Line Search Procedures for the Logarithmic Barrier Function

Walter Murray
Department of Operations Research
Stanford University
Stanford, California 94305

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

Abstract

Barrier methods for constrained optimization, widely applied in the 1960's and 1970's, have recently enjoyed a revival of popularity. For problems involving nonlinear functions, some modern interior methods include a line search with respect to a logarithmic barrier function or related potential function. Standard line search procedures tend to be inefficient in this context for two reasons: an inappropriate choice of initial trial step, and poor approximation of the barrier function by low-order polynomial interpolants. This paper discusses line search strategies specifically designed for the logarithmic barrier function.

1. Barrier Methods for Constrained Optimization

Beginning in 1984 with the work of Karmarkar [Kar84], a resurgence of interest has taken place in barrier methods for constrained optimization. Classical barrier methods, which were popular in the 1960's and early 1970's, treat inequality constraints by creating a transformed function containing a positive singularity ("barrier") at the boundary of the feasible region.

Consider the constrained optimization problem

$$\underset{x\in\Re^n}{\text{minimize}}\ f(x) \quad \text{subject to} \quad c_i(x) \ge 0, \ \ i = 1, \ldots, m. \tag{1.1}$$

We assume that the objective function f and the set of constraint functions $\{c_i\}$ are smooth. A point $\bar{x}$ for which $c_i(\bar{x}) > 0$, $i = 1, \ldots, m$, is said to be *strictly feasible*.

Let x^* denote a local solution of (1.1). The corresponding logarithmic barrier function is

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

where μ is a positive scalar called the *barrier parameter*. Note that this barrier function is defined only at strictly feasible points. Let $x^*(\mu)$ denote an unconstrained minimizer of $B(x, \mu)$. Under certain conditions, it can be shown that

$$\lim_{\mu\to 0} x^*(\mu) = x^*.$$

For details about background and theory, see [FM68], [Fia79] and [Wr92].

Most interior methods proposed since 1984 are designed for the special cases of linear programming (LP) and convex quadratic programming, and the vast majority of papers on these topics have concentrated on proofs of polynomial complexity (see the extensive bibliographies in [GT89] and [Gonz92]). Within the last few years, several interior methods for convex nonlinear programming have been proposed and analyzed from the viewpoint of complexity; see, for example, [NN89], [DRT90], [Jar90], and [MS90]. On the practical side, barrier techniques are increasingly being applied to nonlinearly constrained problems of many varieties, including nonconvex and large-scale instances; see, for example, [JS91], [NS91] and [Ring91].

Barrier methods fell from favor in the 1970's partly because inherent ill-conditioning in the Hessian matrices created difficulties for standard unconstrained methods. A major issue today is the development of strategies for dealing with this ill-conditioning; for a discussion of some approaches, see [Wr92]. A second inconvenient feature of barrier functions is that they tend to cause inefficiencies for general-purpose line search techniques. In the next section, we discuss the role of a line search in optimization algorithms, and explain why special-purpose line search procedures may be helpful in dealing with barrier functions.

Some barrier methods are based on unconstrained minimization of a sequence of barrier functions (1.2) with decreasing values of μ. The line search procedures to be described also apply to linearly and nonlinearly constrained subproblems in which a barrier function serves as a merit function.

2. Line Searches for Barrier Functions

2.1. The role of the line search

In many optimization methods, a typical iteration (say, the kth) defines a new iterate x_{k+1} by taking a positive step α_k from the current point x_k along an n-dimensional search direction p_k, so that

$$x_{k+1} = x_k + \alpha_k p_k. \tag{2.1}$$

We shall assume henceforth that p_k is a descent direction for some differentiable function Φ, i.e., that $p_k^T \nabla\Phi(x_k) < 0$. The process of choosing α_k is usually called the *line search*.

Iterations of the form (2.1) arise in several contexts. In line search methods for unconstrained minimization of Φ, α_k is required to satisfy mathematical conditions of *sufficient decrease* with respect to properties of Φ at x_k (see, for example, [OR70], [DS83], [DS89], [Fle87], [GMW81]). Sufficient decrease conditions require more than simply a reduction in Φ; they demand that Φ decrease "enough" to satisfy certain requirements for convergence of $\{x_k\}$ (see [OR70] for detailed proofs).

Let $\phi(\alpha)$ denote the function

$$\phi(\alpha) = \Phi(x_k + \alpha p_k), \tag{2.2}$$

so that $\phi(\alpha)$ represents the variation in Φ as a function of the step α from x_k. Then

$$\phi(0) = \Phi(x_k), \quad \phi'(0) = p_k^T \nabla\Phi(x_k) \quad \text{and} \quad p_k^T \nabla\Phi(x_k + \alpha p_k) = \phi'(\alpha).$$

The property that p_k is a descent direction for Φ implies that $\phi'(0) < 0$. Two frequently used sufficient decrease conditions are posed in terms of constants σ_1 and σ_2 satisfying $0 < \sigma_1 < \sigma_2 < 1$. The step α_k is accepted only if

$$\phi(\alpha_k) < \phi(0) + \sigma_1 \alpha_k \phi'(0) \quad \text{and} \quad \phi'(\alpha_k) \geq \sigma_2 \phi'(0). \tag{2.3}$$

For other sufficient decrease conditions, see, for example, [OR70], [BN89], [DS83], [DS89], [Fle87] and [GMW81].

Some algorithms for constrained optimization include a line search in the sense that they require a sufficient decrease at each iteration in a merit function intended to measure progress toward the solution. Although the definition of the merit function may vary between iterations, sufficient decrease conditions such as (2.3) are applied during any given iteration. For simplicity, $\Phi(x)$ and the associated one-variable function $\phi(\alpha)$ will always denote the generic functions in which a sufficient decrease is sought.

2.2. The line search iteration

A suitable value of α_k is usually found by an iterative procedure that generates a sequence of trial steps $\{\alpha^{(j)}\}$ for $j = 0, \ldots,$ and terminates by accepting the first $\alpha^{(j)}$ for which the sufficient decrease conditions hold. A certain amount of work is associated with each trial point $\alpha^{(j)}$. For example, simply testing the sufficient decrease conditions (2.3) requires evaluation of Φ and $\nabla\Phi$ at $x_k + \alpha^{(j)} p_k$.

The choice of the initial step $\alpha^{(0)}$ in the line search almost always depends on the strategy that defines p_k. For Newton and quasi-Newton methods (the only methods considered here) the obvious "natural" initial step of unity ($\alpha^{(0)} = 1$) produces the correct convergence rate as well as a sufficient decrease for iterates near the solution (see [DM77]).

