Interior Methods for Constrained Optimization

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

Abstract

Interior methods for optimization were widely used in the 1960s, primarily in the form of barrier methods. However, they were not seriously applied to linear programming because of the dominance of the simplex method. Barrier methods fell from favour during the 1970s for a variety of reasons, including their apparent inefficiency compared with the best available alternatives. In 1984, Karmarkar's announcement of a fast polynomial-time interior method for linear programming caused tremendous excitement in the field of optimization. A formal connection can be shown between his method and classical barrier methods, which have consequently undergone a renaissance in interest and popularity. Most papers published since 1984 have concentrated on issues of computational complexity in interior methods for linear programming. During the same period, implementations of interior methods have displayed great efficiency in solving many large linear programs of ever-increasing size. Interior methods have also been applied with notable success to nonlinear and combinatorial problems. This paper presents a self-contained survey of major themes in both classical material and recent developments related to the theory and practice of interior methods.

1. Introduction to interior methods

1.1. The way we were

Before 1984, the question 'How should I solve a linear program?' would have been answered almost without exception by 'Use the simplex method'. In fact, it would have been extremely difficult to find serious discussion of any method for linear programming (LP) other than the famous simplex method developed by George B. Dantzig in 1947.

As most readers already know, the simplex method is an iterative procedure derived from a fundamental property of essentially all linear programs: an optimal solution lies at a vertex of the feasible region. Beginning with a vertex, the simplex method moves between adjacent vertices, decreasing the objective as it goes, until an optimal vertex is found.

Although nonsimplex strategies for LP were suggested and tried from time to time, such techniques had never approached the simplex method in overall speed and reliability. Hence the simplex method retained unquestioned pre-eminence as the linear programming method of choice for nearly 40 years. (We describe later the persistent unhappiness with the simplex method on grounds of its theoretical complexity.)

Such an exclusive focus on the simplex method had several effects on the field of optimization. Largely for historical reasons, the simplex method is surrounded by a bevy of highly specialized terminology ('basic feasible solution') and pedagogical constructs (the tableau) with little apparent connection to other continuous optimization problems. Many researchers and practitioners consequently viewed linear programming as philosophically distinct from nonlinear programming. This conceptual gap reinforced a tendency to develop 'new' linear programming methods only as variations on the simplex method.

In marked contrast, the field of *nonlinear* optimization was characterized not only by the constant development of new methods with differing flavours, but also by a shift over time in the preferred solution techniques. Since the late 1970s, for example, nonlinearly constrained optimization problems have been solved with sequential quadratic programming (SQP) methods, which involve a sequence of constrained subproblems based on the Lagrangian function. In the 1960s, however, constrained problems were most often converted to *unconstrained* subproblems. Penalty and barrier methods were especially popular, both motivated by minimizing a composite function that reflects the original objective function as well as the influence of the constraints. Classical barrier methods, intended for inequality constraints, include a composite function containing an impassable positive singularity ('barrier') at the boundary of the feasible region, and thereby maintain strict feasibility while approaching the solution.

Although barrier methods were widely used and thoroughly analysed during the 1960s (see Section 3 for details and references), they nonetheless suffered a severe decline in popularity in the 1970s for various reasons, including inherent ill-conditioning as well as perceived inefficiency compared to alternative strategies. By the late 1970s, barrier methods were considered for the most part an interesting but passé solution technique.

As we shall see, the situation today (1991) in both linear and nonlinear programming has altered dramatically since 1984, primarily as a result of dissatisfaction with the theoretical computational complexity of the simplex method.

1.2. Concerns about the simplex method

On 'real-world' problems, the simplex method is invariably extremely efficient, and consistently requires a number of iterations that is a small multiple (2–3) of the problem dimension. Since the number of vertices associated with any LP is finite, the simplex method is also guaranteed under quite mild conditions to converge to the optimal solution. The number of vertices, however, can be exponentially large. The well known 'twisted cube' example of Klee and Minty (1972) is a linear program with n variables and $2n$ inequality constraints for which the simplex method with the standard pivot-selection rule visits each of the 2^n vertices. The *worst-case* complexity of the simplex method (the number of arithmetic operations required to solve a general LP) is consequently *exponential* in the problem dimension. The gigantic gap between the observed and worst-case performance of the simplex method is still puzzling; the issue of whether an (undiscovered) simplex pivot rule could improve its complexity is also unresolved.

As the formal study of computational complexity increased in importance during the 1960s and 1970s, it became a strongly held article of faith among computer scientists that a 'fast' algorithm must be *polynomial-time*, meaning that the number of operations required to solve the problem should be bounded above by a polynomial in the problem size. The simplex method clearly does not satisfy this property. Although practitioners routinely and happily solved large linear programs with the simplex method, the existence of a provably polynomial algorithm remained a major open question.

In 1979, to the accompaniment of worldwide publicity, Leonid Khachian published the first polynomial algorithm for LP. The *ellipsoid method* of Khachian is based on earlier techniques for nonlinear programming developed by other mathematicians, notably Shor, Yudin and Nemirovsky. An interesting feature of Khachian's approach is that it does not rely on combinatorial features of the LP problem. Rather, it constructs a sequence of ellipsoids such that each successive ellipsoid both encloses the optimal solution and undergoes a strict reduction in volume. The ellipsoid method generates improving iterates in the sense that the region of uncertainty surrounding the solution is monotonically 'squeezed'. (Simplex iterates are also improving in the sense that the objective value is decreasing, but they provide no information about the closeness of the current iterate to the solution.)

The crucial elements in polynomiality of the ellipsoid method are what might be termed outer and inner bounds for the solution. The outer bound guarantees an initial enclosing ellipsoid, and the inner bound specifies the size of the final ellipsoid needed to ensure sufficient closeness to the exact solution. Similar features also figure prominently in the complexity analysis of interior methods, and are discussed in Section 6.

Despite its polynomial complexity, the ellipsoid method's performance was extremely disappointing. In practice, the number of iterations tended to be almost as large as the worst-case upper bound, which, although polynomial, is *very* large. The simplex method accordingly retained its position as the clear winner in any comparison of actual solution times. Creation of the ellipsoid method led to an unexpected anomaly in which an algorithm with the desirable theoretical property of polynomiality compared unfavourably in speed to an algorithm with worst-case exponential complexity. The quest therefore continued for an LP algorithm that was not only polynomial, but also efficient in practice.

This search ended in 1984, when Narendra Karmarkar presented a novel interior method of polynomial complexity for which he reported solution times 50 times faster than the simplex method. Once again, international coverage in the popular press surrounded the event, which has had remarkable and lasting scientific consequences.

Karmarkar's announcement led to an explosion of interest among researchers and practitioners, with substantial progress in several directions. Interior methods are indeed 'fast'; extensive numerical trials have shown conclusively that a variety of interior methods can solve many very large linear programs substantially faster than the simplex method. After a formal relationship was shown between Karmarkar's method and classical barrier methods (Gill *et al.*, 1986), much research has concentrated on the common theoretical foundations of linear and nonlinear programming.

Unlike the simplex method, interior techniques can obviously be applied to nonlinear optimization problems. (In fact, they were devised more than 30 years ago for this purpose!) Interior methods have already been developed for quadratic and nonlinear programming, and extensions of the interior approach to difficult combinatorial problems have also been proposed; see Karmarkar (1990).

A fundamental theme permeating the motivation for interior methods is the creation of continuously parametrized families of approximate solutions that asymptotically converge to the exact solution. As the parameter approaches its limit, the paths to the solution trace smooth trajectories whose geometric properties can be analysed. Each iteration of a 'path-following' method constructs a step intended to follow one of these trajectories, moving both 'toward' and 'along' the path. In the first heyday of barrier methods, these ideas led to great interest in extrapolation. Today, they are being generalized and extended to new problem areas; for a discussion of such ideas in linear programming, see Megiddo (1987), Bayer and Lagarias (1989, 1991), and Karmarkar (1990). The field of interior methods seems to offer the continuing promise of original theory and efficient methods.

1.3. Overview

This article covers only a small part of the large and rapidly expanding number of topics related to interior methods. Although the term 'interior methods' is not precisely defined, several themes perceived as disparate before 1984 can now be placed in a unified framework. For reasons of space, we motivate interior methods only through a 'classical' barrier function. Karmarkar's original 1984 algorithm was based on nonlinear projection, a perspective that provides interesting geometric insights. See Gonzaga (1992), Nesterov and Nemirovsky (1989), and Powell (1990) for further interpretations.

Work in interior methods today is a melange of rediscovered as well as new methods, complexity analysis, and sparse linear algebra. The approach taken in this article is to present some initial background on optimization (Section 2), followed by a detailed treatment of the theory of classical barrier methods (Section 3). After reviewing Newton's method (Section 4), we turn in Section 5 to the special case of linear programming, and describe the structure of several interior methods. A particular interior LP method and its complexity analysis are given in detail (Section 6) to give the flavour of such proofs. The practical success of interior methods is dependent on efficient linear algebra; the relevant techniques for linear and nonlinear problems are described in Section 7. Finally, we close by mentioning selected directions for future research.

2. Background in optimization

2.1. Definitions and notation

Optimization problems, broadly speaking, involve finding the 'best' value of some function. A continuous optimization problem has three ingredients: a set of variables, usually denoted by the real n-vector x; an *objective* function $f(x)$ to be optimized (minimized or maximized); and *constraints* (equality and/or inequality) that restrict acceptable values of the variables.

Except for the linear programming case, our main interest is in inequality constraints. We consider a generic optimization problem of the form

$$\underset{x\in\mathcal{R}^n}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c_i(x)\geq 0, \quad i=1,\ldots,m. \tag{2.1}$$

It is assumed throughout that all functions of interest are smooth. This assumption is stronger than necessary, and is imposed mainly to simplify the discussion.

The set of points satisfying the constraints of (2.1) is denoted by

$$\mathcal{F}=\{x \mid c_i(x)\geq 0,\ i=1,\ldots,m\}, \tag{2.2}$$

and is called the *feasible region*. If x is in $\mathcal{F}$, x is said to be *feasible*.

A *linear programming* problem is an optimization problem in which the objective function and all the constraint functions, both equalities and inequalities, are linear. An optimization problem is called a *nonlinear program* if the objective or any constraint function is nonlinear. A *quadratic program* has a quadratic objective and linear constraints.

Several definitions involving sets will be important in our discussion. All sets are in $\mathcal{R}^n$ unless stated otherwise.

Definition 1 *(Interior of a set.) Given a set S, a point x is an* interior point *of S if $x\in S$ and there exists a neighbourhood of x that is entirely contained in S. The* interior *of S, denoted by* $\text{int}(S)$*, is the collection of all interior points of S.*

Definition 2 *(Boundary of a set.) Given a set S, a point x is a* boundary point *of S if every neighbourhood of x contains at least one point in S and at least one point not in S. The* boundary *of S is the collection of all boundary points of S.*

It is straightforward to show that a closed set contains all its boundary points.

For the feasible region $\mathcal{F}$ (2.2) associated with our generic optimization problem, the subset of points in $\mathcal{F}$ for which all the constraint functions are *strictly positive* is denoted by $\text{strict}(\mathcal{F})$ and defined as

$$\text{strict}(\mathcal{F})=\{x \mid c_i(x)>0, \quad i=1,\ldots,m\}. \tag{2.3}$$

A point x in $\text{strict}(\mathcal{F})$ is said to be *strictly feasible*.

Although the sets $\text{strict}(\mathcal{F})$ and $\text{int}(\mathcal{F})$ are identical in many instances, they can be different. For example, consider the single constraint $x_1^2+x_2^2\geq 0$ in $\mathcal{R}^2$. The corresponding feasible region $\mathcal{F}$ includes all of $\mathcal{R}^2$; consequently, every point in $\mathcal{R}^2$ is an interior point, and $\mathrm{int}(\mathcal{F}) = \mathcal{R}^2$. In contrast, the set $\mathrm{strict}(\mathcal{F})$ includes all points in $\mathcal{R}^2$ *except* the origin.

The idea of a level set will be used in several proofs.

Definition 3 *(Level set.) For x in a set S, the* level set *of the function $f(x)$ corresponding to the constant τ is the set of points in S for which the value of f is less than or equal to τ:*

$$\{x \in S \mid f(x) \le \tau\}.$$

For reference, we state formal definitions of local, global and isolated minimizers for the generic problem (2.1). The definitions given here, taken from Fiacco and McCormick (1968), are tailored to our treatment of interior methods, and are slightly different from those in standard textbooks. They can be specialized in an obvious way to include additional restrictions on x.

Definition 4 *(Local constrained minimizer.) The point x^* is a* local (constrained) minimizer *of problem (2.1) if there exists a compact set S such that*

$$x^* \in \mathrm{int}(S) \cap \mathcal{F} \quad \textit{and} \quad f(x^*) = \min\{f(x) \mid x \in S \cap \mathcal{F}\}.$$

Definition 5 *(Global constrained minimizer.) The point x^* is a* global (constrained) minimizer *of problem (2.1) if*

$$x^* \in \mathcal{F} \quad \textit{and} \quad f(x^*) = \min\{f(x) \mid x \in \mathcal{F}\}.$$

Definition 6 *(Isolated constrained minimizer.) A constrained minimizer x^* is* isolated *if there is a neighbourhood of x^* in which x^* is the only constrained minimizer.*

For the nonlinear function $f(x)$, the n-vector $g(x)$ denotes the gradient (vector of first partial derivatives) of f, and the $n \times n$ symmetric matrix $H(x)$ denotes the Hessian (matrix of second partial derivatives) of f. Given a nonlinear constraint function $c_i(x)$, its gradient will be denoted by $a_i(x)$, and its Hessian by $H_i(x)$. For an m-vector $c(x)$ of constraint functions, the $m \times n$ Jacobian matrix of c is denoted by $A(x)$, whose ith row (the transposed gradient of c_i) is $a_i(x)^T$.

2.2. Optimality conditions

We now state optimality conditions for three varieties of nonlinear optimization problems, without any explanation of the origin of these conditions. (Optimality conditions for linear programming are given in Section 5.) Detailed derivations of optimality conditions are given in, for example, Avriel (1976); Fiacco and McCormick (1968); Fletcher (1987); Gill *et al.* (1981); and Luenberger (1984). Optimality conditions are extremely important because they not only allow us to recognize that a solution has been found, but also suggest algorithms for finding a solution.

Unconstrained optimization. The definition of a local unconstrained minimizer will be important in our discussion of barrier functions, where 'unconstrained' implies that no constraints are *locally* relevant.

Definition 7 *(Local unconstrained minimizer.) The point x^* is a local unconstrained minimizer of $f(x)$ if there exists a compact set S such that*

$$x^* \in \mathrm{int}(S) \quad \textit{and} \quad f(x^*) = \{\min f(x) \mid x \in S\}.$$

The following conditions are well known to be necessary for x^* to be an unconstrained minimizer of $f(x)$:

$$g(x^*) = 0 \quad \text{and} \quad H(x^*) \geq 0, \tag{2.4}$$

where the notation '$M \geq 0$' means that the matrix M is positive semi-definite. (Similarly, '$M > 0$' means that M is positive definite.)

Sufficient conditions for x^* to be an isolated unconstrained minimizer of $f(x)$ are

$$g(x^*) = 0 \quad \text{and} \quad H(x^*) > 0. \tag{2.5}$$

The 'order' of an optimality condition refers to the highest order of the derivatives that it contains. For example, the requirement that $g(x^*) = 0$ is a first-order optimality condition.

Linear equality constraints. Consider the problem of minimizing $f(x)$ subject to linear equality constraints:

$$\underset{x \in \mathcal{R}^n}{\text{minimize}} \;\; f(x) \quad \text{subject to} \quad Ax = b, \tag{2.6}$$

where A is a constant $m \times n$ matrix. (Note that A is the Jacobian of the linear constraints $Ax - b = 0$.) Let N denote any matrix whose columns form a basis for the null space of A, i.e. for the subspace of vectors p such that $Ap = 0$. Although the null space itself is unique, in general there are an infinite number of associated bases.

The following conditions are necessary for the point x^* to be a local solution of (2.6):

$$\begin{aligned} Ax^* &= b; && \text{(2.7a)} \\ g(x^*) &= A^T\lambda^* \quad \text{for some } \lambda^*; && \text{(2.7b)} \\ N^T H(x^*) N &\geq 0. && \text{(2.7c)} \end{aligned}$$

Sufficient conditions for x^* to be an isolated solution of (2.6) are that (2.7a–b) hold and that $N^T H(x^*) N$ is positive definite.

The first-order condition $g(x^*) = A^T\lambda^*$ of (2.7b) means that the gradient of f at an optimal point can be expressed as a linear combination of the columns of A^T, and hence lies in the range space of A^T. The *Lagrange multiplier* λ^* represents the set of coefficients in this linear combination, and is unique if A has full row rank.

The *Lagrangian function* for problem (2.6) is

$$L(x, \lambda) = f(x) - \lambda^T (Ax - b), \tag{2.8}$$

where λ is an m-vector. Condition (2.7b) can be interpreted as a statement that the gradient of the Lagrangian function with respect to x vanishes when $\lambda = \lambda^*$.

The relation $g(x^*) = A^T\lambda^*$ is also equivalent to the condition $N^T g(x^*) = 0$, namely, the projection of $g(x^*)$ into the null space of A vanishes. (The vector $N^T g(x)$ is called the *reduced gradient* of f at x.) Condition (2.7b) is therefore analogous to the requirement in the unconstrained case that the gradient itself must be zero.

The matrix $N^T H(x^*) N$ appearing in the second-order optimality condition (2.7c) is the Hessian of f projected into the null space of A, and is called the *reduced Hessian* of f. For linear equality constraints, the reduced Hessian plays the same role in optimality conditions as the full Hessian in the unconstrained case.

The feasibility and first-order optimality conditions (2.7a–b) satisfied by x^* and λ^* can conveniently be summarized as a system of $(n + m)$ nonlinear equations in the variables (x, λ):

$$\Phi(x, \lambda) = \begin{pmatrix} g(x) - A^T\lambda \\ Ax - b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.9}$$

These equations state that the gradient of the Lagrangian function (2.8) and the constraint vector $Ax - b$ should both be zero.

Nonlinear inequality constraints. The final problem category to be discussed is the generic problem with nonlinear inequality constraints:

$$\underset{x \in \mathcal{R}^n}{\text{minimize}} \;\; f(x) \quad \text{subject to} \quad c(x) \geq 0, \tag{2.10}$$

where $c(x)$ consists of m component functions. The constraint $c_i(x) \geq 0$ is said to be *active* at $\bar{x}$ if $c_i(\bar{x}) = 0$ and *inactive* if $c_i(\bar{x}) > 0$. Let $\hat{A}(x)$ denote the Jacobian of the *active* constraints at x, and let $N(x)$ denote a matrix whose columns form a basis for the null space of $\hat{A}$.

Nonlinear constraints can be extremely complicated, and necessary optimality conditions can be stated only after making assumptions about the constraints (called regularity assumptions or constraint qualifications); see, for example, Avriel (1976), Fiacco and McCormick (1968) or Fletcher (1987). The most common form of constraint qualification is an assumption that the gradients of the active constraints are linearly independent (or that the constraints are linear).

The Lagrangian function for problem (2.10) is defined as

$$L(x, \lambda) = f(x) - \lambda^T c(x) \tag{2.11}$$

(see (2.8)). For future reference, we note that the Hessian of the Lagrangian with respect to x, denoted by W, is given by

$$W(x, \lambda) = \nabla^2 L(x, \lambda) = H(x) - \sum_{i=1}^{m} \lambda_i H_i(x). \tag{2.12}$$

If a suitable constraint qualification holds at x^*, the following conditions can be shown to be necessary for the point x^* to be a constrained minimizer of (2.10):

$$\begin{aligned} c(x^*) &\geq 0; && (2.13a)\\ g(x^*) &= A(x^*)^T \lambda^* \quad \text{for some } \lambda^*; && (2.13b)\\ \lambda_i^* c_i(x^*) &= 0, \quad i = 1, \ldots, m; && (2.13c)\\ \lambda_i^* &\geq 0, \quad i = 1, \ldots, m; && (2.13d)\\ N(x^*)^T W(x^*, \lambda^*) N(x^*) &\geq 0. && (2.13e) \end{aligned}$$

Condition (2.13c), which forces at least one of $c_i(x^*)$ and λ_i^* to be zero for every i, is called a *complementarity* condition. In particular, it means that if $c_i(x^*) > 0$, i.e. constraint i is *inactive* at x^*, then λ_i^* must be zero.

Because the multipliers for inactive constraints are zero, the first-order condition (2.13b) states that the gradient of f at x^* is a linear combination of the *active* constraint gradients, so that $N(x^*)^T g(x^*) = 0$. Trivial rearrangement of (2.13b) also reveals that the gradient of the Lagrangian function with respect to x vanishes at x^* when $\lambda = \lambda^*$, i.e. x^* is a stationary point of the Lagrangian function when $\lambda = \lambda^*$. However, x^* is not necessarily a minimizer of the Lagrangian function.

A crucial distinction arising from constraint nonlinearities can be seen in the second-order condition (2.13e), which involves the reduced Hessian of the Lagrangian, rather than the reduced Hessian of f alone. The inclusion of constraint curvature is an essential feature of efficient algorithms for nonlinearly constrained problems.

