu)^{-1}} = (w^T\nabla^2 P(x,\mu)^{-1}w)^{1/2}$. However, our final example (below) illustrates that a rather strong condition on w is necessary indeed to maintain the validity of Theorem 3.2.

Finally, we point out that the second part of condition (29) can be dropped if we work with the modified barrier function (6). Having obtained a sufficiently accurate approximation to the minimizer $x(\mu)$ of $P(\cdot;\mu)$, we may set w as in (5). Then, updating μ to μ_+, we can take a Newton step for the *modified* barrier function $P_w(\cdot;\mu_+)$ toward $x_w(\mu_+)$. Theorem 3.2 holds for this Newton step, even if the second part of condition (29) does not hold. (The Newton direction for the modified barrier function generated in the above way coincides with the primal affine scaling direction.) The analysis of the Newton step for the modified barrier function is similar to the analysis above. Since the barrier function (3) is more commonly used, however, we confined our main result to this function.

- **Example:** Finally, we formulate our introductory example to demonstrate the asymptotic equivalence (28). By introducing an artificial variable, we transform (4) to

$$\min x_3 \text{ subject to } (x_1, x_2) \geq 0, \; x_3 \geq x_1 + \tfrac{1}{2}(x_2 - 1)^2.$$

Then the barrier function $P(x; \mu)$ is

$$P(x; \mu) = x_3 - \mu \ln(x_3 - x_1 - \tfrac{1}{2}(x_2 - 1)^2) - \mu \ln x_1 - \mu \ln x_2,$$

and we find from $\nabla P(x(\mu); \mu) = 0$ that

$$x(\mu) = \begin{pmatrix} \mu \\ \frac{1}{2}(1 + \sqrt{1+4\mu}) \\ 1 + 4\mu - \sqrt{1+4\mu} \end{pmatrix} \approx \begin{pmatrix} \mu \\ 1 + \mu \\ 2\mu \end{pmatrix},$$

so

$$x(\mu) - x^* \approx \mu(1, 1, 2)^T. \tag{38}$$

A little manipulation shows that for $\mu_+ \in (0, \mu)$ and $x = x(\mu)$, we have

$$\nabla P(x; \mu_+) = \begin{pmatrix} 0 \\ 0 \\ 1 - \mu_+/\mu \end{pmatrix}, \qquad \nabla^2 P(x; \mu_+) \approx \frac{\mu_+}{\mu^2} \begin{bmatrix} 2 & \mu & -1 \\ \mu & \mu & -\mu \\ -1 & -\mu & 1 \end{bmatrix},$$

so the Newton step p is approximately

$$p \approx -\frac{\mu}{\mu_+}(\mu - \mu_+) \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}. \tag{39}$$

As claimed, (38) and (39) are consistent with (28).

We conclude this example by illustrating the inaccuracy in the direction p for centered and noncentered points for this particular example. We calculate the direction accuracy

$$\left\| \frac{p}{\|p\|} - \frac{x - x^*}{\|x - x^*\|} \right\|$$

for various values of μ. For each μ, we choose x to be the exactly centered point $x = x(\mu)$ and also the noncentered point

$$\hat{x}(\mu) = \begin{pmatrix} \mu/2 \\ 1 + 2\mu \\ 4\mu \end{pmatrix}.$$

Results are shown in Tables 1 and 2. Note that the directions line up as $\mu \downarrow 0$ in Table 1, while Table 2 shows that the noncentrality of $\hat{x}(\mu)$ introduces a systematic error into the Newton direction p that does not vanish even when $\hat{x}(\mu)$ is extremely close to x^*.

Table 1: Direction accuracy from centered points $x = x(\mu)$

μ	μ_+	Relative Accuracy
1	10^{-1}	$.60 \times 10^{-1}$
1	10^{-2}	$.73 \times 10^{-1}$
10^{-2}	10^{-3}	$.39 \times 10^{-2}$
10^{-2}	10^{-4}	$.43 \times 10^{-2}$
10^{-4}	10^{-5}	$.41 \times 10^{-4}$
10^{-4}	10^{-6}	$.45 \times 10^{-4}$

Table 2: Direction accuracy from non-centered points $x = \hat{x}(\mu)$

μ	μ_+	Relative Accuracy
10^{-2}	10^{-3}	$.83 \times 10^{-1}$
10^{-2}	10^{-4}	$.97 \times 10^{-1}$
10^{-4}	10^{-5}	$.88 \times 10^{-1}$
10^{-4}	10^{-6}	.10
10^{-6}	10^{-7}	$.88 \times 10^{-1}$
10^{-6}	10^{-8}	.10

Acknowledgments

Part of this work was done while the first author was visiting the Institute of Statistical Mathematics in Tokyo, Japan; thanks are due to Shinji Mizuno for his warm hospitality. He is also grateful to Takashi Tsuchiya for helpful discussions about the affine scaling direction and to Professor Kojima for his help with office and computing facilities at the Tokyo Institute of Technology.

References

[1] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968.

[2] S. Lang, *Real Analysis* (2nd ed.), Addison-Wesley, Reading, Massachusetts, 1983.

[3] M. H. Wright, *Why a pure primal Newton barrier step may be infeasible*, Numerical Analysis Manuscript 93–02, AT&T Bell Laboratories, March 1993. To appear in *SIAM Journal on Optimization.*