In the early years of line search methods for unconstrained optimization, most implementations included an "accurate" line search, i.e., α_k was a close approximation to a minimizer of ϕ. Today, however, it is customary to perform an "inexact" or "relaxed" line search, typified by a small value of σ_1 and a large value of σ_2 in (2.3) (say, $\sigma_1 = 10^{-4}$ and $\sigma_2 = 0.99$). An accurate line search nonetheless remains desirable in some situations, such as in large-scale problems where the linear algebra needed to compute p_k may be substantially more time-consuming than the evaluation of Φ at trial steps. See [DS83], [Fle87] and [GMW81] for additional discussion.

Typical relaxed sufficient decrease criteria are satisfied by a *range* of acceptable values for α_k. An effective line search should therefore produce a "good" step (broadly speaking, a "large" reduction in Φ) without requiring "too many" trial steps. The tradeoff is somewhat delicate, and the optimal balance is problem-dependent.

In the popular *Armijo* or *pure backtracking* line search ([OR70], [DS83]), each trial step $\alpha^{(j+1)}$, $j = 0, \ldots,$ is taken as a fixed fraction τ (typically $\tau = \frac{1}{2}$ or $\tau = \frac{1}{10}$) of the previous step $\alpha^{(j)}$. Under reasonably mild conditions, this procedure is guaranteed to produce a sufficient decrease (see, for example, [BN89]). The great advantage of an Armijo line search is its simplicity. In instances where the initial step is almost always accepted—for instance, in minimizing well behaved functions with Newton's method—an Armijo line search is also efficient. A pure backtracking procedure is inappropriate when an accurate line search is needed, since the fixed, strictly decreasing sequence $\{\tau^j \alpha^{(0)}\}$, $j = 1, \ldots,$ of trial steps is unaffected by the behavior of Φ along p_k.

A second standard line search technique is to define the trial steps as iterates from a procedure for minimizing $\phi(\alpha)$ within a specified interval of uncertainty. For unconstrained problems, the initial interval of uncertainty is bounded on the left by $\alpha = 0$ (where $\phi' < 0$) and on the right (usually) by a large step that represents the maximum permitted change in x during a particular iteration. The interval of uncertainty is modified to reflect information about ϕ accumulated as the line search proceeds. For example, if $\phi'(\alpha^{(0)})$ is positive, at least one minimizer must occur in the interval $(0, \alpha^{(0)})$. It should be stressed that univariate minimization is used only as a model subproblem for generating trial steps in the line search; the sequence $\{\alpha^{(j)}\}$ terminates as soon as the desired sufficient decrease conditions are satisfied.

For smooth functions, low-order polynomial interpolation is widely used to construct $\alpha^{(j+1)}$ based on previous trial steps and the associated values of ϕ (see [Br73], [DS83], [DS89], [Fle87], [GM74], [GMW81]). This approach is sometimes combined with backtracking (see, for example, [BN89] and [DS89]), and can be used when an accurate line search is desired. To ensure robustness, the interpolant step may be rejected in favor of a safeguarded step that produces a guaranteed reduction in the interval of uncertainty. This issue and related topics are discussed in [Br73], [DS89] and [GM74].

2.3. Specialized line searches for barrier functions

In the 1960's and 1970's, the first heyday of barrier methods, several special-purpose line search strategies were proposed. In [FlMc69], separate approximations are made to the objective function and the barrier term. In the approach of [LFR73], a special function is developed by substituting linearized versions of the objective function and constraints into the form of the barrier function.

The techniques of both [FlMc69] and [LFR73] apply to a broad class of barrier functions. In contrast, the approximating functions to be described in Section 4 are designed specifically for the logarithmic barrier function, and were first suggested in [Wr76] and [MW76].

More recently, line search strategies for barrier methods have received very little attention, primarily because an extremely simple line search appears to be entirely adequate for the predominant case of linear programming. The typical line search in barrier-related linear programming methods chooses α_k in almost all circumstances as a fixed fraction close to unity (such as 0.95 or 0.99) of the step from x_k along p_k to the boundary of the feasible region. (See the discussion and references in [Gonz92].) Broadly speaking, this strategy succeeds for the LP case because singularity at the boundary forces a minimizer to occur near the boundary in a function that would otherwise behave like a monotonically decreasing linear function. We discuss in Section 3 why a direct generalization of this approach is unlikely to be effective for nonlinear problems.

Several recent papers about interior methods for nonlinear problems have discussed conditions analogous to sufficient decrease that are imposed on the line search to obtain desirable complexity properties. See, for example, [ADRT90] and [Ans90].

Two aspects of the general-purpose line search procedures mentioned in Section 2.2 may be inefficient in barrier methods designed for nonlinear problems (including quadratic programming): the choice of the initial step $\alpha^{(0)}$, and the procedure for generating subsequent trial steps.

Within a barrier method, an initial step of unity may produce a point where at least one constraint is nonpositive. Retaining strict feasibility poses no difficulty when all the constraints are *linear* (as in linear and quadratic programming). For a linear constraint, the nonnegative step along p_k to the boundary of the feasible region ("the step to the nearest constraint") can be computed directly, and then imposed as the right endpoint of the interval of uncertainty in which α_k must lie. For nonlinear constraints, however, the situation is more complicated. Selection of the initial step will be discussed in Section 3.

The second issue involves the procedure for generating trial steps. Special techniques are indicated because low-order polynomial interpolants constructed near $\alpha = 0$ often fail to reflect the characteristic dramatic behavior of a barrier function near the singularity, particularly when the barrier parameter is small. The disparity in an example with $\mu = \frac{1}{2}$ is illustrated in Figure 1; as μ decreases, the differences become even more extreme. Special-purpose interpolating functions containing a logarithmic singularity are described in Section 4.