Sufficient conditions for x^* to be an isolated constrained minimizer of (2.10) are: (i) a suitable constraint qualification applies at x^* (for example, the gradients of the active constraints at x^* are linearly independent); (ii) conditions (2.13a–c) are satisfied; and (iii) the following strengthened versions of (2.13d–e) hold:

$$\lambda_i^* > 0 \quad \text{if} \quad c_i(x^*) = 0; \tag{2.14}$$

$$N(x^*)^T W(x^*, \lambda^*) N(x^*) > 0. \tag{2.15}$$

Inequality (2.14) is called *strict complementarity*, and holds when all Lagrange multipliers associated with active constraints are positive. The property of strict complementarity is often assumed because the presence of a zero multiplier for an active constraint creates complications.

Condition (2.15) is equivalent to the existence of $\alpha > 0$ such that

$$p^T W(x^*, \lambda^*)p \geq \alpha \|p\|^2 \quad \text{for all } p \text{ such that } \hat{A}(x^*)p = 0. \tag{2.16}$$

2.3. Convexity

Most work on interior methods to date has focused on convex optimization problems, of which linear programming is the most obvious instance. As we shall see, many complications that arise for general optimization problems disappear in the presence of convexity. See Rockafellar (1970) for a complete treatment of convex analysis.

Definition 8 *(Convex set.) The set S is convex if, for every x_1 and x_2 in S, and for all θ satisfying $0 \leq \theta \leq 1$, the point $z = (1-\theta)x_1 + \theta x_2$ is also in S.*

Definition 9 *(Convex and concave functions.) The function $f(x)$, defined for x in a nonempty open convex set S, is convex if, for every two points x_1 and x_2 in S, and for all θ satisfying $0 \leq \theta \leq 1$,*

$$f\Big((1-\theta)x_1 + \theta x_2\Big) \leq (1-\theta)f(x_1) + \theta f(x_2). \tag{2.17}$$

(If the set S is not specified, it is assumed to be $\mathcal{R}^n$.) The function f is concave if $-f$ is convex. The function f is strictly convex if the inequality in (2.17) is strict when $x_1 \neq x_2$ and $0 < \theta < 1$.

Several useful results associated with convexity are:

1 the intersection of a finite number of convex sets is convex;

2 all level sets of a convex function are convex;

3 given a set of convex functions $\{\varphi_i(x)\}$, $i = 1, \ldots, m$, the set of points satisfying $\varphi_i(x) \leq 0$ is convex;

4 the smooth function $f(x)$, defined for x in an open convex set S, is convex if its Hessian matrix $H(x)$ is positive semi-definite for all $x \in S$, and strictly convex if $H(x)$ is positive definite for all $x \in S$.

Convex programs are constrained optimization problems with important special properties. It should be stressed that the only *equality* constraints permitted in a convex program are linear constraints.

Definition 10 *(Convex program.) The problem of minimizing $f(x)$ subject to the linear equality constraints $Ax = b$ and the inequality constraints $c_i(x) \geq 0$, $i = 1, \ldots, m$, is a convex program if $f(x)$ is convex and $-c_i(x)$ is convex for $i = 1, \ldots, m$.*

A slight irritation is that our generic form for inequality constraints involves a 'greater than' relation ($c_i(x) \geq 0$) for expositional convenience. Unfortunately, an optimization problem with constraints in this form is a convex program only if each *negative* constraint function $-c_i(x)$ is convex, i.e. if $c_i(x)$ itself is concave. Hence minus signs appear throughout our discussion of the constraints in convex programs.

Using these definitions, it is easy to see that a linear function is convex (and also concave), and that a linear programming problem is a convex program. Two properties that are important in interior methods for linear programming are stated formally in the following theorems; see Fiacco and McCormick (1968) or Fletcher (1987) for details.

Theorem 1 *If x^* is a local constrained minimizer of a convex programming problem, it is also a global constrained minimizer. Further, the set of minimizers of a convex program is convex.*

Theorem 2 *If the optimization problem (2.10) is a convex program, and if x^* satisfies the feasibility and first-order necessary conditions (2.13a–d), then x^* is a global constrained minimizer of (2.10).*

3. Barrier methods

3.1. Intuition and motivation

Suppose that we wish to minimize $f(x)$ subject to a set of inequality constraints $c_i(x) \geq 0$, $i = 1, \ldots, m$. If the constraints affect the solution, either an unconstrained minimizer of $f(x)$ is infeasible (for example, when minimizing x^2 subject to $x \geq 1$), or else $f(x)$ is unbounded below when the constraints are removed (for example, when minimizing x^3 subject to $x \geq 1$). Consequently, if an optimization method tries to achieve a 'large' reduction in the objective function from its value at a feasible point, the iterates tend to move *outside* the feasible region. In fact, many popular algorithms for nonlinearly constrained optimization (such as SQP methods; see, for example, Fletcher (1987), and Gill *et al.* (1981)) typically produce infeasible iterates that approach feasibility only in the limit.

When feasibility at intermediate points is essential – for example, in practical problems where the objective function is meaningless unless the constraints are satisfied – it seems desirable for iterates to approach the constrained solution from the *interior* of the feasible region. *Barrier methods* constitute a well known class of methods with this property.

Barrier methods may be applied only to inequality constraints for which strictly feasible points exist. This property does not hold for all inequality constraints, even if the feasible region is nonempty; for example, consider the constraints $x_1 + x_2 \geq 0$ and $-x_1 - x_2 \geq 0$, for which the feasible region consists of the line $\{x_1 + x_2 = 0\}$.

Given an initial strictly feasible point and mild assumptions about the feasible region, strict feasibility can be retained by minimizing a composite function consisting of the original objective $f(x)$ plus a positive multiple of an infinite 'barrier' at the boundary of strict($\mathcal{F}$). The most effective methods for unconstrained optimization (such as Newton's method; see Section 4) require differentiability. A suitable barrier term is therefore composed of functions that are smooth at strictly feasible points, but contain a positive singularity if any constraint is zero. Under these conditions, a minimizer of the composite function must occur at a strictly feasible point.

When the barrier term is heavily weighted, a minimizer of the composite function will lie, informally speaking, 'far away' from the boundary. If the coefficient of the barrier term is reduced, the singularity becomes less influential, except at points near the boundary; minimizers of the composite function can then move closer (but not 'too close') to the boundary. The weight on the barrier term thus tends to regulate the distance from the iterates to the boundary. In the parlance of modern interior methods, the barrier term forces the iterates to remain *centred* in the strictly feasible region.

As the factor multiplying the barrier term decreases to zero, intuition suggests that minimizers of the composite function will converge to a constrained solution x^* that lies on the boundary of strict($\mathcal{F}$). We shall see later (Sections 3.3 and 3.4) that this intuition can be verified rigorously under reasonably mild conditions.

We stress that there is ample room for many formulations of a 'barrier function', as indicated by the range of definitions in Fiacco and McCormick (1968) and in Nesterov and Nemirovsky (1989). Other varieties of composite functions – called 'potential' and 'centering' functions – have also been proposed for use in interior methods; see, for example, Sonnevend (1986) and Gonzaga (1992). Karmarkar's original (1984) LP algorithm included a logarithmic potential function. The method of centres of Huard (1967) imposes an additional constraint at each iteration based on the current value of the objective function; see Renegar (1988) for an LP method based on this idea.

In all cases, the composite functions display a common motivation of simultaneously reflecting the objective function (thereby encouraging its reduction) as well as forcing iterates to stay 'nicely centred' in the feasible region. They differ, however, in the balance of these sometimes conflicting aims.

3.2. The logarithmic barrier function

For simplicity, we discuss only the simplest barrier function based on a logarithmic singularity, which was not only the most popular in the 1960s, but also has received substantial attention since 1984. The logarithmic barrier function was first defined by Frisch in 1955, and was extensively studied and analysed during the 1960s. Detailed theoretical discussions of classical barrier methods, along with historical background, are given in Fiacco and McCormick (1968) and Fiacco (1979).

The *logarithmic barrier function* associated with minimizing $f(x)$ subject to $c(x) \geq 0$ is

$$B(x, \mu) = f(x) - \mu \sum_{i=1}^{m} \ln c_i(x), \tag{3.1}$$

where the *barrier parameter* μ is strictly positive. (When the meaning is clear, we may write B with a single argument μ or without arguments.) Since the logarithm is undefined for nonpositive arguments, the logarithmic barrier function is defined only in strict$(\mathcal{F})$.

Simply stating the definition (3.1) does not give an adequate impression of the dramatic effects of the imposed barrier. Figure 1 depicts the one-dimensional variation of a barrier function for two values of μ. Even for the modest value $\mu = 0.1$, the (visually) extreme steepness of the singularity is evident.

Figure 1. The one-dimensional behaviour of a barrier function.

The intuitive motivation for a barrier method is that we seek unconstrained minimizers of $B(x, \mu)$ for values of μ decreasing to zero. If the solution x^* of the constrained problem lies on the boundary and exact arithmetic is used, a barrier method can never produce the exact solution. Barrier methods consequently terminate when the current iterate satisfies some approximation to the desired optimality conditions. 'Classical' barrier algorithms as well as many recent interior methods have the following form:

Generic Barrier Algorithm

0. Set x_0 to a strictly feasible point, so that $c(x_0) > 0$, and set μ_0 to a positive value; $k \leftarrow 0$.

1. Check whether x_k qualifies as an approximate local constrained minimizer for the original problem (2.10). If so, stop with x_k as the solution.

2. Compute an unconstrained minimizer $x(\mu_k)$ of $B(x, \mu_k)$.

3. $x_{k+1} \leftarrow x(\mu_k)$; choose $\mu_{k+1} < \mu_k$; $k \leftarrow k+1$; return to Step 1.

In practice, the calculation of $x(\mu_k)$ in Step 2 is carried out approximately, and only a few iterations of an unconstrained method may be performed before the barrier parameter is updated. In the theoretical results given here, we assume that $x(\mu_k)$ is an exact unconstrained minimizer.

We illustrate the behaviour of the generic algorithm on a simple two-variable example:

$$\begin{array}{ll} \text{minimize} & x_1x_2 - \frac{1}{2}x_1^2 - x_2 \\ \text{subject to} & x_1^2 + x_2^2 \le 2 \\ & x_1^2x_2^2 \quad \le 10. \end{array}$$

The first constraint is satisfied inside the circle of radius $\sqrt{2}$ centred at the origin; although the second constraint is redundant, it nonetheless affects each minimizer of the barrier function. The point $x^* = (-1,1)^T$ is an isolated local constrained minimizer at which only the first constraint is active. Figure 2 depicts selected barrier minimizers converging to x^*, which lies on the boundary of the feasible region (depicted as a dashed curve).

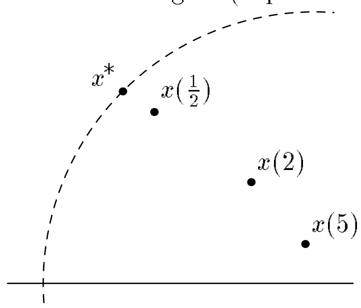

Figure 2. Convergence of barrier minimizers to $x^* = (-1,1)^T$.

The next two sections provide a rigorous foundation for the generic approach, including the assumptions necessary to make it succeed in converging to a solution x^* of the original constrained problem. After establishing local convergence properties, we return in Section 3.5 to a more detailed analysis of the sequence of barrier minimizers.

3.3. Theoretical results for convex programs

Pre-1984 presentations of barrier methods for nonlinear problems typically begin with general results, which are then specialized to convex programs. We have chosen instead to give a self-contained presentation of the convex results first. Readers whose primary interest is in interior methods for linear and convex programming can read this section only and skip to Section 3.5.

Consider the convex programming problem

$$\underset{x\in\mathcal{R}^n}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c_i(x) \ge 0, \quad i = 1,\ldots,m, \tag{3.2}$$

where f and $\{-c_i\}$ are convex. In this section, $\mathcal{F}$ denotes the feasible region for the constraints of (3.2). Recall from Theorem 1 that every local minimizer of a convex program is a global minimizer; hence, if any minimizer exists, the optimal value of f in $\mathcal{F}$ is unique.

An obvious fundamental question involves the conditions under which a solution x^* of (3.2) is the limit of a sequence of unconstrained minimizers of the barrier function. The main assumption needed to prove convergence results is that the set $\mathcal{M}$ of minimizers of (3.2) is *bounded.* (We know already from Theorem 1 that $\mathcal{M}$ is convex.) Boundedness of the set of minimizers holds automatically under the much stronger assumption that the feasible region itself is bounded.

The major results of this section are given in Theorem 5. Two other theorems serve as a prelude.

Theorem 3 (a version of Theorem 24 in Fiacco and McCormick (1968)) shows that, if a set of convex functions defines a *bounded* feasible region, then suitably perturbed versions of the same functions also define a bounded feasible region. The application of this theorem in the proof of Theorem 4 involves a level set derived from the objective function.

Theorem 3 *(Boundedness of perturbed convex sets.) Let $-\varphi_i(x)$ be a convex function for $i = 1, \ldots, m$, and assume that the convex set*

$$\mathcal{N} = \{x \mid \varphi_i(x) \geq 0, \quad i = 1, \ldots, m\}$$

is nonempty and bounded. Then for any set of values $\{\Delta_i\}$, where $\Delta_i \geq 0$, $i = 1, \ldots, m$, the set

$$\{x \mid \varphi_i(x) \geq -\Delta_i, \quad i = 1, \ldots, m\}$$

is bounded.

Proof. The result will follow in an obvious way if verified for $\Delta_1 > 0$ and $\Delta_i = 0$, $i \neq 1$. Given $\Delta_1 > 0$, let $\mathcal{N}_1$ denote the set

$$\mathcal{N}_1 = \{x \mid \varphi_1(x) \geq -\Delta_1 \text{ and } \varphi_i \geq 0, \;\; i = 2, \ldots, m\}.$$

Because $\mathcal{N}_1$ is the intersection of a finite number of convex sets, $\mathcal{N}_1$ is convex.

To prove by contradiction that $\mathcal{N}_1$ is bounded, we assume the contrary: for any point $x_1 \in \mathcal{N}$, there exists a ray emanating from x_1 that does not intersect the boundary of $\mathcal{N}_1$, so that $x_1 + \alpha p$ lies in $\mathcal{N}_1$ for some direction p and any $\alpha \geq 0$. (The fact that any unbounded convex set must contain a ray is standard; see, for example, Grünbaum (1967).)

Because $\mathcal{N}$ is bounded by assumption, there must be a point x_2 on this ray that does *not* lie in $\mathcal{N}$. Let x_2 be such a point, given by $x_2 = x_1 + \alpha_2 p$ for some $\alpha_2 > 0$, for which φ_1 assumes a *negative* value, say $\varphi_1(x_2) = -\delta < 0$, where $\delta < \Delta_1$.

Let x_3 denote a point on the ray that lies *beyond* x_2, i.e. $x_3 = x_1 + \alpha_3 p$, where $\alpha_3 > \alpha_2$. The point x_2 can then be written as

$$x_2 = x_1 + \theta(x_3 - x_1) = (1 - \theta)x_1 + \theta x_3, \tag{3.3}$$

where $0 < \theta < 1$.

Applying Definition 9 of a convex function to the expression (3.3) for x_2, we obtain

$$(1 - \theta)\varphi_1(x_1) + \theta\varphi_1(x_3) \leq \varphi_1(x_2) = -\delta,$$

which gives

$$\theta\varphi_1(x_3) \leq -\delta - (1 - \theta)\varphi_1(x_1).$$

Because $\varphi_1(x_1) \geq 0$ and $0 < \theta < 1$, it follows that

$$\varphi_1(x_3) \leq \frac{-\delta}{\theta}.$$

If θ is sufficiently small, namely $\theta < \delta/\Delta_1$, the value of $\varphi_1(x_3)$ must be strictly less than $-\Delta_1$, which shows that x_3 cannot lie in $\mathcal{N}_1$. This gives the desired contradiction, and shows that $\mathcal{N}_1$ must be bounded. □

The next result is related to Lemma 12 in Fiacco and McCormick (1968), which applies to a general barrier function. Given a convex program with a nonempty strict interior and a bounded set of minimizers, the theorem states that any particular level set of the logarithmic barrier function is bounded and closed. The boundedness property is important because it implies that the set of *minimizers* of the barrier function is bounded.

Theorem 4 *(Compactness of barrier function level sets.) Consider the convex program of minimizing $f(x)$ subject to $c_i(x) \geq 0$, $i = 1, \ldots, m$. Let $\mathcal{F}$ denote the (convex) feasible region. Assume that* strict$(\mathcal{F})$ *is nonempty and that the set of minimizers $\mathcal{M}$ for the convex program is nonempty and bounded. Then for any $\mu_k > 0$ and any constant τ, the level set*

$$S(\tau) = \{x \in \text{strict}(\mathcal{F}) \mid B(x, \mu_k) \leq \tau\}$$

is bounded and closed, where $B(x, \mu_k)$ is the logarithmic barrier function.

Proof. Boundedness of $S(\tau)$ will be established by showing that, under the stated assumptions, the barrier function cannot remain bounded above while its argument becomes unbounded.

Let $\hat{x}$ denote any point in strict$(\mathcal{F})$ (which is assumed to be nonempty). Given any $\epsilon > 0$, let $\hat{D}$ denote the level set defined by the values of $f(\hat{x})$ and ϵ:

$$\hat{D} = \{x \in \mathcal{F} \mid f(x) \leq f(\hat{x}) + \epsilon\}. \tag{3.4}$$

Convexity of f implies that $\hat{D}$ is convex (see Section 2.3). The functions f and $\{c_i\}$ are smooth, so that $\hat{D}$ is closed. The first step in proving the theorem is to show that $\hat{D}$ is bounded, from which it will follow that $\hat{D}$ is compact.

To show that the set $\hat{D}$ is bounded, we invoke Theorem 3. Let f^* denote the minimum value of $f(x)$ for $x \in \mathcal{F}$. Because every local minimizer of a convex program is a global minimizer (see Theorem 1), the quantity $\Delta = f(\hat{x})+\epsilon-f^*$ must be positive. By assumption, the set $\mathcal{M}$, which may be written as $\{x \in \mathcal{F} \mid f(x) \leq f^*\}$, is nonempty and bounded; further, $\mathcal{M}$ is convex because f is convex. We now define the function $\phi(x)$ as $f^* - f(x)$, and observe that $-\phi$ is convex. Theorem 3 then applies to ϕ and the positive perturbation Δ, and implies *boundedness* of the set

$$\{x \in \mathcal{F} \mid \phi(x) \geq -\Delta\} = \{x \in \mathcal{F} \mid f^* - f(x) \geq f^* - f(\hat{x}) - \epsilon\},$$

which is simply a rearranged definition of $\hat{D}$. Consequently, $\hat{D}$ is compact. It is straightforward to see that its boundary, bnd$(\hat{D})$, is also compact. The definition (3.4) of $\hat{D}$ shows that $\hat{x}$ does not lie on the boundary of $\hat{D}$.

Having established the compactness of $\hat{D}$ and its boundary, we can now prove boundedness of $S(\tau)$ by contradiction. Assume the contrary of the desired result, namely that for some $\mu_k > 0$, there is an *unbounded* sequence $\{y_j\}$ of points in strict$(\mathcal{F})$ for which the barrier function values $B(y_j, \mu_k)$ remain bounded above.

For such a sequence, let j be sufficiently large so that y_j lies outside $\hat{D}$. By definition of $\hat{D}$, it must hold that

$$f(y_j) > f(\hat{x}) + \epsilon.$$

Let z_j be the point on the boundary of $\hat{D}$ where the line connecting $\hat{x}$ and y_j intersects the boundary. (Because $\hat{D}$ is convex, z_j is unique.) Let λ_j be the scalar satisfying $0 < \lambda_j < 1$ such that

$$z_j = (1 - \lambda_j)\hat{x} + \lambda_j y_j. \tag{3.5}$$

We have assumed that $\|y_j\|$ is unbounded for sufficiently large j. Since $\|z_j\|$ is finite, (3.5) shows that

$$\lambda_j \to 0 \quad \text{as} \quad j \to \infty. \tag{3.6}$$

Because $\hat{x}$ and y_j are both in strict$(\mathcal{F})$, we know that $c_i(\hat{x}) > 0$ and $c_i(y_j) > 0$ for $i = 1$, $\ldots$, m. Convexity of $-c_i(x)$ combined with (3.5) gives

$$c_i(z_j) \geq (1 - \lambda_j)c_i(\hat{x}) + \lambda_j c_i(y_j) > 0, \tag{3.7}$$

which shows that $z_j \in \text{strict}(\mathcal{F})$. Since z_j is by definition in $\text{bnd}(\hat{D})$, we conclude from (3.4) that $f(z_j) = f(\hat{x}) + \epsilon$. Because f is convex (see Definition 9), (3.5) implies

$$f(z_j) \le (1 - \lambda_j) f(\hat{x}) + \lambda_j f(y_j).$$

Dividing by λ_j and substituting $f(z_j) = f(\hat{x}) + \epsilon$, we obtain a *lower bound* on $f(y_j)$:

$$f(y_j) \ge f(\hat{x}) + \frac{\epsilon}{\lambda_j}. \tag{3.8}$$

It then follows from (3.6) that

$$f(y_j) \to \infty \quad \text{as} \quad j \to \infty,$$

so that the objective function values at $\{y_j\}$ become unbounded.

Turning back to the constraint functions, positivity of λ_j means that the first inequality in (3.7) can be rewritten as

$$c_i(y_j) \le c_i(\hat{x}) + \frac{c_i(z_j) - c_i(\hat{x})}{\lambda_j}. \tag{3.9}$$

Since the set $\text{bnd}(\hat{D})$ is compact, the function $c_i(x) - c_i(\hat{x})$ achieves its maximum for some $x \in \text{bnd}(\hat{D})$. Let d_i denote

$$d_i = \max\{c_i(x) - c_i(\hat{x}) \mid x \in \text{bnd}(\hat{D})\}.$$

We now wish to demonstrate that $d_i \ge 0$. Because $z_j \in \text{bnd}(\hat{D})$ and $c_i(y_j) > 0$, we apply the definition of d_i and relation (3.9) to show that

$$c_i(\hat{x}) + \frac{d_i}{\lambda_j} \ge c_i(y_j) > 0, \quad i = 1, \ldots, m. \tag{3.10}$$

If d_i were negative, the first expression in (3.10) would eventually become negative as $\lambda_j \to 0$, which is impossible. It follows that $d_i \ge 0$ for $i = 1, \ldots, m$.

Finally, the barrier function $B(y_j, \mu_k)$ is formed. Using (3.8), (3.10), monotonicity of the logarithm function, and positivity of μ_k, we have:

$$\begin{aligned} B(y_j, \mu_k) &= f(y_j) - \mu_k \sum \ln c_i(y_j) \\ &\ge f(\hat{x}) + \frac{\epsilon}{\lambda_j} - \mu_k \sum \ln(c_i(\hat{x}) + (d_i/\lambda_j)) \\ &= f(\hat{x}) + \frac{\epsilon - \mu_k \lambda_j \sum \ln(c_i(\hat{x}) + (d_i/\lambda_j))}{\lambda_j}. \end{aligned} \tag{3.11}$$

The logarithm function has the property that, for a positive constant ν and $\delta \ge 0$,

$$\lim_{\lambda \to 0_+} \lambda \ln\left(\nu + \frac{\delta}{\lambda}\right) = 0.$$

Thus the limit of the numerator in (3.11) is ϵ, and the quotient in (3.11) is unbounded above as $\lambda_j \to 0$. It follows that $B(y_j, \mu_k)$ is unbounded above as $j \to \infty$, thereby contradicting our assumption that the barrier function values $\{B(y_j, \mu_k)\}$ are bounded above for an unbounded sequence $\{y_j\}$. This proves that $S(\tau)$ is bounded.

To show that $S(\tau)$ is closed, we prove that it contains all its accumulation points. Let $\{x_j\}$ be a convergent sequence in $S(\tau)$, with limit point $\bar{x}$. It follows from the continuity of f and $\{c_i\}$ in strict$(\mathcal{F})$ that $\bar{x}$ must satisfy $B(\bar{x}, \mu_k) \le \tau$. Further, $\bar{x}$ must either be in strict$(\mathcal{F})$ or else have the property that $c_i(\bar{x}) = 0$ for at least one index i.

If $\bar{x}$ is in strict$(\mathcal{F})$, by definition $\bar{x}$ is in $S(\tau)$. Suppose that $\bar{x}$ is not in strict$(\mathcal{F})$. Then, because $c_i(\bar{x}) = 0$ for some index i, unboundedness of the logarithm for a zero argument and convergence of $\{x_j\}$ to $\bar{x}$ together imply that, for sufficiently large j, the barrier term $-\sum_{i=1}^m \ln c_i(x_j)$ cannot be bounded above. In particular, for any constant γ and sufficiently large j,

$$-\sum_{i=1}^{m} \ln c_i(x_j) > \gamma. \tag{3.12}$$

We now define γ as $\gamma = (\tau - f^*)/\mu_k$; the value of γ is finite because f^* is finite. Since x_j lies in strict$(\mathcal{F})$, we know from the convexity of f that $f(x_j) \ge f^*$, which means that $-f^* \ge -f(x_j)$. Applying this inequality and the definition of γ in (3.12), we obtain

$$-\sum_{i=1}^{m} \ln c_i(x_j) > \frac{\tau - f(x_j)}{\mu_k}.$$

After rearrangement, this relation implies that $B(x_j, \mu_k) > \tau$, i.e., that $x_j \notin S(\tau)$, a contradiction. We conclude that any accumulation point of a sequence in $S(\tau)$ must lie in $S(\tau)$, which means that $S(\tau)$ is closed.

We have shown that $S(\tau)$ is both bounded and closed; its compactness is immediate. $\square$

We are now ready to give the main theorem concerning barrier methods for convex programs. The most important result is (vi), which shows that limit points of a minimizing sequence for the barrier function converge to constrained minimizers of the convex program.

Theorem 5 *(Convergence of barrier methods on convex programs.) Consider the convex program of minimizing $f(x)$ subject to $c_i(x) \ge 0$, $i = 1, \ldots, m$. Let $\mathcal{F}$ denote the feasible region for this problem, and assume that* strict$(\mathcal{F})$ *is nonempty. Let $\{\mu_k\}$ be a decreasing sequence of positive barrier parameters such that* $\lim_{k\to\infty} \mu_k = 0$. *Assume that the set $\mathcal{M}$ of constrained local minimizers of the convex program is nonempty and bounded, and let f^* denote the optimal value of f. Then*

(i) the logarithmic barrier function $B(x, \mu_k)$ is convex in strict$(\mathcal{F})$;

(ii) $B(x, \mu_k)$ has a finite unconstrained minimizer in strict$(\mathcal{F})$ for *every* $\mu_k > 0$, and the set $\mathcal{M}_k$ of unconstrained minimizers of $B(x, \mu_k)$ in strict$(\mathcal{F})$ is convex and compact for every k;

(iii) any unconstrained local minimizer of $B(x, \mu_k)$ in strict$(\mathcal{F})$ is also a *global* unconstrained minimizer of $B(x, \mu_k)$;

(iv) let y_k denote an unconstrained minimizer of $B(x, \mu_k)$ in strict$(\mathcal{F})$; then, for all k,

$$f(y_{k+1}) \le f(y_k) \quad \text{and} \quad -\sum_{i=1}^{m} \ln c_i(y_k) \le -\sum_{i=1}^{m} \ln c_i(y_{k+1});$$

(v) there exists a compact set S such that, for all k, every minimizing point y_k of $B(x, \mu_k)$ lies in $S \cap \text{strict}(\mathcal{F})$;

(vi) any sequence $\{y_k\}$ of unconstrained minimizers of $B(x, \mu_k)$ has at least one convergent subsequence, and every limit point of $\{y_k\}$ is a local constrained minimizer of the convex program;

(vii) let $\{x_k\}$ denote a convergent subsequence of unconstrained minimizers of $B(x, \mu_k)$; then $\lim\limits_{k\to\infty} f(x_k) = f^*$;

(viii) $\lim_{k\to\infty} B_k = f^*$, where B_k denotes $B(x_k, \mu_k)$.

Proof. It is straightforward to prove convexity of $B(x, \mu_k)$ using the convexity of f and $\{-c_i\}$, monotonicity of the logarithm function and Definition 9 of a convex function. Thus (i) is established.

The assumptions of this theorem are the same as those of Theorem 4. Let x_0 denote the strictly feasible point at which the barrier iterations are initiated. For the barrier parameter μ_k and some $\epsilon > 0$, we define the set S_0 as:

$$S_0 = \{x \in \text{strict}(\mathcal{F}) \mid B(x, \mu_k) \le B(x_0, \mu_k) + \epsilon\}.$$

Theorem 4 implies that S_0 is *compact* for all $\mu_k > 0$. It follows that the smooth function $B(x, \mu_k)$ assumes its minimum in S_0, necessarily at an interior point of S_0. We then apply Definition 7 and conclude that $B(x, \mu_k)$ has at least one finite *unconstrained* minimizer.

Because $B(x, \mu_k)$ is convex, any local minimizer is also a global minimizer, so that *every* unconstrained minimizer of $B(x, \mu_k)$ must be in the set S_0. Thus the set $\mathcal{M}_k$ of unconstrained minimizers of $B(x, \mu_k)$ is bounded. The set $\mathcal{M}_k$ is closed because the minimum value of $B(x, \mu_k)$ is unique, and it follows that $\mathcal{M}_k$ is compact. Convexity of $\mathcal{M}_k$ follows from Theorem 1, and result (ii) has been verified.

Result (iii) follows from Theorem 1, and results (i) and (ii).

To show result (iv), let y_k and y_{k+1} denote global minimizers of the barrier function for the barrier parameters μ_k and μ_{k+1}. By definition of y_k and y_{k+1} as minimizers, we have

$$\begin{aligned} f(y_k) - \mu_k \sum_{i=1}^{m} \ln c_i(y_k) &\le f(y_{k+1}) - \mu_k \sum_{i=1}^{m} \ln c_i(y_{k+1}); \\ f(y_{k+1}) - \mu_{k+1} \sum_{i=1}^{m} \ln c_i(y_{k+1}) &\le f(y_k) - \mu_{k+1} \sum_{i=1}^{m} \ln c_i(y_k). \end{aligned} \tag{3.13}$$

We multiply the first of these inequalities by the ratio μ_{k+1}/μ_k, which lies strictly between 0 and 1, add the resulting inequality to the second inequality, cancel the terms involving logarithms and obtain

$$f(y_{k+1})\Big(1 - \frac{\mu_{k+1}}{\mu_k}\Big) \le f(y_k)\Big(1 - \frac{\mu_{k+1}}{\mu_k}\Big).$$

Since $0 < \mu_{k+1} < \mu_k$, it follows that $f(y_{k+1}) \le f(y_k)$. Applying this result in (3.13) and dividing by the positive number μ_k, we obtain

$$-\sum_{i=1}^{m} \ln c_i(y_k) \le -\sum_{i=1}^{m} \ln c_i(y_{k+1}), \tag{3.14}$$

as required for the second part of (iv).

To verify existence of the set S in (v), we use result (iv). Let f_k denote $f(y_k)$. Since $f_{k+1} \le f_k$ for each k, the compact convex level set $\{x \in \mathcal{F} \mid f(x) \le f_k\}$ not only contains all minimizers of $B(x, \mu_k)$, but also contains all minimizers of $B(x, \mu_{k+1})$. The compact convex level set S defined by the strictly feasible point x_0,

$$S = \{x \in \mathcal{F} \mid f(x) \le f(x_0)\}, \tag{3.15}$$

accordingly contains $\mathcal{M}$ as well as all minimizers of $B(x, \mu_k)$ for all k.

Now we show (vi). It follows from the last statement of the preceding paragraph that every minimizer y_k must lie in the compact set S defined by (3.15). We conclude that the sequence $\{y_k\}$ is uniformly bounded, and hence contains at least one convergent subsequence, say with limit point $\hat{x}$. Because y_k lies in S for all k, $\hat{x}$ must be feasible.

To prove that $\hat{x}$ is a local constrained minimizer of the convex program, we assume otherwise, that $\hat{x} \notin \mathcal{M}$. Since every local solution of a convex program is a global solution, this would imply that $f(\hat{x}) > f^*$. A contradiction is now established from this inequality and the definition of $\hat{x}$ as a limit point of a convergent subsequence of minimizers of $B(x, \mu_k)$.

Let $\{x_k\}$ denote a subsequence of $\{y_k\}$ converging to $\hat{x}$. Continuity of f and the relation $f_k \geq f_{k+1}$ imply that, for all k,

$$f(x_k) \geq f(\hat{x}). \tag{3.16}$$

We next show that there must exist a strictly feasible point x_{int} such that

$$f(\hat{x}) > f(x_{\text{int}}).$$

Let x^* denote any point in the set $\mathcal{M}$ of constrained minimizers, so that $f(x^*) = f^*$ and x^* is in the set S defined by (3.15). If x^* itself is strictly feasible, we simply take $x_{\text{int}} = x^*$, since our initial assumption was that $f(\hat{x}) > f(x^*)$.

If x^* is not strictly feasible, x_{int} is found as follows. By assumption, strict$(\mathcal{F})$ is nonempty, and hence contains at least one point, say z; the definition and uniqueness of f^* guarantee that $f(z) \geq f(x^*)$. If $f(z) < f(\hat{x})$, z may be taken as x_{int}. If $f(z) \geq f(\hat{x})$, consider a generic point $\bar{x}$ on the line segment joining x^* and z, defined by $\bar{x} = (1-\lambda)x^* + \lambda z$ for λ satisfying $0 < \lambda < 1$. Because z is strictly feasible and $-c_i$ is convex for $i = 1, \ldots, m$, we have

$$c_i(\bar{x}) \geq (1-\lambda)c_i(x^*) + \lambda c_i(z) > 0,$$

so that $\bar{x}$ is strictly feasible.

Convexity of f implies that

$$f(\bar{x}) \leq (1-\lambda)f(x^*) + \lambda f(z),$$

where $f(z) \geq f(\hat{x}) > f(x^*)$. Using continuity of f, we see that $f(\bar{x}) < f(\hat{x})$ for some suitably small λ, namely λ such that

$$\lambda < \frac{f(\hat{x}) - f(x^*)}{f(z) - f(x^*)} \leq 1. \tag{3.17}$$

For any λ satisfying (3.17), $\bar{x}$ may be taken as x_{int}.

Thus far we have shown that, if $\hat{x}$ is not in the minimizing set $\mathcal{M}$, then a strictly feasible point x_{int} exists such that

$$f(x_k) \geq f(\hat{x}) > f(x_{\text{int}}). \tag{3.18}$$

Since x_k is a global minimizer of $B(x, \mu_k)$,

$$f(x_k) - \mu_k \sum_{i=1}^{m} \ln c_i(x_k) \leq f(x_{\text{int}}) - \mu_k \sum_{i=1}^{m} \ln c_i(x_{\text{int}}). \tag{3.19}$$

The barrier term involving x_{int} in (3.19) is finite, and

$$\lim_{k\to\infty} B(x_{\text{int}}, \mu_k) = f(x_{\text{int}}).$$

If the limit point $\hat{x}$ of $\{x_k\}$ is also strictly feasible, the barrier term involving x_k in (3.19) is similarly finite as $k \to \infty$, and

$$\lim_{k\to\infty} B(x_k, \mu_k) = f(\hat{x}).$$

Letting $k \to \infty$ in (3.19), we obtain the inequality $f(\hat{x}) \leq f(x_{\text{int}})$, which contradicts (3.18).

Suppose, on the other hand, that $\hat{x}$ is not in strict($\mathcal{F}$), so that $c_i(\hat{x}) = 0$ for at least one index i. Adding a barrier term involving $x_{\rm int}$ to both sides of the inequality $f(x_{\rm int}) < f(x_k)$, we have

$$f(x_{\rm int}) - \mu_k \sum_{i=1}^{m} \ln c_i(x_{\rm int}) < f(x_k) - \mu_k \sum_{k=1}^{m} \ln c_i(x_{\rm int}).$$

Combining this inequality with (3.19), rearranging and dividing by μ_k, we obtain

$$f(x_k) - \mu_k \sum_{k=1}^{m} \ln c_i(x_k) < f(x_k) - \mu_k \sum_{k=1}^{m} \ln c_i(x_{\rm int}).$$

Cancelling $f(x_k)$ from both sides then gives

$$-\sum_{k=1}^{m} \ln c_i(x_k) < -\sum_{k=1}^{m} \ln c_i(x_{\rm int}).$$

The sum on the right-hand side involving $x_{\rm int}$ is finite. However, since $\hat{x}$ is not strictly feasible, $-\ln c_i(x_k)$ approaches infinity for at least one i. The left-hand side is therefore unbounded above, and we again have a contradiction.

The conclusion is that $\hat{x}$ lies in $\mathcal{M}$, the set of minimizers. Because $\hat{x}$ is the limit point of $\{x_k\}$, we have obtained the crucial result (vi). For the remainder of the proof, x^* will denote the limit point of $\{x_k\}$.

Part (vii) follows immediately from the fact that $\lim_{k\to\infty} x_k = x^*$.

To show (viii), note first that the optimal value of $B(x, \mu_k)$ is unique, and is equal to $B(x_k, \mu_k)$. We distinguish two cases, depending on whether or not x^* (the limit point of $\{x_k\}$) is strictly feasible.

If x^* is strictly feasible, the sum of logarithms of the constraints at x_k remains finite as $k \to \infty$. It is easy to see that in this case $\lim_{k\to\infty} B(x_k, \mu_k) = f^*$.

Consider the other possibility, that x^* is not strictly feasible. Since at least one constraint is converging to zero, the barrier term of $B(x_k, \mu_k)$ must be *positive* for all sufficiently large k. Combining this property with (3.14), we have

$$0 < -\sum_{i=1}^{m} \ln c_i(x_k) \le -\sum_{i=1}^{m} \ln c_i(x_{k+1}). \tag{3.20}$$

One implication of this result is that, for sufficiently large k,

$$B(x_k, \mu_k) > f(x_k). \tag{3.21}$$

In addition, the minimizing property of x_{k+1}, the first inequality in (3.20), and the relation $\mu_{k+1} \le \mu_k$ together give:

$$\begin{aligned} f(x_{k+1}) - \mu_{k+1} \sum_{i=1}^{m} \ln c_i(x_{k+1}) &\le f(x_k) - \mu_{k+1} \sum_{i=1}^{m} \ln c_i(x_k) \\ f(x_k) - \mu_{k+1} \sum_{i=1}^{m} \ln c_i(x_k) &\le f(x_k) - \mu_k \sum_{i=1}^{m} \ln c_i(x_k), \end{aligned}$$

which shows that, for sufficiently large k,

$$B(x_{k+1}, \mu_{k+1}) \le B(x_k, \mu_k). \tag{3.22}$$

It follows from (3.21) and (3.22) that, for sufficiently large k,

$$f^* \le \cdots \le B_{k+1} \le B_k, \tag{3.23}$$

where B_k denotes $B(x_k, \mu_k)$. The sequence $\{B_k\}$ of barrier function values is consequently nonincreasing and bounded from below, and must converge monotonically from above to a limit, say B^*, where $B^* \geq f^*$.

Suppose that $B^* > f^*$. In this case, we define δ as the positive number $\frac{1}{2}(B^* - f^*)$. It follows from continuity of f and the relation $f^* < B^*$ that there must be a neighbourhood of x^* in which

$$f(x) \leq B^* - \delta \tag{3.24}$$

for all x in the neighbourhood. Consider a particular strictly feasible point $\bar{x}$ in this neighbourhood. (Such a point must exist because $\mathcal{F}$ is convex and strict($\mathcal{F}$) is nonempty.) Strict feasibility of $\bar{x}$ implies that the quantity $\sum_{i=1}^{m} \ln c_i(\bar{x})$ is finite. Because $\mu_k > 0$ and $\mu_k \to 0$, there must be an integer K such that, for $k \geq K$,

$$-\mu_k \sum_{i=1}^{m} \ln c_i(\bar{x}) < \tfrac{1}{2}\delta. \tag{3.25}$$

Since x_k is a global minimizer of $B(x, \mu_k)$, we know that

$$B(x_k, \mu_k) \leq B(\bar{x}, \mu_k) = f(\bar{x}) - \mu_k \sum_{i=1}^{m} \ln c_i(\bar{x}).$$

If we apply (3.24) and (3.25), the result is

$$B(x_k, \mu_k) < B^* - \delta + \tfrac{1}{2}\delta = B^* - \tfrac{1}{2}\delta,$$

which contradicts the monotonic convergence of $\{B_k\}$ to B^* from above. We conclude that $B^* = f^*$, which gives result (viii). $\square$

The implications of this theorem are remarkably strong. For any convex program with a bounded set of minimizers, the barrier function has a finite unconstrained minimizer for *every* value of the barrier parameter, and every limit point of a minimizing sequence for the barrier function is a constrained minimizer. It is not necessarily true in general, however, that every minimizing sequence converges.

To every convex program, there corresponds a related *dual* convex program. For reasons of space, general results from duality theory will not be considered here, except for the following important result: the objective function value at any unconstrained minimizer x_k of $B(x, \mu_k)$ satisfies the inequality $f(x_k) - f^* \leq m\mu_k$, where m is the number of constraints. We know from results (iv) and (vii) of Theorem 5 that $f^* \leq f(x_k)$. Combining these bounds, we have

$$0 \leq f(x_k) - f^* \leq m\mu_k. \tag{3.26}$$

This somewhat surprising property implies that, when a barrier method is applied to a convex program, the deviation of $f(x_k)$ from optimality is always bounded by $m\mu_k$, independently of the particular problem functions. For comments about duality in linear programming, see Section 5.1.

3.4. Results for general nonlinear programs

Once we move from a convex program to a general nonlinear program, matters become far more complicated. In particular, certain topological assumptions are required to avoid pathological cases. Furthermore, the results apply only in a neighbourhood of a constrained minimizer, and involve convergence of subsequences of *global* minimizers of the barrier function. The general approach in this section follows that in Fiacco and McCormick (1968).

At the most basic level, the nice property given by Theorem 4 that the level sets of the barrier function are bounded if the set of constrained minimizers is bounded does not hold for the nonconvex case. If the feasible region is bounded, the barrier function is obviously bounded below. The following example of Powell (1972), however, shows that difficulties may arise when the feasible region is unbounded:

$$\text{minimize} \quad \frac{-1}{x^2+1} \quad \text{subject to} \quad x \geq 1. \tag{3.27}$$

The objective function is bounded below in the feasible region, and the unique solution is $x^* = 1$. In contrast, the barrier function

$$B(x,\mu) = \frac{-1}{x^2+1} - \mu \ln(x-1)$$

is *unbounded below* in the feasible region, although it has a local minimizer that approaches x^* as $\mu \to 0$.

The major local convergence results will be given in Theorem 7. To build up to the statement of this theorem, several preliminary results are required.

The following lemma, an adaptation of Corollary 8 from Fiacco and McCormick (1968), plays the role of Theorem 4 for the convex case. The general result is that, if a continuous function is unbounded above for all sequences of points in $\text{strict}(\mathcal{F})$ and converging to its boundary, then the function must achieve its minimum value at a *strictly interior* point. The obvious application of Lemma 1 is when $B(x,\mu)$ plays the role of φ.

Lemma 1 *Given a set of m smooth constraint functions $\{c_i(x)\}$, $i = 1, \ldots, m$, let* $\text{strict}(\mathcal{F})$ *denote the set defined by (2.3). Let S be a compact set, and assume that the set* $\text{strict}(\mathcal{F}) \cap S$ *is nonempty. Consider any convergent sequence $\{y_k\} \in \text{strict}(\mathcal{F}) \cap S$ whose limit point $\bar{y}$ lies on the boundary of* $\text{strict}(\mathcal{F})$*, i.e. such that*

$$\lim_{k\to\infty} y_k = \bar{y}, \quad \textit{where} \quad \bar{y} \in \text{bnd}(\text{strict}(\mathcal{F})) \cap S. \tag{3.28}$$

Suppose that φ is a continuous function on $\text{strict}(\mathcal{F}) \cap S$ *with the property that $\varphi(y_k)$ is unbounded above as $k \to \infty$ for every sequence $\{y_k\}$ satisfying (3.28). Then the global minimum value of φ in* $\text{strict}(\mathcal{F}) \cap S$*, denoted by φ^*, is finite, and is achieved at some point x^* in* $\text{strict}(\mathcal{F}) \cap S$*:*

$$\min\{\varphi(x) \mid x \in \text{strict}(\mathcal{F}) \cap S\} = \varphi(x^*) = \varphi^*.$$

Proof. Given any point $\hat{x}$ in $\text{strict}(\mathcal{F}) \cap S$, define the associated level set $\hat{W}$ as

$$\hat{W} = \{x \in \text{strict}(\mathcal{F}) \cap S \mid \varphi(x) \leq \hat{\varphi}\},$$

where $\hat{\varphi} = \varphi(\hat{x})$. Because S is compact, $\hat{W}$ is bounded. Compactness of $\hat{W}$ will follow if we show that $\hat{W}$ is closed, i.e., contains all its accumulation points.

Let R denote the closed set

$$R = \text{strict}(\mathcal{F}) \cup \text{bnd}(\text{strict}(\mathcal{F})).$$

Because S is compact and R is closed, the set $R \cap S$ is compact. Consider any convergent sequence $\{x_k\}$ such that $x_k \in \hat{W}$ for all k, with limit point $\bar{x}$. Since $x_k \in \text{strict}(\mathcal{F}) \cap S$, $\bar{x}$ must lie in $R \cap S$. Hence $\bar{x}$ must lie in either $\text{strict}(\mathcal{F}) \cap S$ or $\text{bnd}(\text{strict}(\mathcal{F})) \cap S$.

If $\bar{x}$ is in $\text{bnd}(\text{strict}(\mathcal{F})) \cap S$, then $\{x_k\}$ is a sequence satisfying (3.28), which means that $\varphi(x_k) \to \infty$. Since $\hat{\varphi}$ is an upper bound on the value of φ at any point in $\hat{W}$, we conclude that $x_k \notin \hat{W}$ for sufficiently large k, which is a contradiction. Any limit point $\bar{x}$ of a sequence in $\hat{W}$ therefore cannot be in $\text{bnd}(\text{strict}(\mathcal{F})) \cap S$, and must lie in $\text{strict}(\mathcal{F}) \cap S$.

Because x_k is in $\hat{W}$, the relation $\varphi(x_k) \leq \hat{\varphi}$ holds for all k. Continuity of φ in $\text{strict}(\mathcal{F}) \cap S$ then implies that the limit point $\bar{x}$ satisfies $\varphi(\bar{x}) \leq \hat{\varphi}$, so that $\bar{x}$ possesses both properties required for membership in $\hat{W}$. Since $\{x_k\}$ is an arbitrary convergent sequence in $\hat{W}$, it follows that $\hat{W}$ contains all its accumulation points and is closed.

We know already that $\hat{W}$ is bounded, so that $\hat{W}$ is compact. Because φ is continuous in the compact set $\hat{W}$, it attains its global minimum in $\hat{W}$ at some point x^*. By definition of $\hat{W}$, the value of φ at any point in strict$(\mathcal{F}) \cap S$ but outside $\hat{W}$ must be strictly larger than the smallest value of φ at any point in $\hat{W}$. Hence x^* is the global minimizer of φ in the entire set strict$(\mathcal{F}) \cap S$, which is the desired conclusion. $\square$

A property needed for local convergence is that a particular subset of local constrained minimizers is 'isolated' within the full set of local constrained minimizers. Such a definition is unnecessary for the convex case, since the set of minimizers is convex.

Definition 11 *(Isolated subset.) Let $\mathcal{M}$ and $\mathcal{M}^*$ be sets in $\mathcal{R}^n$ such that $\mathcal{M}^* \subseteq \mathcal{M}$. The set $\mathcal{M}^*$ is called an* isolated *subset of $\mathcal{M}$ if there exists a closed set E such that $\mathcal{M}^* \subset \mathrm{int}(E)$ and $E \cap \mathcal{M} = \mathcal{M}^*$.*

Broadly speaking, $\mathcal{M}^*$ is 'separated' by E from any other points of $\mathcal{M}$. The definition is satisfied if $\mathcal{M}^* = \mathcal{M}$, or if $\mathcal{M}^*$ is an isolated point in $\mathcal{M}$.

The next theorem (a version of Theorem 7 of Fiacco and McCormick (1968)) shows that, if a set of constrained minimizing points is *compact* and *isolated*, there is a compact set S, strictly enclosing the set of minimizers, within which the minimizers are *global*. The role of the set S is critical: if we can restrict attention to points in S, the value of f at any minimizing point in S is a strict lower bound on the value of f at any other feasible (nonoptimal) point in S. For the convex case, a suitable set S is provided 'automatically' by the level set for f at any strictly feasible point; see (3.15).

Theorem 6 *(Existence of compact enclosing set.) Consider the problem of minimizing $f(x)$ subject to $c_i(x) \ge 0$, $i = 1, \ldots, m$. Let $\mathcal{M}$ denote the set of all local constrained minimizers with objective function value f^*, and assume that $\mathcal{M}$ is nonempty. Assume further that the set $\mathcal{M}^* \subseteq \mathcal{M}$ is a nonempty compact isolated subset of $\mathcal{M}$. Then there exists a compact set S such that $\mathcal{M}^*$ lies in $\mathrm{int}(S) \cap \mathcal{F}$, with the property that for any feasible point y in S but not in $\mathcal{M}^*$, $f(y) > f^*$. The points in $\mathcal{M}^*$ are thus global minimizers of the nonlinear program for $x \in S \cap \mathcal{F}$.*

Proof. Applying Definition 11, the assumption that $\mathcal{M}^*$ is an isolated subset of $\mathcal{M}$ implies existence of a closed set E strictly containing $\mathcal{M}^*$ such that $\mathrm{int}(E) \cap \mathcal{M} = \mathcal{M}^*$.

The assumption that $\mathcal{M}^*$ is compact means that we can construct a sequence of strictly nested compact sets $\{S_j\}$ converging to $\mathcal{M}^*$, each strictly containing $\mathcal{M}^*$, namely such that $\mathcal{M}^* \subset \mathrm{int}(S_j) \subset \mathrm{int}(E)$,

$$S_{j+1} \subset S_j, \quad \text{and} \quad \lim_{j \to \infty} S_j = \mathcal{M}^*. \tag{3.29}$$

The proof will show by contradiction that the desired compact set S may be taken as S_j for some finite j.

If this is impossible, then for every j we can find a feasible point x_j with the following properties:

$$x_j \in \mathcal{F} \cap S_j, \quad x_j \notin \mathcal{M}^* \ \text{ and } \ f(x_j) \le f^*. \tag{3.30}$$

Consider this hypothetical sequence $\{x_j\}$. The nested structure of $\{S_j\}$ means that $\{x_j\}$ is bounded and hence has at least one limit point, say $\bar{x}$. It follows from (3.29) and the fact that $\mathcal{M}^*$ is closed that $\bar{x} \in \mathcal{M}^*$, so that $f(\bar{x}) = f^*$ and $\bar{x}$ is a constrained minimizer. Let $\{y_k\}$ denote a subsequence of $\{x_j\}$ converging to $\bar{x}$, where $y_k = x_{j_k}$.

If $f(y_k)$ is strictly less than $f(\bar{x})$ for an infinite number of indices k, then every neighbourhood of $\bar{x}$ contains feasible points with a strictly smaller value of f. This means that $\bar{x}$ cannot be a constrained minimizer (see Definition 4) and gives a contradiction.

We conclude that eventually, for some $k = \bar{k}$, any point in S_{j_k} that qualifies as $y_{\bar{k}}$ (i.e. any feasible point in S_{j_k} for which $f(y_{\bar{k}}) \le f^*$) must have an objective function value *equal to* f^*. This result implies that f^* is the *smallest* value of f achieved at feasible points in $S_{j_{\bar{k}}}$.

Let $\bar{S}$ denote $S_{j_{\bar{k}}}$. Since all subsequent sets S_{j_k} for $k \ge \bar{k}$ are subsets of $\bar{S}$, as is the minimizing set $\mathcal{M}^*$, it follows that f^* is the smallest value of f assumed at any feasible point in S_{j_k} for all $k \ge \bar{k}$.

The strictly nested property of the sets $\{S_j\}$ means that, for sufficiently large j, *any* point x_j satisfying (3.30), not necessarily a member of the subsequence converging to $\bar{x}$, must lie in the interior of the compact set $\bar{S}$. Because x_j is feasible and satisfies $f(x_j) \le f^*$, and the smallest value of f for any feasible point in $\bar{S}$ is equal to f^*, it must be true that $f(x_j) = f^*$. Since x_j lies in the interior of the compact set $\bar{S}$ and $f(x_j) = f^*$, x_j satisfies Definition 4 of a local constrained minimizer with function value f^*, and hence x_j is in $\mathcal{M}$, the set of such minimizers. However, x_j is by definition in S_j, which is contained in the interior of E. Because $\mathrm{int}(E) \cap \mathcal{M} = \mathcal{M}^*$, x_j must be in $\mathcal{M}^*$. It follows that, for sufficiently large j, no points in S_j satisfy (3.30).

We have constructed a compact set $\bar{S}$ that strictly contains $\mathcal{M}^*$. Further, $\bar{S}$ contains no feasible points with objective function values less than f^*, and every feasible point in $\bar{S}$ with objective function value equal to f^* lies in $\mathcal{M}^*$. It follows that any feasible point y in $\bar{S}$ but not in $\mathcal{M}^*$ must satisfy $f(y) > f^*$. The set $\bar{S}$ thus satisfies all the criteria specified for S, and the theorem is proved. $\square$

We now give a fundamental theorem, analogous to Theorems 8 and 10 in Fiacco and McCormick (1968), about local convergence of logarithmic barrier methods. This theorem assumes two important properties: (a) a compactness requirement that the relevant set $\mathcal{M}^*$ of local minimizers is nonempty and compact (in the simplest case, $\mathcal{M}^*$ is a single point); and (b) a topological restriction that at least one of the points in $\mathcal{M}^*$ lies in the closure of $\mathrm{strict}(\mathcal{F})$, i.e. is either strictly feasible or else an accumulation point of $\mathrm{strict}(\mathcal{F})$.

Assumption (b) disallows minimizers that occur at isolated feasible points (points in a neighbourhood containing no other feasible points). For example, consider the constraints $x \ge 1$ and $x^2 - 5x + 4 \ge 0$. The function $x^2 - 5x + 4$ is nonnegative if $x \le 1$ and if $x \ge 4$, so that the feasible points lie in two separated regions. The constraint $x \ge 1$ eliminates all of the region $\{x \le 1\}$ except the single point $x = 1$. The feasible region for *both* constraints therefore consists of the isolated point $\{x = 1\}$ and the set of points $\{x \ge 4\}$. Hence $\mathrm{strict}(\mathcal{F})$ is the set $\{x > 4\}$, and the point $x = 1$ does not lie in the closure of $\mathrm{strict}(\mathcal{F})$.

Barrier methods can be viewed as finding the infimum of f subject to $c(x) > 0$, and consequently cannot converge to minimizers occurring at isolated points. Isolated minimizers do not arise in the convex case because a convex set with a nonempty interior cannot contain an isolated point.

Theorem 7 *(Local convergence for barrier methods.) Consider the problem of minimizing $f(x)$ subject to $c_i(x) \ge 0$, $i = 1, \ldots, m$. Let $\mathcal{F}$ denote the feasible region, and let $\mathcal{M}$ denote the set of minimizers corresponding to the objective function value f^*. Let $\{\mu_k\}$ be a decreasing sequence of positive barrier parameters such that $\lim_{k\to\infty} \mu_k = 0$. Assume that*

(a) there exists a nonempty compact set $\mathcal{M}^*$ of local minimizers that is an isolated subset of $\mathcal{M}$;

(b) at least one point in $\mathcal{M}^*$ is in the closure of $\mathrm{strict}(\mathcal{F})$.

Then the following results hold:

(i) there exists a compact set S strictly containing $\mathcal{M}^*$ such that for any feasible point $\bar{x}$ in S but not in $\mathcal{M}^*$, $f(\bar{x}) > f^*$;

(ii) for all sufficiently small μ_k, $B(x, \mu_k)$ has at least one unconstrained minimizer in $\text{strict}(\mathcal{F}) \cap \text{int}(S)$, and any sequence of global unconstrained minimizers of $B(x, \mu_k)$ in $\text{strict}(\mathcal{F}) \cap \text{int}(S)$ has at least one convergent subsequence;

(iii) let $\{x_k\}$ denote any convergent subsequence of global unconstrained minimizers of $B(x, \mu_k)$ in $\text{strict}(\mathcal{F}) \cap \text{int}(S)$; then the limit point of $\{x_k\}$ is in $\mathcal{M}^*$;

(iv) $\lim_{k\to\infty} f(x_k) = f^* = \lim_{k\to\infty} B(x_k, \mu_k)$.

Proof. Result (i) follows immediately from Theorem 6, which implies the existence of a strictly enclosing compact set S within which all points in $\mathcal{M}^*$ are global constrained minimizers.

Consider the behaviour of the barrier function $B(x, \mu_k)$ in the bounded set $\text{strict}(\mathcal{F}) \cap S$. Continuity of f and $\{c_i\}$ in $\mathcal{F}$ implies that $B(x, \mu_k)$ is continuous in $\text{strict}(\mathcal{F}) \cap S$. The barrier function possesses the properties of φ in Lemma 1, which then implies that $B(x, \mu_k)$ achieves a finite global minimum value at some point in $\text{strict}(\mathcal{F}) \cap S$. (This result is close but not equivalent to (ii), which states that the minimizing point lies in $\text{int}(S)$.) Let y_k be any point in $\text{strict}(\mathcal{F}) \cap S$ for which the minimum value is achieved.

The sequence $\{y_k\}$ is bounded and hence has at least one limit point. Let $\hat{x}$ denote a limit point of $\{y_k\}$. Because y_k is strictly feasible for all k and the set S is compact, it follows that $\hat{x} \in \mathcal{F} \cap S$, so that $\hat{x}$ is feasible.

We wish to show that $\hat{x}$ lies in the set $\mathcal{M}^*$ of constrained minimizers, with $f(\hat{x}) = f^*$. The result will be proved by contradiction, and we accordingly assume the contrary, that $\hat{x} \notin \mathcal{M}^*$.

Since $\hat{x}$ is feasible and in S, result (i) implies that $f(\hat{x}) > f^*$. We next prove that this inequality implies the existence of a strictly feasible point x_{int} in S such that

$$f(\hat{x}) > f(x_{\text{int}}). \tag{3.31}$$

The point x_{int} can be found as follows. We know from assumption (b) that at least one point in $\mathcal{M}^*$ is in the closure of $\text{strict}(\mathcal{F})$. Let x^* denote such a point, which must either lie in $\text{strict}(\mathcal{F})$ or else be an accumulation point of $\text{strict}(\mathcal{F})$. Because $\mathcal{M}^*$ is contained in $\text{int}(S)$, x^* is also in the interior of S.

If x^* itself is strictly feasible, x_{int} may be taken as x^*. If x^* is not strictly feasible, x^* is an accumulation point of $\text{strict}(\mathcal{F})$, which means that every neighbourhood of x^* contains strictly feasible points. Further, every neighbourhood of x^* contains points in S. We know that: f is continuous; $\hat{x}$ is feasible and lies in S; $f(\hat{x}) > f(x^*)$; and x^* is a global constrained minimizer of f for all feasible points in S. Hence there must be a strictly feasible point x_{int} in a neighbourhood of x^* for which $f(x_{\text{int}}) < f(\hat{x})$.

Let $\{x_k\}$ denote a convergent subsequence of $\{y_k\}$ with limit $\hat{x}$. The relation $f(\hat{x}) > f(x_{\text{int}})$ then implies that, for sufficiently large k,

$$f(x_k) > f(x_{\text{int}}). \tag{3.32}$$

Since x_{int} is in $\text{strict}(\mathcal{F}) \cap S$, our definition of x_k as a global minimizer of $B(x, \mu_k)$ in $\text{strict}(\mathcal{F}) \cap S$ implies the inequality

$$f(x_k) - \mu_k \sum_{i=1}^{m} \ln c_i(x_k) \le f(x_{\text{int}}) - \mu_k \sum_{i=1}^{m} \ln c_i(x_{\text{int}}). \tag{3.33}$$

Strict feasibility of x_{int} means that the barrier term involving x_{int} in (3.33) is *finite*, and

$$\lim_{k\to\infty} B(x_{\text{int}}, \mu_k) = f(x_{\text{int}}).$$

Suppose that the limit point $\hat{x}$ of $\{x_k\}$ is also strictly feasible, namely $\hat{x} \in \text{strict}(\mathcal{F}) \cap S$. Then the barrier term involving x_k is finite as $k \to \infty$ and

$$\lim_{k\to\infty} B(x_k, \mu_k) = f(\hat{x}).$$

Letting $k \to \infty$ in (3.33), we obtain the inequality $f(x_k) \le f(x_{\text{int}})$, which contradicts the relation $f(x_k) > f(x_{\text{int}})$ of (3.32).

Suppose on the other hand that $\hat{x}$ is not strictly feasible. Adding a barrier term to both sides of the inequality $f(x_k) > f(x_{\text{int}})$ gives

$$f(x_{\text{int}}) - \mu_k \sum_{i=1}^{m} \ln c_i(x_{\text{int}}) < f(x_k) - \mu_k \sum_{k=1}^{m} \ln c_i(x_{\text{int}}).$$

Combining this inequality with (3.33), rearranging and then dividing by μ_k, we obtain

$$f(x_k) - \mu_k \sum_{k=1}^{m} \ln c_i(x_k) < f(x_k) - \mu_k \sum_{k=1}^{m} \ln c_i(x_{\text{int}}).$$

Cancelling $f(x_k)$ from both sides, the result is

$$-\sum_{k=1}^{m} \ln c_i(x_k) < -\sum_{k=1}^{m} \ln c_i(x_{\text{int}}). \tag{3.34}$$

As before, strict feasibility of x_{int} guarantees that the sum on the right-hand side is fixed and finite. However, since $\hat{x}$ is not strictly feasible, $-\ln c_i(x_k)$ approaches infinity for at least one i. The left-hand side of (3.34) is therefore unbounded above, which again gives a contradiction.

In either case, we have shown that $f(\hat{x}) = f^*$ and hence that $\hat{x} \in \mathcal{M}^*$. Since $\hat{x}$ was taken as any limit point of $\{y_k\}$, we conclude that every limit point of a convergent subsequence of global barrier minimizers lying in $\text{strict}(\mathcal{F}) \cap S$ must be a constrained minimizer with objective value f^*.

Result (ii) is proved by noting that the relation $\hat{x} \in \mathcal{M}^*$ means that $\hat{x} \in \text{int}(S)$. Since $\hat{x}$ is the limit point of $\{x_k\}$, it must hold that x_k is also in $\text{int}(S)$ for sufficiently large k. By definition, x_k is strictly feasible. Hence the global minimum of $B(x, \mu_k)$ in $\text{strict}(\mathcal{F}) \cap S$ is achieved at some point x_k lying strictly inside both $\mathcal{F}$ and S. Applying Definition 7, the global minimizer x_k of $B(x, \mu_k)$ in S is an *unconstrained* minimizer. Results (ii) and (iii) are thus proved.

The first relation in (iv), that $\lim_{k\to\infty} f(x_k) = f^*$, follows because $f(\hat{x}) = f^*$. The second, $\lim_{k\to\infty} B(x_k, \mu_k) = f^*$, follows from the same arguments used in proving (iv) and (viii) of Theorem 5, with the additional restriction here that all points must lie in S. □

At this point we should emphasize what has *not* been proved. Even within the set S, a general global minimizing sequence $\{x_k\}$ of the barrier function is not guaranteed to converge. The properties of local minimizing sequences are even less secure. In particular, it is not true that every limit point of a local minimizing sequence is a constrained minimizer.

For example, consider the problem

$$\text{minimize } x \quad \text{subject to} \quad x^2 \ge 0, \;\; x \ge -\gamma,$$

where $\gamma > 0$ (Moré and Wright, 1990); a similar example is given in Fiacco (1979). The unique solution is obviously the point $x^* = -\gamma$. The barrier function $B(x, \mu)$ has two feasible minimizers:

$$x(\mu) = \frac{3\mu - \gamma \pm \left((3\mu-\gamma)^2 + 8\gamma\mu\right)^{1/2}}{2}.$$

For $\mu \to 0$, the *global* minimizing sequence corresponds to the negative square root and converges to $-\gamma$, the unique solution of the constrained problem. However, the *nonglobal* minimizing sequence of $B(x, \mu)$, corresponding to the positive square root, converges to the origin, which is *not* a constrained minimizer.

Despite these cautions, the bright side is that barrier methods will converge to the solutions of constrained problems for which the usual sufficient conditions do not hold. Barrier methods can converge, for example, when the constrained minimizer is not locally unique. Barrier methods can succeed even when a local constrained minimizer does not satisfy a constraint qualification.

3.5. The barrier trajectory

In this section, we describe conditions under which a sequence $x(\mu)$ of barrier minimizers not only converges to x^*, but also defines a smooth path (the 'barrier trajectory') that is nontangential to the active constraint gradients.

Discussions of the logarithmic barrier function involve special diagonal matrices related to vectors, for which the following notation has become popular. When a lower-case letter refers to a vector, its upper-case version means the diagonal matrix of comparable dimension whose (i, i) element is the ith component of the vector. For example, C denotes the $m \times m$ diagonal matrix of constraint values $\{c_i\}$:

$$C = \operatorname{diag}(c_i) = \begin{pmatrix} c_1 & & & \\ & c_2 & & \\ & & \ddots & \\ & & & c_m \end{pmatrix},$$

and C^{-1} is the diagonal matrix whose ith element is $1/c_i$. Using this convention, we have the general relation $Ce = c$, where e denotes the vector of appropriate dimension whose components are all equal to one, $e = (1, 1, \ldots, 1)^T$.

The gradient of $B(x, \mu)$ with respect to x is

$$\nabla B(x, \mu) = g - \sum_{i=1}^{m} \frac{\mu}{c_i}\, a_i = g - \mu A^T C^{-1} e, \tag{3.35}$$

where all functions are evaluated at x, a_i is the gradient of $c_i(x)$ and A is the Jacobian of $c(x)$. (Recall that the constraint gradients are the transposed rows of A and hence the *columns* of A^T.) The Hessian of $B(x, \mu)$ is

$$\begin{aligned} \nabla^2 B(x, \mu) &= H - \sum_{i=1}^{m} \frac{\mu}{c_i}\, H_i + \sum_{i=1}^{m} \frac{\mu}{c_i^2}\, a_i a_i^T \\ &= H - \sum_{i=1}^{m} \frac{\mu}{c_i}\, H_i + \mu A^T C^{-2} A. \end{aligned}$$

The point $x(\mu)$ is an unconstrained minimizer of $B(x, \mu)$ only if the gradient vanishes at $x(\mu)$. Substituting from (3.35), the following relation must hold at $x(\mu)$:

$$g(\mu) = \sum_{i=1}^{m} a_i(\mu)\, \frac{\mu}{c_i(\mu)}, \tag{3.36}$$

where the argument μ denotes evaluation at $x(\mu)$. Since $\mu > 0$ and $c_i(\mu) > 0$, it follows that the gradient of f at $x(\mu)$ is a *positive* linear combination of the gradients of *all* the constraints.

The first-order optimality conditions (2.13b–d) for nonlinear constraints are

$$g(x^*) = A(x^*)^T\lambda^* = \sum_{i=1}^{m} a_i(x^*)\lambda_i^*, \tag{3.37}$$

where $\lambda_i^* \geq 0$ and $\lambda_i^* c_i(x^*) = 0$, $i = 1, \ldots, m$. At x^*, the gradient of f is thus a nonnegative linear combination of all the constraint gradients, where inactive constraints have zero multipliers.

The similar forms of (3.36) and (3.37) reveal that the ith coefficient $\mu/c_i(\mu)$ in the linear combination (3.36) is directly analogous to the ith Lagrange multiplier λ_i^*. When standard sufficient optimality conditions hold at x^* and the gradients of the active constraints are linearly independent, the multiplier estimates $\mu/c_i(\mu)$ do indeed converge to λ_i^*. In fact, under these conditions a differentiable curve $x(\mu)$ of barrier minimizers, parametrized by μ, exists near $\mu = 0$ and converges to x^*. This curve of minimizers is called the *barrier trajectory*; in linear programming, it is usually known as the *central path*. Its existence and properties define the broad class of 'path-following' algorithms that attempt to follow the trajectory to the solution; see Sections 5.2 and 6.2.

The results of the following theorem are essentially those of Theorem 12 of Fiacco and McCormick (1968).

Theorem 8 *Consider the problem of minimizing $f(x)$ subject to $c_i(x) \geq 0$, $i = 1, \ldots, m$, where $m \geq 1$. Let $\mathcal{F}$ denote the feasible region, and assume that* strict($\mathcal{F}$) *is nonempty. Assume further that x^* is a local constrained minimizer at which*

(a) $g(x^*) = A(x^*)^T\lambda^*$, with $\lambda_i^* c_i(x^*) = 0$;

(b) $\lambda_i^* > 0$ if $c_i(x^*) = 0$;

(c) there exists $\alpha > 0$ such that $p^T W(x^*, \lambda^*)p \geq \alpha\|p\|^2$ for all p satisfying $\hat{A}p = 0$, where $\hat{A}$ denotes the Jacobian of the active constraints at x^* and W is the Hessian of the Lagrangian function (see (2.12));

(d) the gradients of the active constraints at x^* are linearly independent.

Consider a logarithmic barrier method in which $B(x, \mu_k)$ is minimized for a sequence of positive values $\{\mu_k\}$ converging monotonically to zero as $k \to \infty$. Then

(i) there is at least one subsequence of unconstrained minimizers of the barrier function $B(x, \mu_k)$ converging to x^*;

(ii) For such a convergent subsequence $\{x_k\}$,

$$\lim_{k\to\infty} \mu_k/c_i^k = \lambda_i^*, \quad \text{where } c_i^k \text{ denotes } c_i(x_k);$$

(iii) for sufficiently large k, the Hessian matrix $\nabla^2 B(x_k, \mu_k)$ is positive definite;

(iv) a unique, continuously differentiable vector function $x(\mu)$ of unconstrained minimizers of $B(x, \mu)$ exists in a neighbourhood of $\mu = 0$;

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

Proof. The properties assumed about x^* ensure that it is an isolated constrained minimizer. Two implications follow from the linear independence of the active constraint gradients: the Lagrange multipliers λ^* are unique; and every neighbourhood of x^* contains points in strict($\mathcal{F}$), so that x^* is in the closure of the interior of the feasible region. Theorem 7 consequently applies to x^*, and implies that there is at least one subsequence of unconstrained minimizers of $B(x, \mu_k)$ converging to x^*. This proves (i).

Let $\{x_k\}$ denote such a convergent sequence, with redefinition of k as necessary, so that

$$\lim_{k\to\infty} x_k = x^*. \tag{3.38}$$

As convenient, we denote quantities associated with x_k by a subscript or superscript k; the subscript i always denotes the ith component of a vector.

For sufficiently large k, x_k is an unconstrained minimizer of $B(x,\mu_k)$, which means that the gradient (3.35) of the barrier function vanishes at x_k:

$$g^k = \sum_{i=1}^m a_i^k \lambda_i^k, \quad \text{where} \quad \lambda_i^k = \frac{\mu_k}{c_i^k}. \tag{3.39}$$

The quantity λ_i^k is strictly positive for any $\mu_k > 0$.

Suppose that constraint i is *inactive* at x^*. Then, from (3.38),

$$\lim_{k\to\infty} c_i^k = c_i(x^*) > 0, \quad \text{and hence} \quad \lim_{k\to\infty} \lambda_i^k = \lambda_i^* = 0. \tag{3.40}$$

If no constraints are active, we have verified (ii).

Otherwise, let $\mathcal{A}$ denote the set of indices of constraints active at x^*, so that $c_i(x^*) = 0$ for $i \in \mathcal{A}$. Let the positive numbers s_k and v_i^k be defined as

$$s_k = \sum_{i=1}^m \lambda_i^k \quad \text{and} \quad v_i^k = \frac{\lambda_i^k}{s_k}.$$

Note that $v_i^k > 0$ and $\sum_{i=1}^m v_i^k = 1$, so that $v_i^k \le 1$. Since $s_k > 0$ and (3.39) holds at x_k, we have

$$\frac{1}{s_k} g^k - \sum_{i=1}^m a_i^k v_i^k = 0. \tag{3.41}$$

As $k \to \infty$, the sequence $\{v_i^k\}$ is bounded for $i = 1, \ldots, m$, and accordingly contains a convergent subsequence.

The value of $\liminf_{k\to\infty} s_k$, denoted by $\hat{s}$, must be finite. If not, consider (3.41) as $k \to \infty$. Because of (3.38), a_i^k converges to a_i^*, where the superscript $*$ denotes evaluation at x^*. The following relation must hold for any set $\{\hat{v}_i\}$ of limit points of $\{v_i^k\}$:

$$\sum_{i=1}^m a_i^* \hat{v}_i = 0, \quad \text{where} \quad \hat{v}_i \ge 0 \quad \text{and} \quad \sum_{i=1}^m \hat{v}_i = 1. \tag{3.42}$$

Because $\hat{v}_i = 0$ if constraint i is inactive at x^*, relation (3.42) states that a nontrivial linear combination of the active constraint gradients at x^* is zero, which contradicts our assumption of their linear independence.

Finiteness of $\hat{s}$ implies that each component λ_i^k is bounded for all k, and consequently the sequence $\{\lambda_i^k\}$ has at least one accumulation point, say $\bar{\lambda}_i$. It follows from (3.39) and (3.40) that

$$g^* = \hat{A}^T \bar{\lambda}.$$

Because the rows of $\hat{A}$ are linearly independent, the values satisfying this equation are unique, and we conclude that $\bar{\lambda}_i = \lambda_i^*$ for $i \in \mathcal{A}$, which completes the proof of (ii).

We now wish to demonstrate positive-definiteness of the barrier Hessian at x_k. This property will be verified using the asymptotic structure of the Hessian of the barrier function, which approaches the sum of the Hessian of the Lagrangian function and a 'large' matrix in the range space of the active constraint gradients.

Consider the ratio $\mu_k/(c_i^k)^2$, which we denote by d_i^k:

$$d_i^k = \frac{\mu_k}{(c_i^k)^2} = \frac{\lambda_i^k}{c_i^k}. \tag{3.43}$$

When constraint i is active at x^*, i.e. $i \in \mathcal{A}$, result (ii) and assumption (b) of strict complementarity imply that λ_i^k converges to a strictly positive constant. Since c_i^k converges to zero for $i \in \mathcal{A}$, the final ratio in (3.43) is clearly *unbounded*, and

$$\liminf_{k\to\infty} d_i^k = \liminf_{k\to\infty} \frac{\mu_k}{(c_i^k)^2} = \infty \quad \text{for } i \in \mathcal{A}. \tag{3.44}$$

Recall that the Hessian of the barrier function is given by

$$\nabla^2 B(x,\mu) = H - \sum_{i=1}^m \frac{\mu}{c_i} H_i + \sum_{i=1}^m \frac{\mu}{c_i^2} a_i a_i^T.$$

Let H_B^k denote $\nabla^2 B(x_k, \mu_k)$. The limiting properties of this matrix are revealed by expressing it in the following form:

$$H_B^k = W^* + M^* + M_1^k + M_2^k + M_3^k.$$

The first two matrices on the right-hand side depend on x^* and a bounded positive constant γ:

$$\begin{aligned} W^* &= H^* - \sum_{i=1}^m \lambda_i^* H_i^* \\ M^* &= \gamma \sum_{i\in\mathcal{A}} a_i^*(a_i^*)^T = \gamma \hat{A}^T \hat{A}. \end{aligned}$$

The remaining three matrices are expressed as perturbations involving x_k, x^* and γ:

$$\begin{aligned} M_1^k &= H^k - H^* - \Big(\sum_{i=1}^m \lambda_i^k\, H_i^k - \sum_{i=1}^m \lambda_i^* H_i^*\Big) \\ M_2^k &= \gamma \sum_{i\in\mathcal{A}} \Big(a_i^k(a_i^k)^T - a_i^*(a_i^*)^T\Big) \\ M_3^k &= \sum_{i\in\mathcal{A}} (d_i^k - \gamma) a_i^k (a_i^k)^T + \sum_{i\notin\mathcal{A}} d_i^k a_i^k (a_i^k)^T. \end{aligned}$$

The matrix W^* is the Hessian of the Lagrangian function at x^*. For sufficiently large k, the matrices M_1^k and M_2^k can be made arbitrarily small in norm; this statement follows from continuity of the problem functions, convergence of x_k to x^*, convergence of λ_k to λ^*, and boundedness of γ. Because d_i^k is unbounded above for $i \in \mathcal{A}$ (see (3.44)), the quantity $(d_i^k - \gamma)$ is positive for sufficiently large k; hence the matrix M_3^k is the sum of two positive semi-definite matrices and must itself be positive semi-definite.

Positive-definiteness of H_B^k will follow if the matrix $W^* + \gamma \hat{A}^T \hat{A}$ is guaranteed to be positive definite for some constant γ. This property is shown by examining the effect of the matrix in two orthogonal subspaces: the range space of $\hat{A}^T$ and the null space of $\hat{A}$.

It is well known that any n-vector p may be uniquely expressed as the sum of two orthogonal components,

$$p = p_R + p_N,$$

where p_R lies in the range space of $\hat{A}^T$ and p_N lies in the null space of $\hat{A}$. Using this form, the product $p^T(W^* + \gamma\hat{A}^T\hat{A})p$ can be written as

$$p_N^T W^* p_N + 2p_N^T W^* p_R + p_R^T W^* p_R + \gamma p_R^T \hat{A}^T \hat{A} p_R. \tag{3.45}$$

To ensure positive-definiteness, this quantity must be bounded below by a positive number when $p \neq 0$. To develop the bound, we use a relation that holds for any matrix C and vectors x and y:

$$x^T C y \geq -\|C\|\,\|x\|\,\|y\|.$$

Assumption (c) guarantees the existence of $\alpha > 0$ (the smallest eigenvalue of the reduced Hessian of the Lagrangian; see (2.16)) such that

$$p_N^T W^* p_N \geq \alpha\|p_N\|^2.$$

By definition, p_R is in the range of $\hat{A}^T$. Hence, if $p_R \neq 0$, it holds that

$$\hat{A}p_R \neq 0 \quad \text{and} \quad p_R^T \hat{A}^T \hat{A} p_R > \beta\|p_R\|^2$$

for some positive β (the square of the smallest nonzero singular value of $\hat{A}$). Let ω denote $\|W^*\|$.

Applying these inequalities to (3.45), we obtain

$$p^T(W^* + \gamma\hat{A}^T\hat{A})p \geq \alpha\|p_N\|^2 + \gamma\beta\|p_R\|^2 - 2\omega\|p_R\|\,\|p_N\| - \omega\|p_R\|^2.$$

If $\|p_R\| = 0$, so that p lies entirely in the null space of $\hat{A}$, the expression on the right-hand side is simply $\alpha\|p_N\|^2$, which must be positive. Otherwise, if $\|p_R\| \neq 0$, the right-hand side is guaranteed to be positive if γ is bounded below as follows:

$$\gamma > \frac{\omega^2 + \alpha\omega}{\alpha\beta}. \tag{3.46}$$

We have shown that the Hessian of the barrier function at x_k must be positive definite for sufficiently large k, which is result (iii). The point x_k is consequently an *isolated* unconstrained minimizer of $B(x, \mu_k)$ (see (2.5)), and is locally unique.

To verify the existence of a unique, differentiable function $x(\mu)$ in a neighbourhood of $x(\mu_k)$, we apply the *implicit function theorem* (see Ortega and Rheinboldt (1970)) to the $n+1$ variables (x, μ). At (x_k, μ_k), we know from (3.36) that the following system of nonlinear equations has a solution:

$$\Phi(x, \mu) = g(x) - \mu \sum_{i=1}^{m} \frac{1}{c_i(x)}\, a_i(x).$$

The Jacobian of Φ with respect to x is the barrier Hessian H_B^k, which was just shown to be positive definite at (x_k, μ_k). The implicit function theorem then implies that there is a locally unique, differentiable function $x(\mu)$ passing through $x(\mu_k)$ such that $\Phi(x, \mu) = 0$ for all μ in a neighbourhood of μ_k.

Using continuation arguments, it is straightforward to show that the function $x(\mu)$ exists for all $0 < \mu \leq \mu_k$ for sufficiently large k, which gives result (iv).

The final result is immediate from the local uniqueness of $x(\mu)$ and result (i). $\square$

We have now verified the existence of both the barrier trajectory $x(\mu)$ and the associated multiplier estimate $\lambda(\mu)$. A remaining question involves existence and differentiability of the trajectory at x^* itself. For sufficiently small μ, the following $n + m$ equations are satisfied identically at every pair $(x(\mu), \lambda(\mu))$ on the trajectory:

$$g(x) - A(x)^T\lambda = 0; \tag{3.47a}$$
$$\lambda_i c_i(x) = \mu, \quad i = 1, \ldots, m. \tag{3.47b}$$

If we treat the multipliers $\lambda(\mu)$ as *separate variables*, (3.47) can be viewed as a system of nonlinear equations in the $n+m$ variables (x,λ). The Jacobian matrix of this system is given by

$$J(\mu) = \begin{pmatrix} H(\mu) - \sum \lambda_i(\mu) H_i(\mu) & -A(\mu)^T \\ \Lambda(\mu) A(\mu) & C(\mu) \end{pmatrix}, \tag{3.48}$$

where Λ and C are diagonal matrices corresponding to λ and c.

We can again apply the implicit function theorem to deduce the existence of a differentiable trajectory $(x(\mu), \lambda(\mu))$ at (x^*, λ^*) if the matrix (3.48) is nonsingular at $\mu = 0$. Let J^* denote the limiting version of (3.48):

$$J^* = \begin{pmatrix} W^* & -A^{*T} \\ \Lambda^* A^* & C^* \end{pmatrix}.$$

Nonsingularity of J^* will follow if there is no nontrivial solution z to the system $J^* z = 0$. Partitioning z into an n-vector u and an m-vector v and using the form of J^*, the condition $J^* z = 0$ implies that

$$W^* u - A^{*T} v = 0 \quad \text{and} \quad \Lambda^* A^* u + C^* v = 0. \tag{3.49}$$

If $c_i^* > 0$, we know that $\lambda_i^* = 0$, and the second equation in (3.49) then implies that $v_i = 0$ for all inactive constraints. If, on the other hand, $c_i^* = 0$, the same relation implies that $\lambda_i^* (a_i^*)^T u = 0$ for $i \in \mathcal{A}$. It follows that u must lie in the null space of $\hat{A}$, the Jacobian of the active constraint gradients. Combining these properties, we see that

$$v_i = 0 \;\text{ for } i \notin \mathcal{A} \quad \text{and} \quad (a_i^*)^T u = 0 \;\text{ for } i \in \mathcal{A}. \tag{3.50}$$

If $J^* z = 0$, the scalar $z^T J^* z$ must also be zero. Writing out $z^T J^* z$ in terms of u and v, we have

$$z^T J^* z = u^T W^* u - v^T A^* u + v^T \Lambda^* A^* u + v^T C^* v = 0.$$

It follows from (3.50) that $v^T A^* u = 0$, $v^T \Lambda^* A^* u = 0$, and $v^T C^* v = 0$. Therefore, $u^T W^* u = 0$. But by assumption (c), this can be true only if $u = 0$.

If $u = 0$, the first equation in (3.49) implies that $A^{*T} v = \sum_{i=1}^m a_i^* v_i = 0$. Because the components of v corresponding to inactive constraints are zero, it follows that that $\sum_{i \in \mathcal{A}} a_i^* v_i = 0$. Under assumption (d) that the active constraint gradients are linearly independent, this can be true only if $v_i = 0$ for all $i \in \mathcal{A}$. But in this case, $u = 0$ and $v = 0$, which means that $z = 0$. Since $J^* z = 0$ only for a zero vector z, J^* is nonsingular. The implicit function theorem applied to (3.47) thus implies that the trajectory $x(\mu)$ exists and is differentiable at x^*.

The approach of the trajectory $x(\mu)$ to x^* can be analysed as follows. Let $\dot{x}$ denote $\mathrm{d}x(\mu)/\mathrm{d}\mu$, with a similar meaning for λ. Differentiating (3.47) with respect to μ, we see that, for sufficiently small μ, $x(\mu)$ and $\lambda(\mu)$ satisfy the system of differential equations

$$\begin{pmatrix} H - \sum \lambda_i H_i & -A^T \\ \Lambda A & C \end{pmatrix} \begin{pmatrix} \dot{x} \\ \dot{\lambda} \end{pmatrix} = \begin{pmatrix} 0 \\ e \end{pmatrix}, \tag{3.51}$$

with initial conditions $x(0) = x^*$ and $\lambda(0) = \lambda^*$.

Let y denote the vector $\dot{x}$ evaluated at $\mu = 0$, i.e., y is the tangent to the barrier trajectory at x^*. For an *active* constraint i, the second set of equations in (3.51) reveals that

$$\lambda_i^* (a_i^*)^T y = 1, \quad \text{so that} \quad (a_i^*)^T y = \frac{1}{\lambda_i^*} \quad \text{and} \quad \hat{A} y = (\hat{\Lambda}^*)^{-1} e. \tag{3.52}$$

The assumption of strict complementarity means that $\lambda_i^* \neq 0$ for any active constraint. Hence relation (3.52) shows that the barrier trajectory approaches x^* *nontangentially* with respect to the active constraints, i.e. the iterates do not converge 'along' the boundary.

Assume that constraint i is active, and let θ_i denote the angle between y and the normal to constraint i at x^*. It follows from (3.52) that

$$\cos\theta_i \sim \frac{1}{\|a_i^*\|\lambda_i^*}. \tag{3.53}$$

If all active constraint gradients are approximately equal in norm, relation (3.53) shows that the approach of barrier trajectory to x^* is 'closer to tangential' for active constraints with larger multipliers.

These properties are illustrated graphically with a two-variable example:

$$\begin{array}{ll} \text{minimize} & 2x_1x_2 - x_1^2 - x_2 \\ \text{subject to} & x_1^2 + x_2^2 \qquad \le 2 \\ & x_1^2x_2^2 \qquad\quad \le 10 \\ & (x_1 - \frac{1}{2})^2 + (x_2 - 1)^2 \le \frac{9}{4}. \end{array}$$

The first and third constraints (shown as dashed curves in Figure 3) intersect at $x^* = (-1, 1)^T$, which is an isolated local minimizer with Lagrange multipliers $\lambda_1^* = \frac{3}{2}$ and $\lambda_3^* = \frac{1}{3}$. The trajectory of barrier minimizers is depicted as a solid line converging to x^*. As expected, the trajectory approaches both active constraints along a nontangential path. The figure also confirms the prediction of (3.53) concerning the relative angles of approach to these constraints.

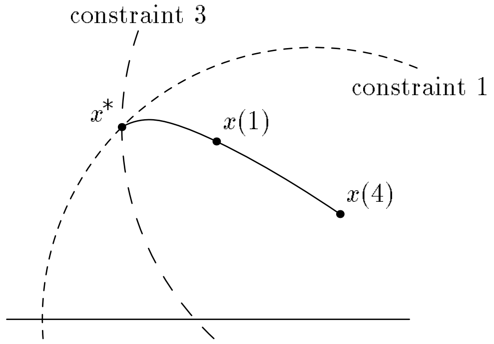

Figure 3. The non-tangential approach of $x(\mu)$ to x^*.

The nontangential property fails to hold without strict complementarity, even if the active constraint gradients are linearly independent. A complete analysis of this case is given in Jittorntrum (1978).

3.6. Properties of the barrier Hessian

The Hessian matrices H_B of the barrier function display a special structure as $\mu \to 0$. The barrier Hessian is given by

$$H_B(\mu) = H - \sum_{i=1}^{m} \frac{\mu}{c_i} H_i + \mu A^T C^{-2} A.$$

As $x(\mu)$ converges to x^*, the first two terms of the Hessian approach the Hessian of the Lagrangian function at x^*; the third matrix is given by

$$\mu A^T C^{-2} A = A^T D^2 A, \quad \text{where} \quad D^2 = \mu C^{-2} = \Lambda C^{-1} = \frac{1}{\mu}\Lambda^2. \tag{3.54}$$

We have already shown that the elements of D corresponding to inactive constraints are converging to zero, and the elements corresponding to active constraints are becoming unbounded (see (3.44)).

Let $\hat{m}$ denote the number of active constraints whose gradients are linearly independent. The Hessian of the barrier function can be characterized in three ways, depending on $\hat{m}$.

If no constraints are active at x^*, then the Hessian of the barrier function converges (as we would expect) to the Hessian of f itself.

At the other extreme, suppose that $\hat{m} = n$, so that the Jacobian of the active constraints has rank n. If $\hat{\lambda}^* > 0$, the barrier Hessian along the trajectory approaches a large multiple of the *nonsingular* matrix $\hat{A}^T(\hat{\Lambda}^*)^2\hat{A}$. In this case, the condition of the limiting Hessian depends on the condition of $\hat{A}$ and the condition of $\hat{\Lambda}^*$, but is not necessarily large. (This situation holds for linear programs in which there are no zero Lagrange multipliers.)

Finally, assume that $0 < \hat{m} < n$. The limiting matrix D^2 then contains $\hat{m}$ unbounded elements and $n - \hat{m}$ zero elements, which means that asymptotically A^TD^2A of (3.54) becomes not only unbounded, but also *rank-deficient.* Murray (1971) showed that $H_B(\mu)$ has $\hat{m}$ unbounded eigenvalues, corresponding to eigenvectors in the range space of $\hat{A}^T$, and $n - \hat{m}$ bounded eigenvalues, corresponding to eigenvectors in the null space of $\hat{A}$. The barrier Hessian accordingly becomes increasingly ill-conditioned for 'small' μ, and is *singular* in the limit. This property is one of the reasons that barrier function methods fell into disfavour in the 1970s, since standard unconstrained methods (such as Newton-based or quasi-Newton methods) tend to experience numerical difficulties when the Hessian is ill-conditioned. Various linear algebraic approaches have been proposed for dealing with this inherent ill-conditioning of the Hessian, and will be discussed in Section 7.2.

Our general analysis of barrier methods will be applied in Section 5 to the special case of linear programming. As background, we briefly summarize the relevant features of Newton's method in Section 4.

4. Newton's method

The application of interior methods to linear programming is heavily based on Newton's method, which we review in this section. Newton's method enters because satisfaction of certain nonlinear equations is definitive in optimality conditions (see Section 2.2), and the most popular technique for solving nonlinear equations is Newton's method.

4.1. Nonlinear equations and unconstrained minimization

We consider two forms for Newton's method. First, let $\Phi(x)$ denote an n-vector of smooth scalar functions $\varphi_i(x)$, $i = 1, \ldots, n$, and let $J(x)$ denote the Jacobian matrix of Φ. We seek a point x^* where $\Phi(x^*) = 0$. If x_k is the current point and $J(x_k)$ is nonsingular, the *Newton step* p_k is the step from x_k to the zero of the local affine model of Φ, and is the unique solution of the linear system

$$J_k p_k = -\Phi_k, \tag{4.1}$$

where J_k denotes $J(x_k)$ and Φ_k denotes $\Phi(x_k)$.

The second form of Newton's method is designed for unconstrained minimization of $f(x)$. Here, a quadratic model of the local variation of f is obtained from the Taylor-series expansion about x_k:

$$f(x_k + p) - f(x_k) \approx g_k^T p + \tfrac{1}{2} p^T H_k p,$$

where $g_k = g(x_k)$ and $H_k = H(x_k)$. If H_k is positive definite, the Newton step p_k is the step from x_k to the minimizer of this model, and satisfies the nonsingular linear system

$$H_k p = -g_k. \tag{4.2}$$

The direction p_k of (4.2), derived for minimization, is equivalent to the Newton step for solving the n-dimensional nonlinear system $g(x) = 0$.

4.2. Local convergence

A pure Newton method for either zero-finding or minimization begins with an initial point x_0, and generates a sequence of Newton iterates $\{x_k\}$, where

$$x_{k+1} = x_k + p_k \tag{4.3}$$

and p_k is defined by (4.1) or (4.2). Newton's method is often regarded as an 'ideal', in large part because of its fast *quadratic* convergence. When x_0 is sufficiently close to the solution and the relevant matrix (Jacobian or Hessian) is nonsingular, the error after each pure Newton step is effectively *squared*:

$$\|x_{k+1} - x^*\| = O(\|x_k - x^*\|^2).$$

Far from the solution, a pure Newton method is unreliable and may fail to converge. The standard way to encourage convergence from a general starting point is to require a reduction at each iteration in some *merit function* that measures progress. The new iterate is then defined by

$$x_{k+1} = x_k + \alpha_k p_k, \tag{4.4}$$

where α_k is a positive scalar called the *step length*.

To guarantee convergence, the step α_k must satisfy conditions known as *sufficient decrease* in $\|\Phi\|$ or f, as appropriate. The process of choosing α_k to produce a sufficient decrease in F is called a *line search*. Standard sufficient decrease conditions are discussed in detail in, for example, Ortega and Rheinboldt (1970). For well-behaved problems, the ultimate quadratic convergence of Newton's method can be retained because the 'natural' step of unity ($\alpha_k = 1$ in (4.4)) asymptotically produces a sufficient decrease.

4.3. Linear equality constraints

Newton's method for minimizing $f(x)$ subject to linear equality constraints can be derived in two ways: solving the nonlinear equations associated with optimality, or solving the constrained minimization subproblem derived from the local quadratic model.

The optimal x^* and multiplier λ^* can be viewed as the solution of the system (2.9) of $n + m$ nonlinear equations in the variables (x, λ). Given x_k and λ_k, we substitute the Jacobian from (2.9) into the generic Newton equation (4.1), which leads to two equivalent linear systems satisfied by the Newton step (p_k, δ_k):

$$\begin{pmatrix} H_k & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p_k \\ \delta_k \end{pmatrix} = \begin{pmatrix} H_k & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p_k \\ -\delta_k \end{pmatrix} = \begin{pmatrix} -g_k + A^T\lambda_k \\ b - Ax_k \end{pmatrix}. \tag{4.5}$$

(The second form has been rewritten with $-\delta_k$ as an unknown so that the matrix is symmetric.) The matrices in (4.5) are nonsingular whenever A has full rank and the reduced Hessian $N^T H_k N$ is positive definite. Another option is to treat the 'new' Lagrange multiplier $\lambda_{k+1} = \lambda_k + \delta_k$ as an unknown, producing the linear system

$$\begin{pmatrix} H_k & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p_k \\ -\lambda_{k+1} \end{pmatrix} = \begin{pmatrix} -g_k \\ b - Ax_k \end{pmatrix}. \tag{4.6}$$

The form (4.6) is often called the *augmented system*; the symmetric indefinite matrix in (4.6) is sometimes called the *KKT matrix*.

Viewed from a minimization perspective, the Newton step p_k is chosen to minimize the Taylor-series quadratic model of f subject to satisfying the constraints $Ax_{k+1} = b$. With this formulation, p_k solves the quadratic program

$$\underset{p \in \mathcal{R}^n}{\text{minimize}} \quad \tfrac{1}{2}p^T H_k p + g_k^T p \quad \text{subject to} \quad Ap = b - Ax_k. \tag{4.7}$$

This subproblem is itself an optimization problem subject to linear equality constraints. The Newton iterate $x_k + p_k$ satisfies the conditions (2.7a–b) for optimality of (4.7) if

$$Ap_k = b - Ax_k \quad \text{and} \quad g_k + H_k p_k = A^T \lambda_{k+1}. \tag{4.8}$$

After rearrangement, we again obtain the same linear system (4.6).

If x_k already satisfies the linear constraints, so that $Ax_k = b$, the Newton step is constrained to lie in the null space of A. In this case, the first equation in (4.8) becomes $Ap_k = 0$. If H_k is nonsingular, we may multiply the second equation by AH_k^{-1} and use the fact that $Ap_k = 0$, yielding the equations

$$\begin{aligned} AH_k^{-1}A^T\lambda_{k+1} &= AH_k^{-1}g_k; \\ p_k &= H_k^{-1}A^T\lambda_{k+1} - H_k^{-1}g_k. \end{aligned}$$

5. Linear programming

The main focus of interior methods since 1984 has been on linear programming. Much of the work on interior methods for LP can be viewed as an application of the general theory for barrier functions (Section 3), with enormous simplifications arising from the special properties of linear programs. Before describing specific interior methods, we give the relevant background on LP, emphasizing the special structure that is relevant to barrier methods.

It should be stressed in advance that hundreds of papers have been and continue to be written about interior LP methods, so that preparation of a complete list of references would be a daunting task. Fortunately, the excellent survey article of Gonzaga (1992) contains an extensive bibliography covering most aspects of the subject through mid-1991. A general bibliography on interior methods has been compiled by Kranich (1991), and can be accessed via electronic mail.

5.1. Background

For various historical and computational reasons, linear programs are widely stated in so-called *standard form*:

$$\underset{x}{\text{minimize}} \ c^T x \quad \text{subject to} \quad Ax = b, \quad x \geq 0, \tag{5.1}$$

where A is $m \times n$. The nonnegativity bound constraints $x \geq 0$ are the only inequalities in a standard-form problem. It is customary to assume that the rows of A are linearly independent. The point x is called *strictly feasible* for the linear program (5.1) if $Ax = b$ and $x > 0$.

A standard-form LP is a convex program (see Definition 10), and first-order conditions are sufficient for optimality. Combining results for linear equalities and general inequalities, we see that the feasible point x^* is a minimizer of the standard-form LP (5.1) if and only if, for some m-vector y^* and n-vector z^*,

$$c = A^T y^* + z^*, \quad z^* \geq 0, \quad \text{and} \quad z_i^* x_i^* = 0 \quad \text{for } i = 1, \ldots, n.$$

The vector z^* is the Lagrange multiplier for the inequality (simple bound) constraints, and y^* is the Lagrange multiplier for the equality constraints. We note that x^* and z^* satisfy a complementarity relation (see (2.13c)); because of the special nature of standard form, the values of the variables $\{x_i\}$ are also the values of the inequality constraints.

A well known property of linear programs is that, if the optimal objective value is finite, a *vertex* minimizer must exist. For a standard-form LP, a point $x^* \geq 0$ satisfying $Ax^* = b$ is a vertex if at least $n-m$ components of x^* are zero. At a *nondegenerate* vertex x^*, exactly $n-m$ components are zero. For details about linear programming and its terminology, see, e.g., Chvátal (1983) or Goldfarb and Todd (1989).

The LP (5.1) is traditionally called the *primal* problem. Its *dual* may be written in the inequality form

$$\underset{y}{\text{maximize}}\ b^Ty \quad \text{subject to} \quad A^Ty \leq c, \tag{5.2}$$

or in standard form:

$$\underset{y,z}{\text{maximize}}\ b^Ty \quad \text{subject to} \quad A^Ty + z = c, \quad z \geq 0. \tag{5.3}$$

The vector z in (5.3) is called the *dual slack*. The solution y^* of the dual is the Lagrange multiplier for the m general equality constraints in the primal, and the primal solution x^* is the Lagrange multiplier for the n equality constraints of the standard-form dual problem (5.3).

The termination criteria in many interior LP methods are based on an important relationship between the primal and dual objective functions. Let x be any primal-feasible point (satisfying $Ax = b$, $x \geq 0$) and y any dual-feasible point (satisfying $A^Ty \leq c$), with z the dual slack vector $c - A^Ty$. It is straightforward to show that

$$c^Tx - b^Ty = x^Tz \geq 0. \tag{5.4}$$

The necessarily nonnegative quantity $c^Tx - b^Ty$ is called the *duality gap*, and is zero if and only if x and y are *optimal* for the primal and dual.

Given a primal-feasible x and a dual-feasible y, the duality gap also provides a computable bound on the closeness of c^Tx to the optimal value c^Tx^*. Assume that $c^Tx - b^Ty = \beta$. Since $c^Tx^* = b^Ty^*$,

$$c^Tx - c^Tx^* = \beta + b^Ty - b^Ty^*.$$

Because the dual objective is maximized by y^* for all dual-feasible y, we know that $b^Ty^* \geq b^Ty$, which means that $b^Ty - b^Ty^* \leq 0$. Similarly, because the primal objective is minimized among all primal-feasible x, $c^Tx - c^Tx^* \geq 0$. Hence a duality gap of β at (x, y) implies

$$0 \leq c^Tx - c^Tx^* \leq \beta. \tag{5.5}$$

5.2. The central path

Suppose that we wish to apply a barrier method to a standard-form LP for which the following assumptions are satisfied:

(a) the set of x satisfying $Ax = b$, $x > 0$, is nonempty;

(b) the set (y, z) satisfying $A^Ty + z = c$, $z > 0$, is nonempty;

(c) $\operatorname{rank}(A) = m$.

Because the inequality constraints in a standard-form problem are exclusively simple bounds, the corresponding logarithmic barrier function is $B(x, \mu) = c^Tx - \mu \sum_{i=1}^n \ln x_i$.

The barrier subproblem involves minimizing $B(x,\mu)$ subject to satisfying the linear equality constraints:

$$\text{minimize} \quad c^T x - \mu \sum_{i=1}^{n} \ln x_i \quad \text{subject to} \quad Ax = b. \tag{5.6}$$

The gradient and Hessian of $B(x,\mu)$ for this case have particularly simple forms:

$$\nabla B(x,\mu) = c - \mu X^{-1} e, \quad \nabla^2 B(x,\mu) = \mu X^{-2}, \tag{5.7}$$

where $X = \mathrm{diag}(x_i)$.

The barrier subproblem (5.6) has a unique minimizer if assumption (b) is satisfied, i.e. there exist points that are strictly feasible for the dual problem. (This result can be deduced from the special nature of linear programs and Theorem 4.) The optimality conditions (2.7) for linear equality constraints imply the existence of y such that the solution of (5.6) satisfies

$$c - \mu X^{-1} e = A^T y \quad \text{or} \quad c = A^T y + \mu X^{-1} e.$$

Defining $z = \mu X^{-1} e$, we may write

$$c = A^T y + z \quad \text{and} \quad Xz = \mu e.$$

These equations are reminiscent of the equations (3.47) that hold along the barrier trajectory, since c is the objective gradient and the variables x are also the inequality constraints. The *central path* for a standard-form LP is defined by the vectors $x(\mu)$, $y(\mu)$ and $z(\mu)$ satisfying

$$\begin{aligned} Ax &= b, \quad x > 0 && (5.8a)\\ A^T y + z &= c, \quad z > 0 && (5.8b)\\ Xz &= \mu e. && (5.8c) \end{aligned}$$

The central path plays a crucial role in many interior LP methods; see Gonzaga (1992) for a detailed survey of methods based on the central path. We stress that relation (5.8c) formally defines the concept of 'centering' x and z, namely using the barrier parameter to control the distance of both vectors from zero (the boundary). Furthermore, the objective value along the path provides an estimate of the deviation from optimality; see (3.26).

5.3. The primal Newton barrier method

Assume that we are given a point x satisfying $Ax = b$ and $x > 0$, and that we wish to apply a barrier method to solve the standard-form LP (5.1). Using the forms (5.7) for the barrier gradient and Hessian, the Newton subproblem (4.7) for (5.6) is

$$\underset{p}{\text{minimize}} \quad \tfrac{1}{2}\mu p^T X^{-2} p + c^T p - \mu p^T X^{-1} e \quad \text{subject to} \quad Ap = 0. \tag{5.9}$$

The first-order optimality criteria of (2.7) applied to (5.9) show that the Newton direction p must satisfy

$$\mu X^{-2} p + c - \mu X^{-1} e = A^T y \tag{5.10}$$

for some Lagrange multiplier vector y. (We use y rather than λ for the Lagrange multiplier to retain consistency with LP notation.) Multiplying through by X^2 and noting that $Xe = x$, we obtain two expressions for p:

$$\begin{aligned} p &= \frac{1}{\mu} X^2 (A^T y - c + \mu X^{-1} e) && (5.11a)\\ p &= x + \frac{1}{\mu} X^2 (A^T y - c). && (5.11b) \end{aligned}$$

An expression for the Lagrange multiplier y is derived by multiplying (5.10) by AX^2 and using the relation $Ap = 0$ to eliminate p:

$$AX^2A^Ty = AX^2c - \mu AXe = AX(Xc - \mu e). \tag{5.12}$$

Because A has full rank and $x \neq 0$, the matrix AX^2A^T is positive definite.

Equation (5.12) has the familiar form of the *normal equations* for a linear least-squares problem with coefficient matrix XA^T. The vector y can therefore equivalently be represented as the solution of

$$\underset{y}{\text{minimize}} \ \ \|XA^Ty - (Xc - \mu e)\|_2^2. \tag{5.13}$$

The residual vector corresponding to (5.13) is given by $r = XA^Ty - (Xc - \mu e)$. Applying (5.11a), we see that the Newton direction p satisfies

$$p = \frac{1}{\mu}Xr,$$

and is a diagonally scaled multiple of the least-squares residual.

Alternatively, we can think of p and y as forming a combined vector of dimension $n + m$ that solves the linear equations corresponding to the augmented system (4.6) for (5.9):

$$\begin{pmatrix} \mu X^{-2} & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p \\ -y \end{pmatrix} = \begin{pmatrix} -c + \mu X^{-1}e \\ 0 \end{pmatrix}. \tag{5.14}$$

Since by construction $Ap = 0$, it follows that $A(x + \alpha p) = b$ for any step α along p whenever $Ax = b$. The pure Newton iterate for this problem (see (4.3)) is $\bar{x} = x + p$; however, a step of unity may violate strict feasibility. Because the constraints are linear, strict feasibility is retained if the step taken along p is less than the step $\hat{\alpha}$ to the boundary of the feasible region. With simple bound constraints, $\hat{\alpha}$ can be calculated directly: for all indices i such that $p_i < 0$, $\hat{\alpha}$ is the smallest value of $-x_i/p_i$.

A model primal Newton barrier algorithm includes both 'outer' and 'inner' iterations. The outer iterations reduce the barrier parameter, and the inner iterations apply Newton's method to solve the current barrier subproblem (5.6).

Primal Newton Barrier Algorithm

$k \leftarrow 0$; $\mu_0 > 0$; x_0 satisfies $Ax_0 = b$ and $x_0 > 0$;
while x_k is not sufficiently close to optimal for the LP **do**
  $x_k^0 \leftarrow x_k$; $i \leftarrow 0$;
  while x_k^i is not sufficiently close to optimal for (5.6) **do**
    Calculate the Newton direction p^i at x_k^i;
    $x_k^{i+1} \leftarrow x_k^i + \alpha^i p^i$, where $\alpha^i < \hat{\alpha}^i$ and $B(x_k^{i+1}, \mu_k) < B(x_k^i, \mu_k)$;
    $i \leftarrow i + 1$;
  end while
  $x_{k+1} \leftarrow x_k^i$;
  Choose $\mu_{k+1} < \mu_k$; $k \leftarrow k + 1$;
end while

The major computational effort associated with a primal barrier algorithm is the calculation of the Newton direction. The equations satisfied by the Newton direction can be written in a variety of theoretically equivalent forms, each of which suggests different linear algebraic techniques; the linear algebra issues will be discussed in Section 7.1.

The unspecified, implementation-dependent aspects of this algorithm include the selection of x_0 and μ_0, the strategy for altering the barrier parameter, and the choice of termination criteria for the inner and outer iterations. With suitable modification, this algorithm does not necessarily require a strictly feasible point. See Gonzaga (1992) for references on approaches that allow a general starting point.

With respect to the choice of the step α, it has been universally observed in practice that a very simple line search strategy of choosing α as a fixed fraction close to one (say, 0.95 or 0.99) of the step $\hat{\alpha}$ to the boundary almost always produces an adequate sufficient decrease (see Section 4.2) in the barrier function. In the extremely rare cases when this step is inadequate, a standard backtracking line search may be used. (For *nonlinear* problems, the issue of the line search becomes more complicated; see Murray and Wright (1991).)

Affine scaling interior methods were originally derived in terms of scaled steepest descent, and at first sight appear unrelated to barrier functions. However, the primal affine scaling method corresponds to defining y from (5.12) with $\mu = 0$. In general, an affine scaling method may be viewed as the limiting case of Newton's method applied to a barrier function. Gonzaga (1992) discusses the history of affine scaling methods.

In the first few years following 1984, affine scaling techniques experienced considerable popularity, in large part because of their simplicity, and were among the most effective in practice. At present, *primal-dual* methods, to be discussed in the next section, are accepted as the most efficient interior methods for LP. Certainly they are the most widely implemented in major software packages.

5.4. Primal-dual barrier methods

The primal Newton barrier algorithm just described finds a Newton step in only the primal variables x; the Lagrange multiplier y arises from the equality-constrained Newton subproblem (5.9). An alternative approach is motivated by finding primal and dual variables x, y, and z that satisfy the (rearranged) nonlinear equations from (5.8) that define the central path:

$$\Phi(x,y,z) = \begin{pmatrix} XZe - \mu e \\ Ax - b \\ A^T y + z - c \end{pmatrix} = 0. \tag{5.15}$$

Note that the second and third equations are linear; all the nonlinearity occurs in the first equation.

Applying Newton's method (4.1) to this system, we obtain

$$J \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} Z & 0 & X \\ A & 0 & 0 \\ 0 & A^T & I \end{pmatrix} \begin{pmatrix} p_x \\ p_y \\ p_z \end{pmatrix} = \begin{pmatrix} \mu e - XZe \\ b - Ax \\ c - A^T y - z \end{pmatrix}, \tag{5.16}$$

where J is the Jacobian of Φ in (5.15) and p_x, p_y and p_z are the Newton directions for x, y and z. Despite the difference in derivation, the linear systems associated with the Newton step in a primal-dual method have the same character as those in a primal method.

The third equation in (5.16) gives the expression

$$p_z = -A^T p_y + c - A^T y - z.$$

Substituting in the first equation to eliminate p_z, we obtain an augmented system involving p_x and p_y:

$$\begin{pmatrix} X^{-1}Z & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} p_x \\ -p_y \end{pmatrix} = \begin{pmatrix} \mu X^{-1}e - c + A^T y \\ b - Ax \end{pmatrix} \tag{5.17}$$

(see (4.6)). Since both x and z are strictly positive, the matrix $X^{-1}Z$ may be written as a positive diagonal matrix D^2, with $d_i^2 = z_i/x_i$.

Using $X^{-1}Z$ as a block pivot to eliminate p_x from the second equation of (5.17), the result is

$$AZ^{-1}XA^Tp_y = AZ^{-1}X(c - \mu X^{-1}e - A^Ty) + b - Ax. \tag{5.18}$$

As in (5.12), the matrix is symmetric and positive definite, with the form AD^2A^T. Once p_y is known, p_z and p_x may be calculated without solving further equations.

Finally, if $b - Ax$ may be written as ADv for some vector v (for example, if $Ax = b$), the equations (5.18) are the normal equations for a linear least-squares problem with matrix DA^T; see (5.13).

Primal-dual algorithms typically have a form similar to the primal algorithm given in Section 5.3. Most implementations choose separate steps for the primal and dual variables, in each case to ensure a sufficient decrease in some suitable merit function. When x and z are respectively primal- and dual-feasible, the easily-computable duality gap provides a guaranteed measure of the deviation from the optimal objective value; see (5.5).

Primal-dual methods of several varieties have been implemented, many with great practical success. An important feature not discussed here is the use of a 'predictor–corrector' technique closely related to extrapolation along the trajectory. For detailed discussion of primal-dual methods, see, for example, Lustig *et al.* (1990) and Mehrotra (1990).

6. Complexity issues

It is interesting as well as ironic that interior methods possess the same property as the simplex method: they are much faster in practice than indicated by complexity analysis. As we shall see, the typical upper bound on the number of iterations required by an interior method is *extremely* large for a problem of even moderate size. However, a lighthearted 'rule of thumb' articulated by several implementors is that interior LP methods tend to converge in an effectively *constant* number of iterations for many problems. This discrepancy has yet to be explained rigorously.

A major reason for the widespread interest in interior methods has been their provable polynomial complexity. Although formal complexity proofs are not typical in the numerical analysis literature, we have included one to indicate its flavour.

6.1. The role of problem size

In complexity proofs, it is standard to assume that *exact arithmetic* is used, and that all data of the problem (i.e. the entries of A, b and c) are *integers*. This is equivalent to assuming that the problem data are rational, since rational values can be rescaled to become integers. We therefore assume, in discussions of complexity only, that the entries of A, b and c are integers.

For a standard-form LP with n variables and m general constraints, the worst-case complexity of the simplex method depends on the number of vertices of the feasible region (which provides an upper bound on the number of iterations) and on the number of arithmetic operations required to perform an iteration. Both of these numbers can be bounded by expressions involving only the dimensions n and m.

When analysing the complexity of interior methods, however, a 'new' integer L makes an appearance. The usual interpretation is that L measures the 'size' of a linear program, and indicates the amount of information needed to represent an encoding of the problem.

The exact definition of L varies somewhat in the literature. For example, in Goldfarb and Todd (1989), the value of L for a standard-form LP with n variables and m equality constraints is defined as

$$L = \sum_{i=0}^{m}\sum_{j=0}^{n}\lceil \log(|a_{ij}| + 1) + 1\rceil, \tag{6.1}$$

where $a_{i0} = b_i$ and $a_{0j} = c_j$. It should be stressed that L can be enormous for problems of even moderate dimension.

The value of L enters the complexity analysis at both lower (termination) and upper (initialization) extremes. The role of L in the initialization of interior methods will be discussed following Theorem 9.

With respect to termination, Khachian (1979) showed that the smallest possible nonzero variation in the objective function between any two distinct vertices is expressible in terms of L. In particular, if x is any vertex, then c^Tx is either equal to the optimal value c^Tx^* or must exceed the optimal value by at least 2^{-2L}. This bound depends on the fact that an optimal vertex is the solution of a linear system involving b and a nonsingular submatrix of A. Under an integrality assumption on the entries of A and b, Cramer's rule shows that the exact solution of such a system is a vector of rational numbers, such that the absolute value and denominator of each component are bounded by $2^{O(L)}$.

A stopping rule that defines acceptable closeness to optimality is needed for interior methods because the exact solution of an LP (a vertex) cannot be produced in a finite number of iterations by a method that generates strictly feasible iterates. Given any feasible point x, a formal 'rounding' procedure requiring $O(n^3)$ operations is known that will produce a vertex $\bar{x}$ for which $c^T\bar{x} \le c^Tx$. If the objective value at an interior point is known to be within 2^{-2L} of the optimal value, a single application of the rounding procedure will produce an optimal vertex. This result follows from the property stated above concerning the minimum nonzero variation in objective values between vertices; see Papadimitriou and Steiglitz (1982) and Gonzaga (1992) for details.

Although the optimal objective value is in general unknown, an interior method that constructs primal- and dual-feasible points can use the duality gap to provide a computable upper bound on the difference between the current and optimal objective values; see (5.5).

6.2. A polynomial-time path-following algorithm

The material in this section closely follows Roos and Vial (1988); similar proofs are given in Monteiro and Adler (1989a). See Gonzaga (1992) for a survey of path-following strategies and the associated complexity bounds.

The argument typifies complexity proofs involving Newton steps and the central path. The fundamental ideas are: first, defining a computable measure of closeness to the central path; second, showing that a Newton step retains a sufficient degree of closeness to the path; and finally, decreasing the barrier parameter at a rate that allows a polynomial upper bound on the number of iterations required to reduce the duality gap to less than $2^{-O(L)}$.

An important element of the proof is a suitable definition of a *proximity measure* $\delta(x,\mu)$, which measures closeness to the central path. This quantity is defined for a strictly feasible x and positive barrier parameter μ. Let $y(x,\mu)$ and $z(x,\mu)$ denote the vectors satisfying $A^Ty + z = c$ for which $\|Xz - \mu e\|$ is minimized. This requirement means that $y(x,\mu)$ and $z(x,\mu)$ solve an optimization problem with a quadratic objective function and linear equality constraints:

$$\underset{y,z}{\text{minimize}}\ \tfrac{1}{2}z^TX^2z - \mu x^Tz \quad \text{such that} \quad A^Ty + z = c. \tag{6.2}$$

Problem (6.2) is merely a conceptual formalism; the required vectors are implicit in the calculations (5.11) and (5.12) associated with the primal Newton direction:

$$y(x,\mu) = y \quad \text{and} \quad z(x,\mu) = c - A^Ty.$$

The projected Newton direction p (5.11) may consequently be written as

$$p = x - \frac{1}{\mu}X^2z, \tag{6.3}$$

where z is the optimal $z(\mu)$ from (6.2).

The proximity measure $\delta(x,\mu)$ is defined as

$$\delta(x,\mu) = \|X^{-1}p\| = \left\|\frac{Xz}{\mu} - e\right\|, \quad \text{where} \quad z = z(x,\mu). \tag{6.4}$$

If x solves (5.6), i.e. x lies on the central path, then $\delta(x,\mu) = 0$. The value of δ is thus a scaled measure of the Newton step, and indicates the distance of x from the central path.

It is convenient to use the n-vector s whose ith component is $s_i = x_i z_i/\mu$, so that

$$s = \frac{Xz}{\mu} \quad \text{and} \quad z = \mu X^{-1} s. \tag{6.5}$$

The definitions of δ and s imply the relations

$$\delta^2 = \sum_{i=1}^{n}\left(\frac{x_i z_i}{\mu} - 1\right)^2 = \|s-e\|^2 = \sum_{i=1}^{n}(s_i - 1)^2. \tag{6.6}$$

It follows from (6.3) and (6.5) that the next iterate $\bar{x}$ of a pure Newton method is

$$\bar{x} = x + p = 2x - \frac{X^2 z}{\mu} = 2x - Xs. \tag{6.7}$$

Component-wise, the new iterate satisfies

$$\bar{x}_i = 2x_i - x_i s_i = (2 - s_i)x_i. \tag{6.8}$$

Before treating the algorithm itself, we show that the duality gap is bounded if the proximity measure is sufficiently small.

Lemma 2 *(Bounds on the duality gap.) If x is strictly feasible, $\delta(x,\mu) < 1$ and the vectors y and z solve (6.2), then y is dual-feasible (i.e., $A^T y \le c$) and*

$$\mu\Big(n - \delta(x,\mu)\sqrt{n}\Big) \;\le\; c^T x - b^T y \;\le\; \mu\Big(n + \delta(x,\mu)\sqrt{n}\Big).$$

Proof. Since $x > 0$ and $\delta(x,\mu) < 1$, it follows from the first equation in (6.6) that $x_i z_i \ge 0$. Hence $z \ge 0$, which means that $A^T y \le c$ and y is dual-feasible.

Because x and y are primal- and dual-feasible, we know from (5.4) that the duality gap is given by $x^T z$, with $z = c - A^T y$. By definition of δ and e,

$$\delta(x,\mu)\sqrt{n} = \left\|\frac{Xz}{\mu} - e\right\|\|e\|.$$

Applying the Cauchy-Schwarz inequality, we obtain

$$\delta(x,\mu)\sqrt{n} = \left\|\frac{Xz}{\mu} - e\right\|\|e\| \ge \left|e^T\left(\frac{Xz}{\mu} - e\right)\right| = \left|\frac{x^T z}{\mu} - n\right|,$$

which leads to

$$n - \delta(x,\mu)\sqrt{n} \;\le\; \frac{x^T z}{\mu} \;\le\; n + \delta(x,\mu)\sqrt{n}.$$

Multiplying by μ gives the desired result. □

We now begin a sequence of lemmas that prove crucial relationships about Newton steps and the proximity measure. First, we show that, if the proximity measure is sufficiently small for given x and μ, the proximity measure for the same value of μ is squared at the next Newton iterate.

Lemma 3 *(Quadratic convergence of proximity measure.) Let x satisfy $Ax = b$, $x > 0$, and assume that $\delta(x,\mu) < 1$. Then the next Newton iterate $\bar{x}$ (6.7) also satisfies $A\bar{x} = b$ and $\bar{x} > 0$. Further, $\delta(\bar{x},\mu) \le \delta(x,\mu)^2$.*

Proof. Because $\delta(x,\mu) < 1$, it follows from (6.6) that $|s_i - 1| < 1$, so that $0 < s_i < 2$ for $1 \le i \le n$. Relation (6.8) then implies that $\bar{x} > 0$. The fact that $A\bar{x} = b$ is immediate from the construction (5.9) of p to satisfy $Ap = 0$.

Because $\delta(\bar{x},\mu)$ is the smallest value of $\|\bar{X}z/\mu - e\|$ for all vectors y and z satisfying $A^T y + z = c$, we have

$$\delta(\bar{x},\mu) \le \left\| \frac{\bar{X}z}{\mu} - e \right\|.$$

Using the relations $z = \mu X^{-1}s$ and $\bar{x}_i = 2x_i - x_i s_i$ gives

$$\frac{\bar{X}z}{\mu} = \bar{X}X^{-1}s = (2X - XS)X^{-1}s = 2s - S^2 e.$$

Therefore, $\delta(\bar{x},\mu) \le \|2s - S^2 e - e\|$, which means that

$$\delta(\bar{x},\mu)^2 \le \sum_{i=1}^{n} (2s_i - s_i^2 - 1)^2 = \sum_{i=1}^{n} (s_i - 1)^4 \le \left(\sum_{i=1}^{n} (s_i - 1)^2 \right)^2 = \delta(x,\mu)^4.$$

The condition $\delta(x,\mu) < 1$ thus ensures that the pure Newton iterates converge quadratically to the point $x(\mu)$ on the central path. □

The next lemma develops a bound on the proximity criterion corresponding to a reduced value of the barrier parameter.

Lemma 4 *(Effect of a reduction in μ.) If θ satisfies $0 < \theta < 1$ and $\bar{\mu}$ is defined as $(1-\theta)\mu$, then*

$$\delta(x,\bar{\mu}) \le \frac{\delta(x,\mu) + \theta\sqrt{n}}{1-\theta}.$$

Proof. It follows from the definition of δ that, for $z = z(x,\mu)$,

$$\delta(x,\bar{\mu}) \le \left\| \frac{Xz}{\bar{\mu}} - e \right\|.$$

Let

$$\nu = \frac{1}{1-\theta} = \frac{\mu}{\bar{\mu}},$$

so that $\nu > 1$ and $\nu - 1 = \theta/(1-\theta)$. Then

$$\delta(x,\bar{\mu}) \le \left\| \frac{Xz}{\bar{\mu}} - e \right\| = \left\| \nu\frac{Xz}{\mu} - e \right\| = \|\nu s - e\|.$$

Applying the triangle inequality gives

$$\begin{aligned} \delta(x,\bar{\mu}) &\le \|\nu(s-e) + (\nu-1)e\| \\ &\le \nu\|s-e\| + (\nu-1)\|e\| \\ &= \frac{\delta(x,\mu)+\theta\sqrt{n}}{1-\theta}, \end{aligned}$$

which is the desired result. □

We now combine the preceding two lemmas to obtain a bound on the proximity measure for the Newton iterate with a barrier parameter that has been reduced by a factor related to the problem dimension.

Lemma 5 *(Bounds on the proximity measure.) Assume that $\delta(x,\mu) \le \frac{1}{2}$, and let*

$$\theta = \frac{1}{6\sqrt{n}}, \quad \textit{so that} \quad \theta\sqrt{n} = \frac{1}{6}.$$

When $\bar{x}$ is the Newton iterate (6.7) and $\bar{\mu} = (1-\theta)\mu$, then $\delta(\bar{x},\bar{\mu}) \le \frac{1}{2}$.

Proof. Applying first Lemma 4 and then Lemma 3, we have

$$\begin{aligned} \delta(\bar{x},\bar{\mu}) &\le \frac{\delta(\bar{x},\mu)+\theta\sqrt{n}}{1-\theta} \le \frac{\delta(x,\mu)^2+\theta\sqrt{n}}{1-\theta} \\ &\le \frac{\frac{1}{4}+\frac{1}{6}}{1-\theta} = \frac{\frac{5}{12}}{1-\theta} \\ &\le \tfrac{1}{2}, \end{aligned}$$

where the last inequality holds because $1/(1-\theta) \le \frac{6}{5}$. □

An approximate path-following method based on reducing the barrier parameter and taking a single Newton step is obviously suggested by these results. Assume that we are given a strictly feasible x_0 and barrier parameter μ_0 such that $\delta(x_0,\mu_0) \le \frac{1}{2}$; the latter condition can always be satisfied, as we shall discuss after the proof of Theorem 9. The following algorithm constructs a sequence of pairs (x_k,μ_k) such that every x_k is strictly feasible, $\mu_k > 0$, $\delta(x_k,\mu_k) \le \frac{1}{2}$, and $\mu_k \to 0$ as $k \to \infty$.

Let q be an accuracy parameter, to be described later, and define θ as $1/(6\sqrt{n})$.

Algorithm I

$k \leftarrow 0$;
while $n\mu_k > \mathrm{e}^{-q}$
  $\mu_{k+1} \leftarrow (1-\theta)\mu_k$;
  $x_{k+1} \leftarrow 2x_k - X_k^2 z_k/\mu_k$, where $z_k = z(x_k,\mu_k)$ of (6.2);
  $k \leftarrow k+1$;
end while

An upper bound for the number of iterations required by Algorithm I is given in the following theorem.

Theorem 9 *(Worst-case number of iterations.) Define $q_0 = \lceil \ln(n\mu_0) \rceil$. Then Algorithm I will terminate after at most $6(q+q_0)\sqrt{n}$ steps, and the final iterate x and the corresponding y obtained from (6.2) satisfy*

$$c^T x - b^T y \le \frac{3}{2}\,\mathrm{e}^{-q}.$$

Proof. We know from Lemma 3 that every iterate x_k is strictly feasible, and from Lemma 5 that $\delta(x_k,\mu_k) \le \frac{1}{2}$. The algorithm terminates when k satisfies $n\mu_k \le \mathrm{e}^{-q}$, where by construction $\mu_k = (1-\theta)^k\mu_0$. Applying the definition of q_0, termination will occur when

$$n\mu_k = n(1-\theta)^k\mu_0 \le (1-\theta)^k \mathrm{e}^{q_0} \le \mathrm{e}^{-q}.$$

Taking logarithms, the termination condition is

$$-\,k\,\ln(1-\theta) > q + q_0. \tag{6.9}$$

Since $-\ln(1-\theta) > \theta$ for all $\theta < 1$, the inequality (6.9) holds if $k\theta > q + q_0$. Using the definition of θ, we see that the algorithm terminates if k satisfies

$$k > 6\sqrt{n}(q + q_0),$$

which gives the first desired result.

For the final iterate x_k, let $y_k = y(x_k, \mu_k)$. Lemma 2 implies that y_k is dual feasible and that

$$c^T x_k - b^T y_k \le \mu_k \Big(n + \delta(x_k, \mu_k)\sqrt{n} \Big).$$

Since $n\mu_k \le \mathrm{e}^{-q}$ and $\delta(x_k, \mu_k) \le \frac{1}{2}$, rearrangement gives

$$c^T x_k - b^T y_k \le \mathrm{e}^{-q} \Big(1 + \frac{\delta(x_k, \mu_k)}{\sqrt{n}} \Big) \le \frac{3}{2}\, \mathrm{e}^{-q},$$

which completes the proof. □

To conclude that this bound is polynomial, we need to connect L (6.1) to both the initial barrier parameter μ_0 (which defines q_0) and to the accuracy parameter q.

The value of μ_0 is related to L in detailed proofs by Monteiro and Adler (1989a, b), who show that any LP is polynomially equivalent (i.e. can be transformed in a polynomial number of operations) to an LP of similar size. For this related LP, a strictly feasible x_0 and an initial $\mu_0 = 2^{O(L)}$ are known such that x_0 lies on the central path. With these choices of μ_0 and x_0, the value of q_0 is $O(L)$, and $\delta(x_0, \mu_0) = 0$. It should be emphasized that this value of μ_0 is enormous, and would never be used in a practical algorithm.

Turning now to the accuracy parameter q, we know from our initial discussion of L that the algorithm should terminate when the duality gap is less than $2^{-O(L)}$. In Algorithm I, the duality gap is tested against e^{-q}. Consequently, q should be chosen as $O(L)$ to ensure that the rounding procedure will produce an optimal vertex from the final iterate.

With both q and q_0 taken as $O(L)$, the bound of Theorem 9 is indeed $O(\sqrt{n}L)$ iterations. Finally, each iteration of the algorithm requires $O(n^3)$ operations, to calculate p, y and z from (5.11) and (5.12). The total computational effort for Algorithm I is therefore $O(n^{3.5}L)$, which is (as promised) a polynomial in the problem size.

Polynomiality has been proved for a wide variety of interior methods for linear and quadratic programming. See, for example, Monteiro and Adler (1989b), in which the nature of the 'rounding' required for QP is described in detail. Various authors have proposed interior path-following methods for convex nonlinear problems satisfying certain assumptions. Recent discussion of these approaches is given in, for example, Nesterov and Nemirovsky (1989), den Hertog *et al.* (1990), Mehrotra and Sun (1990), and Jarre (1991). With these methods, polynomial bounds can be proved only on the number of iterations, since no rounding procedure exists for general nonlinear problems.

7. Linear algebraic issues

A persuasive argument can be made that the *practical* success of interior methods depends on numerical linear algebra. For very large problems, even (say) 40 iterations of an interior method would be inordinately time-consuming if the associated linear systems could not be solved efficiently and reliably.

7.1. Linear algebra in interior LP methods

The linear systems in interior methods for linear programming have a strikingly different nature from those associated with the simplex method. At each simplex iteration, two (transposed) square $m \times m$ systems are solved, and the matrix changes by only a single column per iteration. Typical implementations of the simplex method perform an initial sparse LU factorization of the basis, followed by Forrest–Tomlin or Bartels–Golub updates. As an aside, we stress that linear algebra in the 'real' simplex method bears almost no resemblance to a typical textbook tableau. Recent discussions of selected linear algebraic issues in the simplex method may be found in Bixby (1990), Duff *et al.* (1986) and Forrest and Tomlin (1990, 1991).

Interior LP methods have been of practical interest mainly for large problems, and we henceforth assume that the matrix A is large and sparse. In most implementations to date, the Newton direction is calculated from equations arising in two theoretically equivalent formulations:

(i) *Normal-equation* form, involving an $m \times m$ symmetric positive-definite matrix AD^2A^T (see (5.12) and (5.18));

(ii) *Augmented system* form, containing an $(n+m)$-dimensional specially-structured symmetric indefinite matrix (see (5.14) and (5.17)).

Least-squares problems such as (5.13) have primarily been solved by conversion to (i) or (ii), although some interest remains in application of sparse QR factorizations. A complete discussion of the relevant linear algebraic issues for all these approaches is given in Björck (1991), along with an extensive bibliography.

With either (i) or (ii), the following features are important:

- The $n \times n$ matrix D changes completely at every iteration, but its elements are converging to quantities associated with x^*;
- The Newton direction need not necessarily be computed with high accuracy, since it is only a means to follow the path. Unless there is a complete breakdown in accuracy, the line search ensures progress for any direction of descent with respect to the particular merit function.

The simplest and by far the most popular linear algebraic technique for the normal-equation approach is direct solution: we explicitly form AD^2A^T and compute its Cholesky factorization,

$$AD^2A^T = R^TR,$$

where R is upper triangular. Sparse Cholesky factorizations have been widely studied and carefully implemented in several sparse matrix packages. Comprehensive discussions are given in, for example, George and Liu (1981) and Duff *et al.* (1986).

Most implementations of a sparse Cholesky factorization perform an initial symbolic *analyse phase* that constructs a pivot ordering intended to produce a sparse factor R. When AD^2A^T is sufficiently positive definite, all pivoting orders are numerically stable, so that the ordering need not be altered later. Because only the diagonal scaling D changes at each iteration of an interior LP method, a single analyse phase suffices for all iterations. After a suitable ordering is determined, the triangular matrix R is calculated using the numerical values in AD^2A^T.

Standard ordering heuristics, most commonly minimum degree and minimum local fill, have been very effective in interior methods. The calculation of R has been organized in both 'left-looking' and 'right-looking' versions. The best choice of ordering and organization has been found, not surprisingly, to depend on details of the hardware such as vectorization and memory hierarchy.

Although the barrier Hessian is nonsingular at strictly interior iterates, it becomes asymptotically singular when the linear program is dual degenerate. Many (some would say most) real linear programs display a high degree of dual degeneracy, leading to obvious ill-conditioning in AD^2A^T. The small number of observed numerical difficulties with the normal-equation approach has therefore been a continuing surprise. A careful error analysis is likely to explain this phenomenon, but it remains slightly mysterious at this time.

The major practical difficulty with forming the Cholesky factorization of AD^2A^T is known as the 'dense column' problem. If any columns of A contain a relatively large number of nonzeros, the matrix AD^2A^T is much denser than A. (If A has even one entirely dense column, AD^2A^T fills in completely.) To retain efficiency, some strategy must be developed to detect and treat dense columns separately.

Suppose that A is partitioned into two subsets of columns, with a similar partition of D:

$$A = (A_1 \;\; A_2), \quad \text{so that} \quad AD^2A^T = A_1D_1^2A_1^T + A_2D_2^2A_2^T,$$

where A_2 contains the dense columns. The hope is to solve systems involving AD^2A^T without forming the matrix explicitly, using a Cholesky factorization of the 'sparse part':

$$A_1D_1^2A_1^T = R_1^TR_1.$$

A direct strategy can be devised by observing that the solution p of $AD^2A^Tp = d$ also satisfies

$$\begin{pmatrix} A_1D_1^2A_1^T & A_2 \\ A_2^T & -D_2^{-2} \end{pmatrix} \begin{pmatrix} p \\ z \end{pmatrix} = \begin{pmatrix} d \\ 0 \end{pmatrix}. \tag{7.1}$$

It is well known that the extended system (7.1) can be solved if we can solve linear systems involving $A_1D_1^2A_1^T$ and the (negative) Schur complement

$$C = D_2^{-2} + A_2^T(A_1D_1^2A_1^T)^{-1}A_2.$$

The matrix C can be expressed in terms of R_1 as

$$C = D_2^{-2} + A_2^T(R_1^TR_1)^{-1}A_2 = D_2^{-2} + U^TU,$$

where $U = R_1^{-T}A_2$. The desired vector p is found by solving (in order)

$$R_1^Tv = d, \quad Cz = U^Tv, \quad R_1p = v - Uz.$$

If the column dimension of A_2 is small, the positive-definite matrix C can be formed and factorized without undue effort. This technique for dealing with dense columns is discussed in, for example, Marxen (1989) and Choi *et al.* (1990).

A second approach involves applying an *iterative* technique such as the conjugate gradient method. Given that the equations need not be solved exactly, there is some hope that the required number of iterative steps will not be too large on average. Because AD^2A^T is often ill-conditioned, preconditioning is essential. An obvious source for the preconditioner is a 'sparsified' Cholesky factorization of AD^2A^T, such as the factorization of $A_1D_1^2A_1^T$; see, for example, Gill *et al.* (1986). Other strategies combining direct and iterative techniques have also been devised; see Lustig *et al.* (1990).

A drawback with either strategy is that the matrix $A_1D_1^2A_1^T$ remaining after removal of dense columns has frequently been found to be extremely ill-conditioned or even numerically singular. A second problematic aspect is that heuristic criteria must be developed to identify which columns qualify as dense.

We now turn to formulation (ii) – solving an augmented system in which the matrix has one of the forms

$$K = \begin{pmatrix} D^{-2} & A^T \\ A & 0 \end{pmatrix} \quad \text{or} \quad M = \begin{pmatrix} \beta I & \tilde{D}A^T \\ A\tilde{D} & 0 \end{pmatrix}. \tag{7.2}$$

The second matrix arises from a least-squares formulation, and the scalar β is a scaling factor included to improve stability. Its selection is a compromise between preserving sparsity and maintaining stability; see Arioli *et al.* (1989) and Björck (1991).

Both K and M are symmetric but obviously indefinite. (We shall refer to K in the discussion, but most comments apply also to M.) The standard direct method for solving systems of this form involves calculation of the symmetric indefinite factorization

$$P^T K P = L B L^T,$$

where P is a permutation matrix, L is unit lower-triangular, and B is block-diagonal, with 1×1 or 2×2 blocks. For dense problems, P is chosen using a stability criterion that determines whether to use a 1×1 or 2×2 pivot; see Bunch and Kaufman (1977).

In contrast to the positive-definite case, it cannot be guaranteed that all pivoting orders for a symmetric indefinite matrix are numerically stable. The analyse phase for the symmetric indefinite factorization thus attempts to choose a pivot ordering based solely on sparsity that will lead to low fill-in in L. When the factorization itself is computed with the actual numerical values, interchanges that alter the predicted pivot sequence may be required to retain numerical stability.

The augmented system approach involves an increase in dimension compared to the normal equations, as well as a more complicated factorization. Nonetheless, solving the augmented system should be more reliable numerically, particularly in avoiding instabilities attributable to dense columns. Very promising results have been reported by Fourer and Mehrotra (1990).

As a compromise between approaches (i) and (ii), some suggestions have been made for working with 'partially reduced' augmented systems of the form K in (7.2). The idea is to perform a block pivot in K with the 'good' part of D, simultaneously producing a smaller system and retaining numerical stability. In any such approach, the dense columns of A are placed in the portion of K that is not factorized; see, for example, Vanderbei (1991) and Gill *et al.* (1991). Alternatively, the indefinite system can be solved using an iterative method with a sparse preconditioner; see Gill *et al.* (1990).

Although taking advantage of symmetry often leads to savings in storage and computation, some linear algebra issues are simplified by ignoring symmetry. An approach that deserves exploration is the use of unsymmetric but highly structured systems, such as (5.16).

Much opportunity clearly remains for improvements and refinements in the linear algebraic aspects of interior LP methods.

7.2. Linear algebra for nonlinear problems

For nondegenerate linear programs, the results of Section 3.6 show that the barrier Hessian is asymptotically nonsingular, since $\hat{A}$ (the Jacobian matrix of the active constraints) has rank n. As soon as we consider *nonlinear* problems (including quadratic programming), however, in general the Hessian of the barrier function becomes increasingly ill-conditioned as the solution is approached along the trajectory Since the exact solution of an ill-conditioned problem is by definition extremely sensitive to small changes in the data, interior methods might appear to be fundamentally unsound.

Fortunately, a more optimistic view is justified by several observations. Inherent ill-conditioning afflicts the barrier Hessian only 'near' the solution, which is precisely where asymptotic properties of the Lagrangian function and the barrier trajectory apply. In particular, a 'good' step toward x^* from a point sufficiently near x^* is *not* poorly determined. The ill-conditioning is consequently an artifact of the barrier transformation rather than inherent to the constrained problem. In effect, the ill-conditioning gradually and implicitly reveals subspace information whose asymptotic nature is known.

If the correct active set is identified, a highly accurate approximation to the Newton step can be calculated in two orthogonal 'pieces' lying in the range of $\hat{A}^T$ and the null space of $\hat{A}$, where the condition of the relevant equations reflects that of the original problem; see Wright (1976). But since a definitive property of interior methods is that they do *not* make an explicit identification of the active set, it is arguably inappropriate to make such a prediction. More recent work on this issue has several flavours.

The Newton equations can be solved using a rank-revealing Cholesky (or modified Cholesky) factorization with symmetric interchanges (Higham, 1990), where linear algebraic criteria are invoked to define numerical rank-deficiency. When the condition of the Hessian becomes excessively large, its Cholesky factors lead to bases for the required range and null spaces (Wright, 1991).

If the nonlinear constraints are formulated in 'standard form', namely as $c(x) = 0$, $x \geq 0$, the barrier transformation applies only to the simple bounds. (Inequality constraints can always be converted to standard form by adding nonnegative slack variables.) The resulting Hessian of the barrier function asymptotically approaches the Hessian of the Lagrangian plus a diagonal matrix, some of whose entries are becoming unbounded. In this form, the ill-conditioning is concentrated entirely in large diagonal elements of the Hessian, and does not affect the sensitivity of the solution of the associated KKT system (Ponceleón, 1990).

Finally, in the spirit of seeking nonsymmetric matrices that may avoid difficulties with symmetric forms, we recall from Section 3.5 that the matrix

$$\begin{pmatrix} H - \sum \lambda_i H_i & -A^T \\ \Lambda A & C \end{pmatrix},$$

which arises in a primal-dual characterization of the barrier trajectory, is *nonsingular* at x^*, and does not suffer inevitable ill-conditioning. A nonlinear primal-dual algorithm can thus be developed in which the linear systems are unsymmetric but well-conditioned; see McCormick (1991).

It is still unknown which, if any, of these strategies will be most successful in overcoming the difficulties with conditioning that plagued barrier methods for nonlinear problems in the 1960s and 1970s.

8. Future directions

Many issues remain to be resolved for interior methods, even for linear programming. At the most basic level, the problem categories for which simplex and interior methods are best suited are not well understood. In addition, the gap between worst-case and average-case performance has not been satisfactorily explained.

One great strength of the simplex method is its efficient 'warm start' capability. Many large linear programs do not arise only once, in isolation, but are modified versions of an underlying model. After each change in the model, the resulting LP is re-solved. Because the simplex method can make effective use of *a priori* information, it is not uncommon for the solution to be found in a very small number of simplex iterations – many fewer than if the problem were solved from scratch. In contrast, the very nature of interior methods is to move away from the boundary and then approach the solution along a central path. No effective strategy has yet been devised for allowing interior methods to exploit 'strong hints' about the constraints active at the solution.

For nonlinear problems, researchers are returning with fresh enthusiasm to old topics, such as the treatment of ill-conditioning, the choice of merit function, and termination of the solution of each barrier subproblem. The work of Nesterov and Nemirovsky (1989) suggests new, previously unconsidered, barrier functions, which may be of practical as well as theoretical significance.

It seems safe to predict that the field of interior methods will continue to produce interesting research to suit every taste.

Acknowledgments

The figures in this paper were produced with the MetaPost system of John Hobby; his patient help with MetaPost and fine points of LaTeX is gratefully acknowledged. I thank Ken Clarkson, David Gay, Arieh Iserles, Jeff Lagarias, Mike Powell and Norm Schryer for helpful comments on content and exposition. Special thanks go to Philip Gill for his detailed reading of the manuscript and numerous suggestions for improvements.

References

[1] M. Arioli, I. S. Duff and P. P. M. de Rijk (1989), 'On the augmented system approach to sparse least-squares problems', *Numer. Math.* **55**, 667–684.

[2] M. Avriel (1976), *Nonlinear Programming: Analysis and Methods*, Prentice-Hall (Englewood Cliffs, NJ).

[3] D. A. Bayer and J. C. Lagarias (1989), 'The nonlinear geometry of linear programming, I: Affine and projective scaling trajectories', *Trans. Am. Math. Soc.* **314**, 499–526.

[4] D. A. Bayer and J. C. Lagarias (1991), 'Karmarkar's linear programming algorithm and Newton's method', *Math. Program.* **50**, 291–330.

[5] R. Bixby (1990), 'Implementing the simplex method: the initial basis', Report 90-32, Department of Mathematical Sciences, Rice University, Houston, Texas.

[6] A. Björck (1991), 'Least squares methods', Institute of Technology, Linköping University, Linköping, Sweden.

[7] J. R. Bunch and L. C. Kaufman (1977), 'Some stable methods for calculating inertia and solving symmetric linear systems', *Math. Comput.* **31**, 162–179.

[8] I. C. Choi, C. L. Monma and D. F. Shanno (1990), 'Further development of a primal-dual interior point method', *ORSA J. Comput.* **2**, 304–311.

[9] V. Chvátal (1983), *Linear Programming*, W. H. Freeman (New York).

[10] D. den Hertog, C. Roos and T. Terlaky (1990), 'On the classical logarithmic barrier function method for a class of smooth convex programming problems', Report 90-01, Faculty of Technical Mathematics and Informatics, Delft University of Technology, Delft, Holland.

[11] I. S. Duff, A. M. Erisman and J. K. Reid (1986), *Direct Methods for Sparse Matrices*, Oxford University Press (London).

[12] A. V. Fiacco (1979), 'Barrier methods for nonlinear programming', in *Operations Research Support Methodology* (A. G. Holzman, ed.), Marcel Dekker (New York), 377–440.

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

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

[15] J. J. H. Forrest and J. A. Tomlin (1990), 'Vector processing in simplex and interior methods for linear programming', *Annals of Operations Research* **22**, 71–100.

[16] J. J. H. Forrest and J. A. Tomlin (1991), 'Implementing the simplex method for the Optimization Subroutine Library', Research Report RJ 8174, IBM Almaden Research Centre, San Jose, California.

[17] R. Fourer and S. Mehrotra (1990), 'Performance of an augmented system approach for solving least-squares problems in an interior-point method for linear programming', Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

[18] K. R. Frisch (1955), 'The logarithmic potential method of convex programming', Report, University Institute of Economics, Oslo, Norway.

[19] J. A. George and J. W. H. Liu (1981), *Computer Solution of Large Sparse Positive Definite Systems*, Prentice-Hall (Englewood Cliffs, NJ).

[20] P. E. Gill, W. Murray, D. B. Ponceleón, and M. A. Saunders (1990), 'Preconditioners for indefinite systems arising in optimization', Report SOL 90-8, Department of Operations Research, Stanford University, Stanford, California.

[21] 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.

[22] P. E. Gill, W. Murray, M. A. Saunders, J. A. Tomlin and M. H. Wright (1986), 'On projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method', *Math. Program.* **36**, 183–209.

[23] P. E. Gill, W. Murray and M. H. Wright (1981), *Practical Optimization*, Academic Press (London and New York).

[24] D. Goldfarb and M. J. Todd (1989), 'Linear programming', in *Optimization* (G. L. Nemhauser, A. H. G. Rinnooy Kan and M. J. Todd, eds.), North Holland (Amsterdam and New York), 73–170.

[25] C. C. Gonzaga (1992), 'Path following methods for linear programming', *SIAM Review* **34**, to appear.

[26] B. Grünbaum (1967), *Convex Polytopes*, John Wiley and Sons (London).

[27] N. J. Higham (1990), 'Analysis of the Cholesky decomposition of a semi-definite matrix', in *Reliable Numerical Computation* (M. G. Cox and S. Hammarling, eds.), Clarendon Press (Oxford), 161–185.

[28] P. Huard (1967), 'Resolution of mathematical programming with nonlinear constraints by the method of centres', in *Nonlinear Programming* (J. Abadie, ed.), North Holland (Amsterdam and New York), 207–219.

[29] F. Jarre (1991), 'Interior-point methods for convex programming', Report SOL 90-16, Department of Operations Research, Stanford University, Stanford, California.

[30] K. Jittorntrum (1978), *Sequential Algorithms in Nonlinear Programming*, Ph. D. thesis, Australian National University.

[31] N. K. Karmarkar (1984), 'A new polynomial time algorithm for linear programming', *Combinatorica* **4**, 373–395.

[32] N. K. Karmarkar (1990), 'Riemannian geometry underlying interior-point methods for linear programming', in *Mathematical Developments Arising from Linear Programming* (J. C. Lagarias and M. J. Todd, eds.), American Mathematical Society (Providence, RI), 51–75.

[33] L. G. Khachian (1979), 'A polynomial algorithm in linear programming', *Doklady Akademiia Nauk SSSR* **244**, 1093–1096 (in Russian); English translation in *Sov. Math. Dokl.* **20**, 191–194.

[34] V. Klee and G. J. Minty (1972), 'How good is the simplex algorithm?', in *Inequalities III* (O. Shisha, ed.), Academic Press (New York), 159–175.

[35] E. Kranich (1991), 'Interior point methods for mathematical programming: a bibliography', Report 171, Universität Hagen, Hagen, Germany. This bibliography can be accessed electronically by sending email to 'netlib@research.att.com' with message 'send intbib.tex from bib' and/or 'send intbib.bbl from bib'.

[36] D. G. Luenberger (1984), *Introduction to Linear and Nonlinear Programming*, Addison-Wesley (Menlo Park, CA).

[37] I. Lustig, R. E. Marsten and D. F. Shanno (1990), 'On implementing Mehrotra's predictor–corrector interior point method for linear programming', Report SOR 90-03, Department of Civil Engineering and Operations Research, Princeton University, Princeton, New Jersey.

[38] A. Marxen (1989), 'Primal barrier methods for linear programming', Ph. D. thesis, Stanford University, Stanford, California.

[39] 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.

[40] N. Megiddo (1987), 'Pathways to the optimal set in linear programming', in *Progress in Mathematical Programming* (N. Megiddo, ed.), Springer-Verlag (New York), 131–158.

[41] S. Mehrotra (1990), 'On the implementation of a (primal-dual) interior method', Report 90-03, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

[42] S. Mehrotra and J. Sun (1990), '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.), American Mathematical Society (Providence, RI), 265–284.

[43] R. D. C. Monteiro and I. Adler (1989a), 'Interior path following primal-dual algorithms, Part I: Linear programming', *Math. Program.* **44**, 27–41.

[44] R. D. C. Monteiro and I. Adler (1989b), 'Interior path following primal-dual algorithms, Part II: Convex quadratic programming', *Math. