Why a Pure Primal Newton Barrier Step May Be Infeasible

Margaret H. Wright
AT&T Bell Laboratories
Murray Hill, New Jersey 07974

Abstract

Modern barrier methods for constrained optimization are sometimes portrayed conceptually as a sequence of inexact minimizations, with only a very few Newton iterations (perhaps just one) for each value of the barrier parameter. Unfortunately, this rosy image does not accurately reflect reality when the barrier parameter is reduced at a reasonable rate. We present local analysis showing why a pure Newton step in a long-step barrier method for nonlinearly constrained optimization may be seriously infeasible, even when taken from an apparently favorable point. The features described are illustrated numerically and connected to known theoretical results for convex problems satisfying self-concordancy assumptions. We also indicate the contrasting nature of an approximate step to the desired minimizer of the barrier function.

1. Introduction

1.1. Background

Interior methods, most commonly based on barrier functions, have been applied with great practical success in recent years to many constrained optimization problems, especially linear and quadratic programming, and their popularity continues to grow. For general nonlinearly constrained problems, an obvious approach is to use Newton's method for unconstrained minimization of the classical logarithmic barrier function.

For some special problem classes, various authors have proved that a pure Newton step is guaranteed to remain feasible and to produce a reduction in the barrier function when a distance measure for the current point—usually, a particular norm of the Newton step—is small enough. Such a characterization was given for linear programming problems in [Gon91] and [RV89]. Similar criteria for the Newton step in quadratic and certain convex nonlinear programs are developed in, for example, [NN89], [ADRT90], [Jar90], [DRT92], and [DH92]. The results in these papers do not, however, explain why the pure Newton step is unacceptable when the given norm is not sufficiently small.

This paper analyzes why a pure Newton step for a typical long-step "primal" barrier subproblem, i.e. an unconstrained minimization subproblem expressed in the original problem variables, is likely to be infeasible, even under circumstances that appear at first to be favorable—in particular, when the current point lies on the barrier trajectory. Broadly speaking, infeasibility arises because of two factors: the asymptotic role of the active constraint Jacobian matrix in the barrier gradient and Hessian; and the relationship among the optimal multipliers, the active constraint values and the barrier parameter.

1.2. Notation and assumptions

The problem of interest is

$$\underset{x \in \mathcal{R}^n}{\text{minimize}}\ f(x) \quad \text{subject to} \quad c_j(x) \geq 0,\ \ j = 1, \ldots, m, \tag{1.1}$$

where f and $\{c_j\}$ are smooth.

Much of our notation is standard. A local solution of (1.1) is denoted by x^*; $g(x)$ is the gradient of $f(x)$, and $H(x)$ its (symmetric) Hessian; $a_j(x)$ and $H_j(x)$ are the gradient and Hessian of $c_j(x)$; $A(x)$ is the $m \times n$ Jacobian matrix of the constraints, with jth row $a_j(x)^T$. The *Lagrangian function* associated with (1.1) is $L(x, \lambda) = f(x) - \lambda^T c(x)$. The Hessian of the Lagrangian with respect to x is $\nabla^2 L(x, \lambda) = H - \sum_{j=1}^m \lambda_j H_j(x)$.

We let $\hat{m}$ denote the number of constraints active at x^*, and $\mathcal{A}$ the set containing the indices of the active constraints. (It will usually be assumed that $\hat{m} > 0$.) Our discussion will consider only strictly feasible points, so that "active" and "inactive" refer to properties of constraints at x^*, and $\mathcal{A}$ is fixed for any given problem. At the point x, $\hat{A}(x)$ (the "Jacobian of the active constraints") is the $\hat{m} \times n$ matrix whose jth row is the gradient of the jth active constraint evaluated at x. The matrix $Z(x)$ refers to a matrix whose columns form an orthonormal basis for the null space of $\hat{A}(x)$, so that $\hat{A}(x)Z(x) = 0$ and $Z(x)^T Z(x) = I$.

Standard sufficient optimality conditions are assumed to hold at x^*:

(i) $g(x^*) = A^T(x^*)\lambda^*$, where λ^* is called the Lagrange multiplier vector;

(ii) $\lambda_j^* c_j(x^*) = 0$ for $j = 1, \ldots, m$;

(iii) $\lambda_j^* > 0$ if $j \in \mathcal{A}$, i.e. strict complementarity holds at x^*;

(iv) $\hat{A}(x^*)$ has full row rank;

(v) $Z^{*T} W^* Z^*$ is positive definite, where Z^* denotes $Z(x^*)$ and $W^* = \nabla^2 L(x^*, \lambda^*)$.

Under these conditions, x^* is an isolated local constrained minimizer of (1.1) and λ^* is unique; see, for example, [FM68] or [Fle87].

The logarithmic barrier function associated with (1.1) is

$$B(x, \mu) = f(x) - \mu \sum_{j=1}^m \ln c_j(x), \tag{1.2}$$

where μ is a positive scalar called the *barrier parameter*. This barrier function is defined only at strictly feasible points; it will be assumed that at least one point $\bar{x}$ exists where $c(\bar{x}) > 0$.

The gradient of the barrier function (1.2), denoted by g_B, is

$$g_B(x, \mu) = g(x) - \sum_{j=1}^m \frac{\mu}{c_j(x)}\, a_j(x) = g(x) - \mu A^T(x) C^{-1}(x) e, \tag{1.3}$$

where $e = (1, 1, \ldots, 1)^T$. The final form in (1.3) uses the convention that an uppercase version of a letter denoting a vector means the diagonal matrix whose diagonal elements are those of the vector. The barrier Hessian, denoted by H_B, has the form

$$H_B(x, \mu) = H(x) - \sum_{j=1}^m \frac{\mu}{c_j(x)}\, H_j(x) + \mu A^T(x) C^{-2}(x) A(x). \tag{1.4}$$

Assumptions (i)–(v) are well known to imply that, for sufficiently small μ:

(a) an isolated local unconstrained minimizer, denoted by either $x(\mu)$ or x_μ, of the barrier function (1.2) exists, at which

$$g_B(x_\mu, \mu) = 0, \quad \text{so that} \quad g(x_\mu) = \sum_{j=1}^{m} \frac{\mu}{c_j(x_\mu)}\, a_j(x_\mu); \tag{1.5}$$

(b) $H_B(x_\mu, \mu)$ is positive definite, and its smallest eigenvalue is bounded away from zero;

(c) if μ is regarded as a continuous parameter, x_μ defines a smooth trajectory converging to x^* as $\mu \to 0$. The points $\{x_\mu\}$ are said to lie on the *barrier trajectory*.

For proofs and additional details about the logarithmic barrier function, see, for example, [FM68], [Gon92] and [Wr92a].

We use the following standard notation; see [PS82]. Let ϕ be a function of a positive variable h, with p fixed. If there exists a constant $\kappa_u > 0$ such that $|\phi| \le \kappa_u h^p$ for all sufficiently small h, then $\phi = O(h^p)$. If there exists a constant $\kappa_l > 0$ such that $|\phi| \ge \kappa_l h^p$ for all sufficiently small h, then $\phi = \Omega(h^p)$. If there exist constants $\kappa_l > 0$ and $\kappa_u > 0$ such that $\kappa_l h^p \le |\phi| \le \kappa_u h^p$ for all sufficiently small h, then $\phi = \Theta(h^p)$.

2. The Primal Newton Barrier Direction

2.1. Newton's method applied to barrier functions

Suppose that we wish to minimize the barrier function (1.2) using Newton's method, and that the barrier Hessian is positive definite. Let x be the current iterate; the next iterate is then defined as

$$x + \alpha p_N, \quad \text{where} \quad H_B p_N = -g_B, \tag{2.1}$$

with H_B and g_B evaluated at x. The vector p_N is called the (primal) *Newton direction*; the modifier "primal" is used to emphasize that only the x variables are treated as independent.

In unconstrained minimization, the positive step α in (2.1) is chosen using a line search to produce a sufficient decrease in the function being minimized. For the barrier function (1.2), α must also retain strict feasibility in the next iterate, so that $c(x + \alpha p_N) > 0$.

An iteration for which $\alpha = 1$ in (2.1) is said to involve a *pure* Newton step. When the Hessian at the solution is positive definite, a successful application of Newton's method typically takes only pure Newton steps near the solution; choosing $\alpha = 1$ in this region is known to produce quadratic convergence as well as a sufficient decrease (see, for example, [DS83], [Fle87]).

Despite the favorable behavior of Newton's method for general problems, Newton's method has been known for many years to be problematical when applied to barrier functions because of inherent ill-conditioning in the Hessian. In [Mur71, Loot69], it was shown that, when $0 < \hat{m} < n$, the barrier Hessian is ill-conditioned at points on the barrier trajectory for sufficiently small μ, and is asymptotically singular. (This ill-conditioning is one of the factors that led to the decline in popularity of barrier methods in the 1970s.) Recently, it was proved in [Wr92b] that the barrier Hessian is ill-conditioned in an entire neighborhood of x^*. Although this property might appear to imply that the Newton direction cannot be computed without substantial numerical error, it is also shown in [Wr92b] how a highly accurate approximation of the Newton step can be calculated near x^*. "Active-set" strategies for overcoming the ill-conditioning for nonlinearly constrained problems are discussed in [Wr76] and [NS91]; see [GMPS91] for a technique designed for linear and quadratic programs in standard form. Unfortunately, accurate calculation of the Newton direction does not guarantee an effective algorithm, as we shall show in the remainder of this paper.

2.2. Long- and short-step methods

A classical barrier algorithm (see, e.g., [FM68]) typically calculates accurate minimizers of the barrier function for a decreasing sequence of barrier parameters, and so moves from $x(\tilde{\mu})$ to $x(\mu)$, where $\tilde{\mu}$ exceeds μ by some "reasonable" factor, say 10. In more recent practical algorithms, the idea is to improve efficiency by performing only an "inexact" minimization of the barrier function for each particular barrier parameter. For any given value μ, Newton iterations of the form (2.1) are executed until some measure of improvement has been achieved; the barrier parameter is then reduced and the process repeated. The hope is that only a very small number of Newton iterations (perhaps even one) will be needed for each value of barrier parameter.

The complexity analyses of [Gon91], [RV89], [NN89], [ADRT90], [Jar90], [DRT92] and [DH92] (among others) reflect a broad classification of barrier algorithms as "short-step" and "long-step"; complexity results of this form are surveyed in [Gon92] and [DH92]. In short-step methods, only a single pure Newton step is performed for each value of the barrier parameter, which is then multiplied by a factor less than but very close to one, say $1 - 1/(9\sqrt{n})$. In long-step methods, the barrier parameter is reduced by a more generous factor, say 1/10, but the analysis assumes that several Newton steps (some involving a line search) are carried out for a given barrier parameter. Practical methods for nonlinearly constrained problems are invariably long-step methods.

Nonlinear problems are treated in [NN89], [Jar90], [DRT92] and [DH92] under certain assumptions on the problem functions—for example, convexity and κ-self-concordancy. (For many problems, $\kappa = 1$.) The complexity of long-step methods for suitable convex nonlinear programs can then be analyzed using the "H-norm", a distance measure defined using the positive definite barrier Hessian:

$$\|p\|_H^2 = \frac{1}{\mu}\, p^T H_B p; \tag{2.2}$$

note that our barrier function (1.2) differs by a factor of $1/\mu$ from those in the cited papers. When p is the Newton direction p_N, it follows from (2.1) that the H-norm of p_N satisfies

$$\|p_N\|_H^2 = -\frac{1}{\mu}\, p_N^T g_B. \tag{2.3}$$

The proofs in [NN89], [Jar90], [DRT92] and [DH92] show that a pure Newton step whose H-norm is sufficiently small—say, less than $1/(3\kappa)$—is guaranteed to be strictly feasible and to reduce the barrier function. However, these results do not explain *why* a pure Newton step may be unsuccessful in a long-step method for a general problem; we now broadly analyze the reasons.

2.3. Properties of the barrier gradient and Hessian

It is assumed henceforth that we are performing Newton's method to minimize $B(x,\mu)$ for a general nonlinear problem, that the current iterate is x, and that $\hat{m} > 0$. (Because we do *not* assume convexity and self-concordancy, only local results can be obtained.) The barrier Hessian is presumed to be positive definite at any point of interest; this property is assured for points sufficiently close to the barrier trajectory (see result (b) in Section 1.2).

We consider strictly feasible points x "close" to x^* in the sense that

$$c(x) > 0 \quad \text{and} \quad \|x - x^*\| \le \delta \tag{2.4}$$

for suitable small δ. For small enough δ in (2.4), the inactive constraints at x are $\Theta(1)$. Under assumptions (i)–(v) of Section 1.2, it can be shown that $\|x^* - x(\mu)\| = \Theta(\mu)$; see, e.g., [FM68] or [Wr92a] for details. Since our intent is to move from x toward $x(\mu)$, $x(\mu)$ should be closer to x^* than x. We thus assume that

$$\mu = O(\delta). \tag{2.5}$$

Assuming that x satisfies (2.4) and μ satisfies (2.5), we first examine the structure of the barrier gradient (1.3). Full column rank of $\hat{A}^T(x^*)$, continuity of $g(x)$ and $A(x)$, and optimality conditions (i)–(ii) of Section 1.2 imply that the objective gradient satisfies

$$g(x) = \hat{A}^T(x)\hat{\lambda}^* + O(\delta), \quad \text{so that} \quad g(x) \approx \hat{A}^T(x)\hat{\lambda}^*. \tag{2.6}$$

In addition, the elements of $C(x)$ corresponding to inactive constraints are $\Theta(1)$, so that the quantity $\mu/c_j(x)$ in (1.3) is $\Theta(\mu)$ when constraint j is inactive. We conclude from the form of (1.3) and (2.6) that, near x^* (not necessarily near the barrier trajectory), the barrier gradient lies almost entirely in the range of $\hat{A}^T(x)$:

$$g_B \approx \hat{A}^T\hat{\lambda}^* - \mu\hat{A}^T\hat{C}^{-1}e. \tag{2.7}$$

Now consider the barrier Hessian (1.4). Near x^*, the portion of the matrix $\mu A^T C^{-2} A$ corresponding to inactive constraints is $\Theta(\mu)$. In analyzing the remainder of the barrier Hessian, we make the further assumption that x is close enough to the barrier trajectory so that the smallest active constraint value is not "too small" compared to μ; formally, this property means that

$$\min_{j\in\mathcal{A}} c_j(x) = \Omega(\mu), \quad \text{so that} \quad \max_{j\in\mathcal{A}} \frac{\mu}{c_j(x)} = O(1). \tag{2.8}$$

When (2.8) applies at x,

$$\|H(x) - \sum_{j=1}^{m} \frac{\mu}{c_j(x)}\, H_j(x)\| = O(1),$$

since the quotient μ/c_j is $\Theta(\mu)$ for inactive constraints and $O(1)$ for active constraints. Finally, the matrix $\mu\hat{A}^T\hat{C}^{-2}\hat{A}$ is $O(1/\mu)$ and hence dominates the barrier Hessian, which accordingly resembles a "large" matrix whose column space is the range of $\hat{A}^T$:

$$H_B \approx \mu\hat{A}^T\hat{C}^{-2}\hat{A}. \tag{2.9}$$

A more detailed discussion of the nature of the barrier Hessian under various assumptions is given in [Wr92b].

2.4. Approximating the Newton equations

We have just derived approximate expressions for g_B and H_B that apply when x and μ satisfy (2.4), (2.5) and (2.8). If we use (2.9) and (2.7) in (2.1), the Newton equations "look like" the following relation, which involves only vectors in the range of $\hat{A}^T$:

$$\mu\hat{A}^T\hat{C}^{-2}\hat{A}p_N \approx -\hat{A}^T\hat{\lambda}^* + \mu\hat{A}^T\hat{C}^{-1}e. \tag{2.10}$$

Since $\hat{A}^T$ has full column rank at x^*, it also has full column rank near x^*, and may be cancelled from both sides of (2.10). The Newton equations are thus (approximately)

$$\mu\hat{C}^{-2}\hat{A}p_N \approx -\hat{\lambda}^* + \mu\hat{C}^{-1}e. \tag{2.11}$$

Suppose that the jth constraint is active. The corresponding Newton "almost-equation" from (2.11) is

$$\begin{aligned} \frac{\mu}{c_j^2}\, a_j^T p_N &\approx -\lambda_j^* + \frac{\mu}{c_j}, \quad \text{or} \\ a_j^T p_N &\approx c_j - \frac{c_j^2\lambda_j^*}{\mu}. \end{aligned} \tag{2.12}$$

We now show why, if x is a previous point on the trajectory (or close to such a point), a pure Newton step is likely to be infeasible. Suppose that x is very close to to $x(\tilde{\mu})$ for some suitably small but "old" barrier parameter $\tilde{\mu}$, where $\tilde{\mu} > \mu$. The full rank of $\hat{A}(x^*)$, optimality conditions (i)–(ii) and relation (1.5) imply that

$$\left| \frac{\tilde{\mu}}{c_j} - \lambda_j^* \right| = O(\tilde{\mu}), \tag{2.13}$$

which means that

$$\frac{\tilde{\mu}}{c_j} \approx \lambda_j^*.$$

Substituting for λ_j^* in (2.12), we obtain a relation that holds approximately for the Newton direction calculated at $x(\tilde{\mu})$ with barrier parameter μ:

$$a_j^T p_N \approx -c_j \left(\frac{\tilde{\mu}}{\mu} - 1 \right). \tag{2.14}$$

When $\tilde{\mu}$ exceeds μ by some reasonable factor, i.e. the ratio $\tilde{\mu}/\mu$ is greater than (say) 2, the relationship (2.14) strongly suggests that $x + p_N$ will be infeasible, for the following reason. The step p_l from x to a zero (the boundary) of the locally linearized jth constraint satisfies $c_j + a_j^T p_l = 0$, which represents a relation similar in form to (2.14), but with a coefficient of unity on $-c_j$:

$$a_j^T p_l = -c_j.$$

Since the desired minimizer $x(\mu)$ is strictly inside the boundary, the step p_μ from a point near $x(\tilde{\mu})$ to $x(\mu)$ must move toward but "stop short" of the boundary. One would therefore expect that, for each active constraint j,

$$a_j^T p_\mu \approx -\gamma c_j, \quad \text{with} \quad 0 \le \gamma < 1. \tag{2.15}$$

In contrast, the factor multiplying $-c_j$ on the right-hand side of relation (2.14) is *larger* than one. Hence the Newton step is likely to move *beyond* the boundary, and consequently to be infeasible.

Suppose, for example, that μ is smaller than $\tilde{\mu}$ by a factor of 10; this would be typical in a practical (long-step) barrier algorithm. According to (2.14), the Newton direction satisfies

$$a_j^T p_N \approx -9 c_j, \tag{2.16}$$

and will thus tend to produce substantial infeasibility.

2.5. A numerical example

Consider the following (non-convex) numerical example:

$$\begin{array}{rll} \text{minimize} & -\frac{1}{8}x_1 + 2x_2 - x_3 & \\ \text{subject to} & -\frac{1}{2}x_1^2 - x_2^2 - x_3^2 + 5\frac{1}{8} & \ge 0 \\ & x_2^3 + 1 & \ge 0 \\ & x_1^2 + x_2^2 + x_3 - \frac{1}{2} & \ge 0. \end{array} \tag{2.17}$$

The optimal solution of interest is $x^* = (\frac{1}{2}, -1, 2)^T$, where the first and second constraints are active, with Lagrange multiplier vector $\lambda^* = (\frac{1}{4}, \frac{1}{2}, 0)^T$. All calculations given in this paper were performed on a Silicon Graphics 4D/240S, using IEEE standard double-precision floating-point arithmetic (around sixteen decimal digits). All displayed numbers are correctly rounded to the number of digits shown.

Let the vector $d(x, \mu)$ be defined as $d_j(x, \mu) = \mu / c_j(x)$, $j = 1, \ldots, m$. For $\mu = 10^{-3}$,

$$x_\mu = \begin{pmatrix} 0.50108 \\ -0.99933 \\ 1.9992 \end{pmatrix}, \quad c(x_\mu) = \begin{pmatrix} 3.9969 \times 10^{-3} \\ 1.9964 \times 10^{-3} \\ 2.7489 \end{pmatrix}, \quad d(x_\mu) = \begin{pmatrix} 0.25019 \\ 0.50089 \\ 3.6378 \times 10^{-4} \end{pmatrix}.$$

For $\mu = 10^{-4}$,

$$x_\mu = \begin{pmatrix} 0.50011 \\ -0.99993 \\ 1.9999 \end{pmatrix}, \quad c(x_\mu) = \begin{pmatrix} 3.9997 \times 10^{-4} \\ 1.9996 \times 10^{-4} \\ 2.7499 \end{pmatrix}, \quad d(x_\mu) = \begin{pmatrix} 0.25002 \\ 0.50009 \\ 3.6365 \times 10^{-5} \end{pmatrix}.$$

For $\mu = 10^{-5}$,

$$x_\mu = \begin{pmatrix} 0.500010 \\ -0.999993 \\ 1.99999 \end{pmatrix}, \quad c(x_\mu) = \begin{pmatrix} 3.99997 \times 10^{-5} \\ 1.99996 \times 10^{-5} \\ 2.74999 \end{pmatrix}, \quad d(x_\mu) = \begin{pmatrix} 0.250002 \\ 0.500009 \\ 3.63638 \times 10^{-6} \end{pmatrix}.$$

If we choose x as $x(10^{-3})$, and pick $\mu = 10^{-4}$, so that $\tilde{\mu}/\mu = 10$, we have

$$p_N = \begin{pmatrix} 2.1584 \times 10^{-2} \\ -6.0164 \times 10^{-3} \\ 3.2153 \times 10^{-3} \end{pmatrix}, \quad \hat{A} p_N = \begin{pmatrix} -3.5696 \times 10^{-2} \\ -1.8025 \times 10^{-2} \end{pmatrix},$$

which gives

$$\frac{a_1(x)^T p_N}{c_1(x)} = -8.931, \quad \frac{a_2(x)^T p_N}{c_2(x)} = -9.029, \tag{2.18}$$

as predicted by (2.16). Taking a step of unity along the Newton direction leads, as expected, to an infeasible point. Relationship (2.14) also applies for $x = x(10^{-3})$ and $\mu = 10^{-5}$, with $\tilde{\mu}/\mu = 100$; in this case,

$$\frac{a_1(x)^T p_N}{c_1(x)} = -97.8, \quad \frac{a_2(x)^T p_N}{c_2(x)} = -99.5.$$

When the barrier Hessian is positive definite, the H-norm of p_N (2.3) used in complexity analyses should be helpful as a local estimate of the "quality" of the Newton step. Applying (2.3) in problem (2.17) at $x(10^{-3})$ with $\mu = 10^{-4}$, we find that

$$\|p_N\|_H^2 = 161.5,$$

which is certainly *not* small.

Although cutting back the step taken along the Newton direction will eventually restore strict feasibility, such a strategy may end up at a point very near the boundary, where the Hessian will tend to be even more ill-conditioned (see [Wr92b]) and where the H-norm of the next Newton step is unlikely to be small. A special line search (see, e.g., [MW92]) can help to produce a "good" next iterate. Our first preference, however, would be to find a direction along which a unit step can be taken with impunity in a close neighborhood of the solution, where all the asymptotic properties of a "good" Newton method should apply.

2.6. The barrier trajectory direction

In light of the unfavorable relation (2.14), an obvious question is what value $a_j^T p$ "should" have if p is the step to $x(\mu)$ from $x(\tilde{\mu})$ rather than the Newton step. Recall that, along the barrier trajectory, the ratio μ/c_j converges to λ_j^*. We would like to choose p such that

$$\frac{\mu}{c_j(x+p)} \approx \lambda_j^*. \tag{2.19}$$

If, for an *active* constraint j, a good estimate of λ_j^* is available, say λ_j, a search direction p could be required to satisfy a relation like (2.19), but involving a linearized version of the constraint:

$$\frac{\mu}{c_j + a_j^T p} \approx \lambda_j, \quad \text{or} \quad a_j^T p \approx -c_j + \frac{\mu}{\lambda_j}. \tag{2.20}$$

Linear constraints based on (2.20) appear in the barrier trajectory algorithm proposed by [Wr76], in which the search direction solves an equality-constrained quadratic program with constraints

$$\hat{A}p = -c + \mu\hat{\Lambda}^{-1}e,$$

where $\hat{A}$ is a prediction of the active set and $\hat{\Lambda}$ is a set of associated multiplier estimates.

To see the form of the relationship (2.20) for points on the trajectory, suppose that the current point is $x(\tilde{\mu})$ for some suitably small $\tilde{\mu}$. The obvious candidate for the jth multiplier estimate is $\tilde{\mu}/c_j$; see (2.13). With this choice for λ_j, (2.20) becomes

$$a_j^T p = -c_j\Big(1 - \frac{\mu}{\tilde{\mu}}\Big), \tag{2.21}$$

and p is likely to stop short of the boundary (as in (2.15)) rather than produce infeasibility (as for the Newton direction; see (2.14)).

For problem (2.17), the step p_{34} from $x(10^{-3})$ to $x(10^{-4})$ and the step p_{35} from $x(10^{-3})$ to $x(10^{-5})$ are

$$p_{34} = \begin{pmatrix} -9.6706\times 10^{-4} \\ -5.9926\times 10^{-4} \\ 7.2091\times 10^{-4} \end{pmatrix} \quad\text{and}\quad p_{35} = \begin{pmatrix} -1.0651\times 10^{-3} \\ -6.5925\times 10^{-4} \\ 7.9316\times 10^{-4} \end{pmatrix}.$$

When $x = x(10^{-3})$ and $\mu = 10^{-4}$,

$$\frac{a_1(x)^T p_{34}}{c_1(x)} = -0.8996, \quad \frac{a_2(x)^T p_{34}}{c_2(x)} = -0.8993;$$

since $\mu/\tilde{\mu} = 0.1$, (2.21) holds approximately for the active constraints. Similarly, when $x = x(10^{-3})$ and $\mu = 10^{-5}$, with $\mu/\tilde{\mu} = .01$,

$$\frac{a_1(x)^T p_{35}}{c_1(x)} = -0.9896, \quad \frac{a_2(x)^T p_{35}}{c_2(x)} = -0.9893,$$

again approximating (2.21).

2.7. When a pure Newton step works well

Because the barrier Hessian is positive definite at $x(\mu)$, a pure Newton step must eventually be "good" when x is sufficiently close to $x(\mu)$. Given the poor results of Section 2.4 when x is taken as $x(\tilde{\mu})$ (a point that might intuitively appear to be close to $x(\mu)$), an obvious issue is the meaning of "sufficiently close". We have already mentioned the work of [NN89], [Jar90], [DRT92] and [DH92], where it is shown for certain convex problems that the Newton step is guaranteed to be successful when its H-norm is small. It is interesting to consider what information about the Newton direction can be deduced from the properties involving $\hat{A}$ given in Section 2.4.

Consider relation (2.12) for the Newton step in the form

$$\mu\Big(1 - \frac{a_j^T p_N}{c_j}\Big) \approx c_j\lambda_j^*.$$

If $|a_j^T p_N|$ is sufficiently small relative to c_j, so that

$$\left|\frac{a_j^T p_N}{c_j}\right| = \beta, \quad\text{where}\quad 0 < \beta < 1, \tag{2.22}$$

we may make the approximation

$$\frac{1}{1-\dfrac{a_j^T p_N}{c_j}} \approx 1 + \frac{a_j^T p_N}{c_j} + \dots. \tag{2.23}$$

When (2.23) holds, we may rewrite property (2.12) satisfied by the Newton direction p_N as

$$\mu \approx c_j \lambda_j^* \Big(1 + \frac{a_j^T p_N}{c_j}\Big) \approx c_j \lambda_j^* + \lambda_j^* a_j^T p_N,$$

or, after rearrangement,

$$a_j^T p_N \approx -c_j + \frac{\mu}{\lambda_j},$$

which is the same as relation (2.20) derived from properties of the barrier trajectory. Following this interpretation, the Newton direction can be a "good" approximation of the step to $x(\mu)$ only when the ratio $|a_j^T p_N/c_j|$ is small enough for all active constraints. This corresponds to the property that a unit step along the direction stops short of the boundary of the linearized active constraints; see (2.15). It should be observed that any property involving $a_j^T p$ for the active constraints can be viewed as a condition on the component of the search direction in the range space of $\hat{A}^T$, since $\hat{A}p$ is unaffected by the portion of p in the null space of $\hat{A}$.

A property with the same flavor as (2.22) can also be derived by considering an approximation to the H-norm of the Newton step. Because the barrier Hessian is dominated by $\mu \hat{A}^T \hat{C}^{-2} \hat{A}$ near the trajectory (see (2.9)), the H-norm of the Newton step satisfies

$$\frac{1}{\mu}\, p_N^T H_B p_N \approx p_N^T \hat{A}^T \hat{C}^{-2} \hat{A} p_N = \sum_{j \in \mathcal{A}} \Big(\frac{a_j^T p_N}{c_j}\Big)^2,$$

and will be "small" only when the ratio $a_j^T p_N/c_j$ is small in magnitude for every active constraint.

To illustrate these estimates, consider problem (2.17), with $\mu = 10^{-4}$ and a starting point of $x(10^{-3})$. At the first iteration, the ratio $|a_j^T p_N/c_j|$ is approximately 9 for both active constraints (see (2.18)), the squared H-norm of p_N is 161.5, and the pure Newton step is infeasible. For the second iteration, we have

$$\frac{a_1^T p_N}{c_1} = -3.40, \quad \frac{a_2^T p_N}{c_2} = -3.36, \quad \|p_N\|_H^2 = 2.29,$$

and once again the Newton direction is infeasible. At iteration 3, the estimates start to become "small", namely

$$\frac{a_1^T p_N}{c_1} = -0.339, \quad \frac{a_2^T p_N}{c_2} = -0.285, \quad \|p_N\|_H^2 = .209.$$

At this iteration (and all subsequent iterations for this value of μ), the pure Newton step is feasible and produces a sufficient decrease in the barrier function.

3. Conclusions

Complexity analyses of barrier methods for nonlinear convex programming (in particular, [NN89], [Jar90], [DRT92] and [DH92]) indicate that pure Newton steps cannot necessarily be taken successfully unless the barrier parameter is reduced by only a tiny amount. Using local analysis for general nonlinearly constrained problems, we have seen why, following a computationally reasonable (long-step) reduction in the barrier parameter from $\tilde{\mu}$ to μ, a pure Newton step is almost certain to be infeasible even when the initial point is $x(\tilde{\mu})$.

This situation seems unsatisfactory because asymptotic results (see Section 2.6) indicate that, under these circumstances, we should be able to calculate a good step toward the next barrier minimizer. One way to interpret the difficulty is that a formulation in terms of only the primal variables essentially ignores available information about the Lagrange multiplier estimates.

If one is willing to predict the active set, a barrier trajectory algorithm (see [Wr76] and [MW78]) might be appropriate. Strategies based on maintaining multiplier estimates for all the constraints also seem promising; for example, a recently proposed interior method for nonlinear convex programming [JS91] includes tests on the quality of multiplier estimates in its calculation of the search direction. Another strategy not requiring prediction of the active set is to apply Newton's method to solve some form of the nonlinear primal-dual equations, involving both x and multiplier estimates λ, that hold along the barrier trajectory. For example, the standard primal-dual equations appear in the algorithm proposed by McCormick [McC91] for convex programming:

$$g(x) = A(x)^T \lambda \qquad \text{and} \qquad c_j \lambda_j = \mu, \quad j = 1, \ldots, m.$$

An obvious advantage of a primal-dual formulation is that the associated matrix is not inherently ill-conditioned as the solution is approached. Primal-dual methods for general nonlinearly constrained problems will be considered in a future paper.

References

[ADRT90] K. M. Anstreicher, D. den Hertog, C. Roos, and T. Terlaky (1990), A long step barrier method for convex quadratic programming, Technical Report 90-53, Faculty of Mathematics and Informatics, Delft University of Technology, Delft, The Netherlands; to appear in *Algorithmica*.

[DH92] D. den Hertog (1992), *Interior Point Approach to Linear, Quadratic and Convex Programming: Algorithms and Complexity*, Ph. D. thesis, Delft University of Technology, Delft, The Netherlands.

[DRT92] D. den Hertog, C. Roos and T. Terlaky (1992), A large-step analytic center method for a class of smooth convex programming problems, *SIAM J. Optim.* 2, 55–70.

[DS83] J. E. Dennis, Jr. and R. B. Schnabel (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, New Jersey.

[FM68] A. V. Fiacco and G. P. McCormick (1968), *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons (New York). Republished in 1990 by SIAM (Philadelphia).

[Fle87] R. Fletcher (1987), *Practical Methods of Optimization* (second edition), John Wiley and Sons (Chichester).

[GMPS91] P. E. Gill, W. Murray, D. B. Ponceleón, and M. A. Saunders (1991), Solving reduced KKT systems in barrier methods for linear and quadratic programming, Report SOL 91-7, Department of Operations Research, Stanford University, Stanford, California.

[Gon91] C. C. Gonzaga (1991), Large step path-following methods for linear programming, part I: barrier function method, *SIAM J. Optim.* 1, 268–279.

[Gon92] C. C. Gonzaga (1992), Path following methods for linear programming, *SIAM Review* 34, 167–224.

[Jar90] F. Jarre, Interior-point methods for convex programming, Report SOL-16, Department of Operations Research, Stanford University, Stanford, California.

[JS91] F. Jarre and M. A. Saunders (1991), Practical aspects of an interior-point method for convex programming, Report SOL 91-9, Department of Operations Research, Stanford University, Stanford, California.

[Loot69] F. A. Lootsma (1969), Hessian matrices of penalty functions for solving constrained optimization problems, *Philips Res. Repts 24*, 322–331.

[McC91] G. P. McCormick (1991), The superlinear convergence of a nonlinear primal-dual algorithm, Report T-550/91, Department of Operations Research, George Washington University, Washington, DC.

[Mur71] W. Murray (1971), Analytical expressions for the eigenvalues and eigenvectors of the Hessian matrices of barrier and penalty functions, *J. Optim. Theory Appl.* 7, 189–196.

[MW78] W. Murray and M. H. Wright (1978), Projected Lagrangian methods based on the trajectories of penalty and barrier functions, Report SOL 78-23, Department of Operations Research, Stanford University, Stanford, California.

[MW92] W. Murray and M. H. Wright (1992), Line search procedures for the logarithmic barrier function, Numerical Analysis Manuscript 92-01, AT&T Bell Laboratories, Murray Hill, New Jersey. (To appear in *SIAM J. Optim.*)

[NS91] S. G. Nash and A. Sofer (1991), A barrier method for large-scale constrained optimization, Report 91-10, Department of Operations Research and Applied Statistics, George Mason University, Fairfax, Virginia.

[NN89] Y. Nesterov and A. Nemirovsky (1989), *Self-Concordant Functions and Polynomial-Time Methods in Convex Programming*, USSR Academy of Science (Moscow); to be published by SIAM (Philadelphia), 1993.

[PS82] C. R. Papadimitriou and K. Steiglitz (1982), *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall, Englewood Cliffs, New Jersey.

[RV89] C. Roos and J. P. Vial (1989), Long steps with the logarithmic penalty barrier function in linear programming, in *Economic Decision-Making: Games, Economics and Optimization* (J. Gabszevwicz, J. F. Richard and L. Wolsey, eds.), 433–441, Elsevier (Amsterdam).

[Wr76] M. H. Wright (1976), *Numerical Methods for Nonlinearly Constrained Optimization*, Ph. D. thesis, Stanford University, California.

[Wr92a] M. H. Wright (1992), Interior methods for constrained optimization, in *Acta Numerica 1992* (A. Iserles, ed.), Cambridge University Press (New York), 341–407.

[Wr92b] M. H. Wright (1992), Determining subspace information from the Hessian of a barrier function, Numerical Analysis Manuscript 92–02, AT&T Bell Laboratories, Murray Hill, New Jersey.