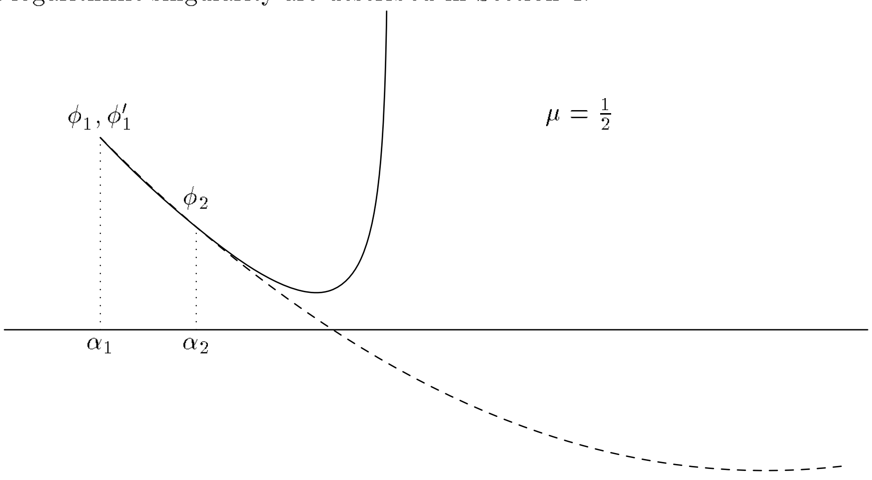

Figure 1. A barrier function $\phi(\alpha)$ with barrier parameter $\mu = \frac{1}{2}$ is shown as a solid line. The dashed line represents the quadratic function that interpolates $\phi(\alpha_1)$, $\phi'(\alpha_1)$ and $\phi(\alpha_2)$. The minimizer of the quadratic (at the extreme right of the figure) is a substantial (and infeasible) overestimate of the minimizer of the barrier function, even for this relatively large value of μ.

3. Choosing the Initial Trial Step

3.1. Difficulties arising from nonlinearity

Before discussing our approach to choosing the initial trial step, we stress that the issue is not simply retention of strict feasibility. If it were, a suitably tiny step along p_k would always suffice. As noted in Section 2.2, a practical line search strategy should retain efficiency throughout the overall minimization. In particular, the natural initial step of unity should be taken near the solution to retain the correct asymptotic convergence rate. If this step is infeasible, an Armijo-like strategy of trying successively smaller multiples of the initial step will "waste" constraint evaluations in the frequent situation when the step to the boundary of the feasible region is much smaller than unity. We therefore consider how the available information at $\alpha = 0$ can be applied to make a sensible choice for the initial step.

The difficulties arising from constraint nonlinearities are revealed by considering the following hypothetical generalization of the linear-constraint strategy: calculate γ_k, the smallest positive step along p_k such that some constraint vanishes at $x_k + \gamma_k p_k$, and choose the right endpoint of the initial interval of uncertainty as a value slightly less than γ_k. Although this idealized strategy would retain feasibility, its implementation is problematic for two reasons.

First, the question arises of how to calculate γ_k. For nonlinear constraints, γ_k must be determined iteratively—for example, by applying a numerical zero-finding procedure to compute a positive step along p_k (if any) for which each constraint vanishes, and taking γ_k as the smallest of these values. Since an arbitrary nonlinear function may "wiggle", however, a generic zero-finding procedure will not necessarily locate the *smallest* positive step along p_k to a constraint zero. Thus there is no guarantee that the correct value of γ_k has been obtained.

Second, it is unclear whether the expense of trying to obtain an accurate value of γ_k is worthwhile. Except for the special case of linear programming, there is no *a priori* reason to expect a barrier function to achieve its minimum close to the boundary. In fact, a longstanding "rule of thumb" for nonlinear problems is that a premature move too close to the boundary can lead to a long sequence of iterates "trapped" near the singularity by the need to decrease the barrier function at each iteration (see, for example, [FlMc69], [Wr76]). Hence an accurate calculation of γ_k could not only be expensive in terms of constraint evaluations, but is also unlikely to be cost-effective.

3.2. A computational procedure

Let α_u denote the standard choice of initial step for the algorithm used to define p_k. (For Newton and quasi-Newton methods, $\alpha_u = 1$.) Ideally, the step α_u is taken whenever possible to preserve the asymptotic rate of convergence. The strategy given here produces an easily-computed estimate of γ_k, which is then used to decide whether the initial trial step $\alpha^{(0)}$ should be taken as α_u or as some smaller value.

For simplicity of notation, we consider only a single iteration, and drop the subscript k, so that x denotes the current iterate, and so on. A stationary point α of the barrier function along p satisfies the following nonlinear equation:

$$p^T \nabla B(x + \alpha p) = p^T g(x + \alpha p) - \sum_{j=1}^{m} \frac{\mu}{c_j(x + \alpha p)} \, p^T a_j(x + \alpha p) = 0 \tag{3.1}$$

We now derive an estimate of the solution of (3.1) based on first-derivative information at x.

For each constraint j that is locally decreasing along p, we compute σ_j, the (necessarily positive) step to the zero of the linearized constraint along p:

$$\sigma_j = -\frac{c_j}{a_j^T p} \quad \text{for} \quad a_j^T p < 0. \tag{3.2}$$

If $a_j^T p \ge 0$ for all j, the initial step $\alpha^{(0)}$ is taken as the standard choice α_u. Otherwise, define $\bar{\jmath}$ and $\bar{\alpha}$ as

$$\bar{\alpha} = \sigma_{\bar{\jmath}} = \min \sigma_j. \tag{3.3}$$

If there is a tie, pick $\bar{\jmath}$ as the index with the most negative value of $a_{\bar{\jmath}}^T p$. In general, $\bar{\alpha}$ is an *estimate* of γ. (If all the constraints happen by chance to be linear, $\bar{\alpha}$ is the exact value γ.) Calculation of $\bar{\alpha}$ requires only information that is already available, namely the values of the constraints and their gradients at x.

An estimate α_b of the solution of (3.1) is obtained by: (i) ignoring all constraints except the $\bar{\jmath}$th (in effect, assuming that the influences of all other constraints are negligible compared to that of the nearest singularity), and (ii) making local linear approximations to f and $c_{\bar{\jmath}}$. (The latter approximations are exact if f and $c_{\bar{\jmath}}$ are linear.) These simplifications lead to the relation

$$g^T p - \frac{\mu}{c_{\bar{\jmath}} + \alpha_b a_{\bar{\jmath}}^T p}\, a_{\bar{\jmath}}^T p = 0,$$

which gives

$$\alpha_b = \frac{\mu}{g^T p} - \frac{c_{\bar{\jmath}}}{a_{\bar{\jmath}}^T p} = \bar{\alpha} + \frac{\mu}{g^T p}. \tag{3.4}$$

If $\bar{\alpha}$ were equal to the ideal value γ, a minimizer of the barrier function along p would necessarily lie to the left of $\bar{\alpha}$. To ensure that $\alpha_b < \bar{\alpha}$, relation (3.4) is used only when $g^T p < 0$, i.e., when the objective function is decreasing along p. If $g^T p \ge 0$, α_b is taken as a fraction of $\bar{\alpha}$, (say, $\alpha_b = .99\bar{\alpha}$).

The estimate α_b is next compared to the standard initial step α_u, and $\alpha^{(0)}$ is taken as the smaller of α_b and α_u:

$$\begin{aligned} \alpha^{(0)} &= \alpha_b \quad \text{if } \alpha_b < \alpha_u; \\ &= \alpha_u \quad \text{otherwise.} \end{aligned}$$

To illustrate calculation of the initial trial step, we consider a simple example:

$$\text{minimize} \quad x_1 x_2^2 \quad \text{subject to} \quad -x_1^2 - x_2^2 + 2 \ge 0 \;\; \text{and} \;\; -x_1^2 x_2^2 + 10 \ge 0. \tag{3.5}$$

Note that these particular constraints would more naturally be expressed in "less than or equal" form. The second constraint is redundant, but is included to show that the barrier function reflects all constraints at every iteration.

The numerical results in this paper have been rounded to four decimal digits, and were calculated in double precision (around 16 decimal digits) on a Silicon Graphics 4D/240S. The numbers given correspond to the point $x = (-.81, -1.1)^T$, for which the constraint values are $c_1 = .1339$ and $c_2 = 9.206$.

When the barrier parameter $\mu = .01$, the Newton direction is $p = (-.8602, -1.459)^T$. The exact step γ to the nearest constraint and the exact minimizer α^* of the barrier function along p are

$$\gamma = .02857 \qquad \text{and} \qquad \alpha^* = .02601. \tag{3.6}$$

The quantities σ_j defined by (3.2) for the two constraints are $\sigma_1 = .02909$ and $\sigma_2 = 2.428$; the value of $g^T p$ is -3.640. Applying (3.3) and (3.4), we have

$$\bar{\alpha} = \sigma_1 = .02909 \qquad \text{and} \qquad \alpha_b = .02634. \tag{3.7}$$

Since $\alpha_u = 1$ and $\alpha_b < \alpha_u$, the initial step $\alpha^{(0)}$ is taken as α_b. The values of the constraints for the step α_b are $c_1 = .01065$ and $c_2 = 9.101$, and the value of the barrier function is -1.056.

To ensure feasibility, all nonlinear constraints must be evaluated at every trial step $\alpha^{(j)}$. (Strictly speaking, it is not necessary to check feasibility for *linear* constraints for any step less than $\bar{\alpha}$.) If any infeasibilities occur at a trial step, the right endpoint of the interval of uncertainty is updated. Secant estimates of zeros of violated constraints can be calculated from the values of the constraint functions at the infeasible point and the most recent feasible step to its left (which may be $\alpha = 0$). The smallest (positive) estimated step to a constraint zero and the corresponding constraint index then play the roles of $\bar{\alpha}$ and $\bar{\jmath}$ in computing a new estimate of α_b from (3.4).

4. Special Interpolating Functions

The functions to be described in this section are specifically designed for interpolating the logarithmic barrier function. (Similar functions could be developed for other barrier or potential functions.) Minimizers of these special functions need to be calculated by iteration, but the associated computational effort is small. The interpolation procedures can be incorporated into standard line searches based on safeguarded parabolic or cubic interpolation (see, for example, [DS83] or [GM74]) with slight modifications to the structure of a typical iteration; see Section 5 for further comments.

In this discussion, let $\phi(\alpha)$ denote the barrier function $B(x_k + \alpha p_k, \mu)$ treated as a function of the single variable α, where x_k is the current iterate, p_k is the search direction, and μ is the barrier parameter. We assume that p_k is a descent direction with respect to the barrier function, so that $\phi'(0) < 0$. Because $\alpha_k \geq 0$, singularity in the barrier function is relevant only if it occurs at some positive step along p_k.

4.1. Linear plus logarithmic singularity

The first interpolating function has the form

$$\ell(\alpha) = a + b\alpha - \mu \ln(d - \alpha),$$

where d is the positive step to the singularity. Because ℓ is defined by three parameters, its role is analogous to that of a parabolic interpolant for a general smooth function. A parabolic fit is typically used when gradients are relatively expensive compared to function values, so that gradients are computed (or approximated by finite differences) only at x_k, and not at subsequent trial steps during a given line search. (If gradients are available at every trial step, a cubic interpolant would normally be used; the corresponding function with a logarithmic singularity is described in Section 4.2.)

The first and second derivatives of $\ell(\alpha)$ are given by

$$\ell' = b + \frac{\mu}{d - \alpha} \quad \text{and} \quad \ell'' = \frac{\mu}{(d - \alpha)^2}.$$

Because its second derivative is always positive, ℓ is convex. The barrier function is decreasing at $\alpha = 0$, and we impose this same property on ℓ, so that

$$\ell'(0) = b + \frac{\mu}{d} < 0. \tag{4.1}$$

Since $d > 0$ and $\mu > 0$, it follows from (4.1) that $b < 0$. A minimizer of ℓ occurs at α_ℓ satisfying

$$\ell'(\alpha_\ell) = b + \frac{\mu}{d - \alpha_\ell} = 0, \quad \text{so that} \quad \alpha_\ell = d + \frac{\mu}{b} > 0. \tag{4.2}$$

Determination of b and d depends on the data to which $\ell(\alpha)$ is fitted during the line search. Non-acceptance of $\alpha^{(0)}$ means that the first line search iteration is based on fitting $\ell(\alpha)$ to $\phi(0)$, $\phi'(0)$ and $\phi(\alpha^{(0)})$. Thereafter, unless infeasibility is encountered at a step less than $\alpha^{(0)}$, $\ell(\alpha)$ is fitted to the three best values of ϕ.

For simplicity, we assume in this discussion that the trial steps involved in the interpolation are numbered as α_1, α_2, and so on.

4.1.1. Two function values, one first derivative. To match ϕ_1 and ϕ'_1 at α_1 and ϕ_2 at α_2, the parameters a, b and d in $\ell(\alpha)$ must satisfy

$$\phi_1 = a + b\alpha_1 - \mu \ln(d - \alpha_1) \tag{4.3a}$$

$$\phi'_1 = b + \frac{\mu}{d - \alpha_1} \tag{4.3b}$$

$$\phi_2 = a + b\alpha_2 - \mu \ln(d - \alpha_2). \tag{4.3c}$$

There is no loss of generality in assuming that $\alpha_1 = 0$. In fact, it is advantageous numerically during a line search to "shift" the points so that the left endpoint of the interval of uncertainty is locally treated as the origin (see [Br73]). Relations (4.3a) and (4.3b) give

$$a = \phi_1 + \mu \ln d \quad \text{and} \quad b = \phi_1' - \frac{\mu}{d}.$$

Substituting these values into (4.3c), we obtain

$$\phi_2 = \phi_1 + \mu \ln d + \alpha_2 \phi_1' - \frac{\alpha_2 \mu}{d} - \mu \ln(d - \alpha_2).$$

Rearrangement produces a nonlinear equation satisfied by d:

$$\ln\left(\frac{d - \alpha_2}{d}\right) + \frac{\alpha_2}{d} + k_1 = 0, \tag{4.4}$$

where

$$k_1 = \frac{1}{\mu}(\phi_2 - \phi_1 - \alpha_2 \phi_1'). \tag{4.5}$$

Making the change of variable $y = \alpha_2/d$, so that $0 < y < 1$ for the admissible range of d, the equation (4.4) becomes

$$\xi_1(y) = y + \ln(1 - y) + k_1 = 0. \tag{4.6}$$

It is straightforward to see that (i) $\xi_1(0) = k_1$ and $\lim_{y \to 1_-} \xi_1(y) = -\infty$, and (ii) $\xi_1' < 0$ and $\xi_1'' < 0$ for y in $(0, 1)$. The equation $\xi_1(y) = 0$ therefore has a unique zero in $(0, 1)$ if $k_1 > 0$. The condition $k_1 > 0$ holds if ϕ_2 lies above the line through ϕ_1 with slope ϕ_1', i.e., the barrier function is higher at α_2 than its linear interpolant at α_1. If this is not the case, the effect of the singularity is not locally revealed, ℓ is a poor approximation to the barrier function, and safeguarded quadratic interpolation should be used.

Although the equation (4.6) can be treated directly, exponentiation and rearrangement give the even simpler form

$$\psi_1(y) = y + e^{-y} e^{-k_1} - 1 = 0. \tag{4.7}$$

The function ψ_1 is nearly linear, and Newton's method applied to find its zero accordingly converges very rapidly. Because the effects of y and k_1 are separated in (4.6), the expression $y + \ln(1 - y)$ can be tabulated for a small selection of values of y in $(0, 1)$. Once k_1 is calculated, a good starting point for the Newton iteration can be found by table lookup.

Consider the example (3.5), where the two values of α are $\alpha_1 = 0$ and $\alpha_2 = \alpha_b = .02634$. (The choice of α_b is explained in Section 3.) The data to be matched by $\ell(\alpha)$ are

$$\phi_1 = -.9822, \quad \phi_1' = -3.292, \quad \text{and} \quad \phi_2 = -1.056.$$

The value of k_1 defined by (4.5) is 1.312. Starting with $y = 0.9$ (chosen by table lookup), a single Newton iteration produces $y = .8893$, with $\psi_1 = 6.279 \times 10^{-6}$. We obtain $d = .02962$, $b = -3.630$, and (from (4.2)) $\alpha_\ell = .02687$. The value of α_ℓ is an overestimate of the exact minimizer $\alpha^* = .02601$, but is less than $\gamma = .02857$ and hence feasible (see (3.6)). The minimizer of the interpolating quadratic fitted to the same data is the *infeasible* step $\alpha = .08710$, so that $\ell(\alpha)$ gives a much better estimate of α^* for this example.

4.1.2. Three function values.

Given three function values ϕ_i at α_i, $i = 1, 2, 3$, the equations satisfied by a, b and d are

$$\phi_i = \ell(\alpha_i) = a + b\alpha_i - \mu \ln(d - \alpha_i), \quad i = 1, \ldots, 3. \tag{4.8}$$

We assume without loss of generality that $\alpha_1 = 0$ and $\alpha_3 > \alpha_2 > \alpha_1$.

Elimination of a and b from the three equations in (4.8) leads to a nonlinear equation satisfied by d:

$$\frac{1}{\alpha_3}\ln\left(\frac{d-\alpha_3}{d}\right)-\frac{1}{\alpha_2}\ln\left(\frac{d-\alpha_2}{d}\right)+k_2=0, \tag{4.9}$$

where

$$k_2=\frac{1}{\mu}\left(\frac{\phi_3-\phi_1}{\alpha_3}-\frac{\phi_2-\phi_1}{\alpha_2}\right). \tag{4.10}$$

Let $\nu=\alpha_3/\alpha_2$ and $y=\alpha_3/d$, so that $\nu>1$ and $0<y<1$ for all admissible d. Multiplying (4.9) by α_3 and rewriting in terms of y and ν, we obtain

$$\xi_2(y)=\ln(1-y)-\nu\ln\left(1-\frac{y}{\nu}\right)+k_2\alpha_3=0.$$

It is straightforward to show that (i) $\xi_2(0)=k_2\alpha_3$ and $\lim_{y\to 1_-}\xi_2(y)=-\infty$; and (ii) $\xi_2'(y)<0$ and $\xi_2(y)''<0$ for all y in $(0,1)$. These properties imply that ξ_2 has a unique zero in $(0,1)$ if $k_2>0$. The condition $k_2>0$ means that ϕ_3 lies above the line connecting the points (α_1,ϕ_1) and (α_2,ϕ_2), so that the barrier function at α_3 has increased more rapidly than the linear interpolant of ϕ between α_1 and α_2.

The following equivalent and somewhat simpler equation is obtained by rearrangement and exponentiation:

$$\psi_2(y)=1-y-\beta\,(1-\frac{y}{\nu})^\nu=0,$$

where $\beta=e^{-k_2\alpha_3}$. The unique zero of ψ_2 in $(0,1)$ can be determined using Newton's method. It is not convenient to find an initial Newton iterate from a table because of the dependence on both β and ν. However, the Newton iterations for ψ_2 can safely be initiated with the starting value $y=1-\beta$. The parameters in $\ell(\alpha)$ can then be computed from the relations

$$d=\alpha_3/y,\quad a=\phi_1+\mu\ln d,\quad \text{and}\quad b=\frac{1}{\alpha_2}\left(\phi_2+\mu\ln(d-\alpha_2)-a\right),$$

and the minimizer α_ℓ is given by (4.2).

Consider the example (3.5), with the three values of α taken as $\alpha_1=0$, $\alpha_3=\alpha_b$ (see (3.7)), and $\alpha_2=\frac{1}{2}\alpha_3$, for which $\phi_2=-1.025$. The value of k_2 defined by (4.10) is 43.80, and $\beta=.3154$. Starting with $y=.6846$, two iterations of Newton's method applied to $\psi_2(y)$ give $y=.9056$ and $\psi_2=-2.435\times10^{-6}$. The parameters of $\ell(\alpha)$ are

$$d=.02909,\quad a=-1.018,\quad \text{and}\quad b=-3.233.$$

Applying relation (4.2), we obtain $\alpha_\ell=.02638$, which is a (feasible) overestimate of the exact value $\alpha^*=.02601$ (see (3.6)). The minimizer of the quadratic function fitted to the same three points is the infeasible step .05519.

4.2. Quadratic plus logarithmic singularity

The second interpolating function has the form

$$q(\alpha)=a+b\alpha+c\alpha^2-\mu\ln(d-\alpha).$$

Since $q(\alpha)$ contains four parameters, it is the analogue of a cubic polynomial in standard line searches. In general, the change from a parabolic to cubic interpolant adds complications because of the need to choose the correct stationary point of the cubic (see, for example, [GM74]). A similar phenomenon occurs when moving from $\ell(\alpha)$ to the more complex function $q(\alpha)$.

4.2.1. Finding the correct root. Differentiating q, we obtain

$$q' = b + 2c\alpha + \frac{\mu}{d-\alpha}, \quad q'' = 2c + \frac{\mu}{(d-\alpha)^2}, \quad q''' = \frac{2\mu}{(d-\alpha)^3}.$$

The barrier function is decreasing at $\alpha = 0$, and we consequently require that $q'(0) < 0$. Hence $b + \mu/d < 0$, which means that $bd + \mu < 0$ and $b < 0$. Since q is unbounded above as $\alpha \to d$, a minimizer of q must exist in the interval $(0, d)$.

A stationary point α_q of q occurs when $q' = 0$, which means that α_q satisfies the quadratic equation

$$2c\alpha^2 + (b - 2cd)\alpha - \mu - bd = 0. \tag{4.11}$$

If $c = 0$, so that there is no quadratic term, $q(\alpha)$ reduces to the linear-log function $\ell(\alpha)$ (Section 4.1).

If $c \neq 0$, there are two roots r_1 and r_2 whose product satisfies

$$r_1 r_2 = \frac{-(bd+\mu)}{2c}. \tag{4.12}$$

The values of r_1 and r_2 are given by the usual quadratic formula, expressed in two forms:

$$\begin{aligned} r_1, r_2 &= \frac{2cd - b \pm \sqrt{(b-2cd)^2 + 8c(\mu + bd)}}{4c} && (4.13\text{a}) \\ &= \frac{2cd - b \pm \sqrt{(b+2cd)^2 + 8\mu c}}{4c}. && (4.13\text{b}) \end{aligned}$$

The discriminant in (4.13b) is positive, so that both roots are real. We know already that at least one root is positive, and now need to determine which root lies in the interval $(0, d)$.

If $c < 0$, relation (4.12) shows that the roots are of opposite sign (since $bd + \mu < 0$). Further, we know that $2cd - b > 0$ and $8c(\mu + bd) > 0$. Because there is a division by c, relation (4.13a) reveals that the root corresponding to the positive square root is negative and therefore unacceptable as α_q.

If $c > 0$, we know from (4.12) that both roots are positive. Further, $2cd - b = 2cd + |b|$, and

$$+\sqrt{(b+2cd)^2 + 8\mu c} \; > \; \max\,(2cd - |b|, |b| - 2cd). \tag{4.14}$$

If $2cd > |b|$, the combination of (4.13b) and (4.14) shows that choosing the positive square root produces a root that exceeds d and thus cannot be α_q. Similarly, if $|b| > 2cd$, it follows that $|b|/c > 2d$; manipulation of (4.13b) again shows that the root associated with the positive square root exceeds d.

It follows that, when $c \neq 0$, the desired root α_q of (4.11) corresponds to the negative square root in (4.13).

4.2.2. Two function values, two first derivatives. Here, the four pieces of data that determine $q(\alpha)$ are the values of ϕ and ϕ' at two points. We assume that $\alpha_2 > \alpha_1$ and that $\alpha_1 = 0$. The four equations defining the parameters of q are

$$\begin{aligned} \phi_1 &= a - \mu \ln d && (4.15\text{a}) \\ \phi_1' &= b + \frac{\mu}{d} && (4.15\text{b}) \\ \phi_2 &= a + b\alpha_2 + c\alpha_2^2 - \mu \ln(d - \alpha_2) && (4.15\text{c}) \\ \phi_2' &= b + 2c\alpha_2 + \frac{\mu}{d - \alpha_2}. && (4.15\text{d}) \end{aligned}$$

The first two equations provide expressions for a and b in terms of d:

$$a = \phi_1 + \mu \ln d \qquad \text{and} \qquad b = \phi_1' - \frac{\mu}{d}. \tag{4.16}$$

Substitution for b in (4.15d) gives an expression for c in terms of d:

$$c = \frac{1}{2\alpha_2}\left(\phi_2' - \phi_1' + \frac{\mu}{d} - \frac{\mu}{d-\alpha_2}\right). \tag{4.17}$$

We then substitute the formulas for a, b and c into (4.15c), re-arrange and obtain the nonlinear equation

$$\ln\left(\frac{d-\alpha_2}{d}\right) + \frac{\alpha_2}{2}\left(\frac{1}{d} + \frac{1}{d-\alpha_2}\right) - k_3 = 0, \tag{4.18}$$

where

$$k_3 = \frac{1}{\mu}\left(\frac{\alpha_2}{2}(\phi_1' + \phi_2') - (\phi_2 - \phi_1)\right). \tag{4.19}$$

Introducing the variable $z = 1 - \alpha_2/d$, so that $0 < z < 1$, the equation (4.18) becomes

$$\psi_3(z) = \ln z + \frac{1}{2}\left(\frac{1}{z} - z\right) - k_3 = 0. \tag{4.20}$$

Care must be taken in solving for a zero of this equation because of unboundedness in the logarithm and reciprocal functions for z near zero. It is straightforward to show that (i) $\lim_{z\to 0_+}\psi_3(z) = +\infty$ and $\psi_3(1) = -k_3$, and (ii) $\psi_3'(z) < 0$ and $\psi_3''(z) > 0$ for z in $(0,1)$. We conclude that $\psi_3(z)$ has a unique zero in $(0,1)$ if $k_3 > 0$. The condition $k_3 > 0$ means that ϕ is locally rising more rapidly than a quadratic, in the sense that the average of ϕ_1' and ϕ_2' exceeds the slope of the straight line joining ϕ_1 and ϕ_2.

Newton's method is guaranteed to converge monotonically if the initial iterate is chosen so that $\psi_3 > 0$. Except when k_3 is extremely large, this can easily be assured by choosing the starting point from a suitable table. If k_3 is large, z must be small; hence we can simply set $z = \epsilon$ for some suitably small ϵ whenever k_3 is greater than a specified value.

Given z, the values of d, c and b and may be obtained from the relations $d = \alpha_2/(1-z)$, (4.16) and (4.17), and then used in a numerically satisfactory form to compute the root α_q.

Consider the example (3.5), with $q(\alpha)$ fitted to

$$\begin{aligned} \alpha_1 &= 0, & \phi_1 &= -.9822, & \phi_1' &= -3.292, \\ \alpha_2 &= .02634, & \phi_2 &= -1.056, & \phi_2' &= .5864. \end{aligned}$$

The value of k_3 defined by (4.19) is 3.798. Starting with $z = .07$ (chosen by table lookup), two iterations of Newton's method applied to ψ_3 produce $z = .07832$, with $\psi_3 = 1.486 \times 10^{-7}$. We then obtain $d = .02858$, $c = -4.518$, $b = -3.642$, and $\alpha_q = .02600$.

The cubic fitted to the same data has a minimizer at $\alpha = .02507$. The minimizers of $q(\alpha)$ and the cubic are rather similar for this example, in part because the positive value of ϕ_2' forces the minimizers of both $q(\alpha)$ and the cubic to lie to the left of α_2.

In contrast, consider an instance in which $\phi_2' < 0$, so that the minimizer is not explicitly bracketed. Using the same example, we now choose $\alpha_2 = \frac{1}{2}\alpha_b$, for which $\phi_2 = -1.025$ and $\phi_2' = -3.112$. The minimizer of $q(\alpha)$ is now $\alpha_q = .02604$ (still a reasonably good estimate of α^*), whereas the minimizer of the interpolating cubic is the infeasible step .05615.

4.2.3. Two function values, one first derivative, one second derivative. The approximating function $q(\alpha)$ is also appropriate when second derivatives of the problem functions are available. In particular, when Newton's method is used to minimize the barrier function, the second derivative of ϕ at $\alpha = 0$ is known. The function $q(\alpha)$ can then be chosen to match the data ϕ_1, ϕ_1', ϕ_1'' and ϕ_2.

The defining equations are

$$\begin{aligned} \phi_1 &= a - \mu \ln d, \quad \phi_1' = b + \frac{\mu}{d}, \quad \phi_1'' = 2c + \frac{\mu}{d^2}, \\ \phi_2 &= a + b\alpha_2 + c\alpha_2^2 - \mu\ln(d-\alpha_2). \end{aligned} \tag{4.21}$$

After the first three equations are arranged to express a, b and c in terms of d, the fourth equation yields

$$\ln(d-\alpha_2) - \ln d + \frac{\alpha_2}{d} + \frac{\alpha_2^2}{2d^2} = \frac{1}{\mu}(\phi_1 + \alpha_2\phi_1' + \tfrac{1}{2}\alpha_2^2\phi_1'' - \phi_2). \tag{4.22}$$

If we define $z = 1 - \alpha_2/d$, so that $0 < z < 1$, the relation (4.22) becomes

$$\psi_4(z) = \ln z + 1 - z + \tfrac{1}{2}(1-z)^2 + k_4 = 0, \tag{4.23}$$

where k_4 is given by

$$k_4 = \frac{1}{\mu}(\phi_2 - \phi_1 - \alpha_2\phi_1' - \tfrac{1}{2}\alpha_2^2\phi_1''). \tag{4.24}$$

Analysis of ψ_4 shows that (i) $\lim_{z\to 0_+}\psi_4(z) = -\infty$ and $\psi_4(1) = k_4$, and (ii) $\psi_4'(z) > 0$ and $\psi_4''(z) < 0$ for all z in $(0,1)$. It follows that ψ_4 has a unique zero in $(0,1)$ if $k_4 > 0$. Positivity of k_4 means that the value of ϕ at α_2 exceeds that of the quadratic fitted to ϕ_1, ϕ_1' and ϕ_1'' at α_1.

Newton's method applied to ψ_4 will converge monotonically if the starting point is such that $\psi_4(z) < 0$. As in Section 4.2.2, this property can be assured by appropriate selection of a starting point based on the value of k_4 and a small table. To avoid difficulties when z is very close to zero, a suitable cutoff value of k_4 can be chosen. For example, if $k_4 = 10.01$, the solution of $\psi_4(z) = 0$ is $z = 10^{-5}$; hence z can simply be taken as 10^{-5} if $k_4 \geq 10.01$.

Once z has been computed, d is given by $\alpha_2/(1-z)$, and b and c are easily obtained from (4.21). The value of α_q is then calculated as described in Section 4.2.1.

Once again, consider example (3.5). Since p is the Newton direction, the second derivative of ϕ at $\alpha = 0$ is simply $\phi''(0) = -\phi'(0)$. The values to which $q(\alpha)$ is fitted are

$$\begin{aligned} &\alpha_1 = 0, &&\phi_1 = -.9822, &&\phi_1' = -3.292, &&\phi_1'' = 3.292,\\ &\alpha_2 = .02634, &&\phi_2 = -1.056. \end{aligned}$$

The value of k_4 defined by (4.24) is 1.197. Starting with $z = .07$ (chosen by table lookup), two iterations of Newton's method applied to ψ_4 produce $z = .07861$, with $\psi_4 = -2.552 \times 10^{-5}$. We then have $d = .02859$, $c = -4.470$, $b = -3.642$, and $\alpha_q = .02601$.

The cubic fitted to the same data has a minimizer at the infeasible step $\alpha = .04011$. In contrast to the example in Section 4.2.2, the cubic here does not reflect the (positive) derivative of ϕ at α_2.

5. Implementation

The procedures of Sections 3 and 4 are designed to be incorporated within a standard safeguarded polynomial line search. The special functions of Section 4 are fitted iteratively, with appropriate adjustment of the interval of uncertainty, exactly as in a general-purpose line search, until a sufficient decrease condition is satisfied (see Section 2.2). For very small values of the barrier parameter, care must be taken when the estimated difference between the singularity and the minimizer is less than the smallest spacing permitted between trial steps. (See [Br73] for a discussion of the latter value.)

When minimizing barrier functions, extra housekeeping, such as testing for feasibility, must be carried out whatever the line search procedure. Fitting the specialized functions requires more work than polynomial interpolation, since each Newton step for estimating the location of the singularity requires calculation of a logarithm or an exponential. Typically, however, the amount of extra work is small; a good starting guess is easily obtained, and one or two Newton iterations usually converge to sufficient accuracy.

As we have stressed several times, for nonlinear problems the characteristic singularity of the barrier function does not necessarily affect every line search. For each trial step, a decision must accordingly be made as to whether to compute the next trial point using the special functions of Section 4 or standard polynomials. The underlying guideline is to use the special functions if $\alpha^{(0)}$ is taken as α_b (see Section 3), or if a constraint violation occurs at any trial step. Furthermore, each special function is meaningful only when a certain constant depending on the given data (k_1–k_4; see Sections 4.1 and 4.2) is positive. If $k_i \leq 0$, a standard polynomial interpolant should be constructed rather than a special function.

When a highly inexact line search is performed, the initial step $\alpha^{(0)}$ itself, obtained as described in Section 3, often yields a sufficient decrease—particularly near x^* (when $\alpha^{(0)} = \alpha_u$), or when α_b is small. In such circumstances, a complete line search involves only a single trial point. Far from x^*, however, or when an accurate line search is desirable (for example, in methods that closely follow the trajectory of the barrier function), the techniques described here should help to make the line search efficient.

In our own experiments (to be described in a later report) and in those reported by [NS91], a specialized line search significantly reduces the average number of evaluations of the problem functions in a typical barrier function minimization. Such results support the usefulness of the suggested techniques.

References

[Ans90] K. M. Anstreicher, On long step path following and SUMT for linear and quadratic programming, Report, George Washington University, Washington, DC, 1990.

[ADRT90] K. M. Anstreicher, D. den Hertog, C. Roos and T. Terlaky, A long step barrier method for convex quadratic programming, Report 90-53, Delft University of Technology, Delft, The Netherlands, July 1990.

[Br73] R. P. Brent, *Algorithms for Minimization without Derivatives*, Prentice-Hall, Englewood Cliffs, New Jersey, 1973.

[BN89] R. H. Byrd and J. Nocedal, A tool for the analysis of quasi-Newton methods with application to unconstrained optimization, *SIAM Journal on Numerical Analysis* 26, 727–739, 1989.

[DRT90] D. den Hertog, C. Roos and T. Terlaky, On the classical logarithmic barrier function method for a class of smooth convex programming problems, Report 90-28, Delft University of Technology, Delft, The Netherlands, April 1990.

[DM77] J. E. Dennis, Jr., and J. J. Moré, Quasi-Newton methods, motivation and theory, *SIAM Review* 19, 46–89, 1977.

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

[DS89] J. E. Dennis, Jr. and R. B. Schnabel, A view of unconstrained optimization, in *Optimization* (G. L. Nemhauser, A. H. G. Rinnooy Kan and M. J. Todd, eds.), North-Holland, New York, 1–72, 1989.

[Fia79] A. V. Fiacco, Barrier methods for nonlinear programming, in *Operations Research Support Methodology* (A. G. Holzman, ed.), Marcel Dekker, New York, 1979.

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

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

[FlMc69] R. Fletcher and A. P. McCann, Acceleration techniques for nonlinear programming, in *Optimization* (R. Fletcher, ed.), Academic Press, New York, 203–213, 1969.

[GM74] P. E. Gill and W. Murray, Safeguarded steplength algorithms for optimization using descent methods, Report NAC 37, National Physical Laboratory, Teddington, England, 1974.

[GMW81] P. E. Gill, W. Murray and M. H. Wright, *Practical Optimization*, Academic Press, London, 1981.

[GMSW89] P. E. Gill, W. Murray, M. A. Saunders and M. H. Wright, Constrained nonlinear programming, in *Optimization* (G. L. Nemhauser, A. H. G. Rinnooy Kan and M. J. Todd, eds.), North-Holland, New York, 171–210, 1989.

[GT89] D. Goldfarb and M. J. Todd, Recent developments in linear programming, in *Optimization* (G. L. Nemhauser, A. H. G. Rinnooy Kan and M. J. Todd, eds.), North-Holland, New York, 73–170, 1989.

[GL91] D. Goldfarb and S. Liu, An $O(n^3 L)$ primal interior point method for convex quadratic programming, *Mathematical Programming* 49, 325–340, 1991.

[Gonz92] C. C. Gonzaga, Path-following methods for linear programming, *SIAM Review* 34, to appear, 1992.

[Jar90] F. Jarre, Interior-point methods for convex programming, Report SOL 90-16, Department of Operations Research, Stanford University, Stanford, California, 1990.

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

[Kar84] N. K. Karmarkar, A new polynomial-time algorithm for linear programming, *Combinatorica* 4, 373–395, 1984.

[LFR73] L. S. Lasdon, R. L. Fox and M. W. Ratner, An efficient one-dimensional search procedure for barrier functions, *Mathematical Programming* 4, 279–295, 1973.

[MS90] S. Mehrotra and J. Sun, An interior point algorithm for solving smooth convex programs based on Newton's method, in *Mathematical Developments Arising from Linear Programming* (J. C. Lagarias and M. J. Todd, eds.), Contemporary Mathematics, Volume 114, 265–294, American Mathematical Society, Providence, Rhode Island, 1990.

[MW76] W. Murray and M. H. Wright, Efficient linear search algorithms for the logarithmic barrier function, Report SOL 76–18, Department of Operations Research, Stanford University, Stanford, California, 1976.

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

[NN89] J. E. Nesterov and A. S. Nemirovsky, Self-concordant functions and polynomial-time methods in convex programming, Central Economic and Mathematical Institute, USSR Academy of Science, Moscow, USSR, 1989.

[OR70] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[Ring91] U. T. Ringertz, Optimal design of nonlinear shell structures, Report FFA TN 91-18, Aeronautical Research Institute of Sweden, 1991.

[Wr76] M. H. Wright, Numerical methods for nonlinearly constrained optimization, Ph. D. thesis, Computer Science Department, Stanford University, Stanford, California. 1976.

[Wr92] M. H. Wright, Interior methods for constrained optimization, *Acta Numerica* 1, 1992, to appear.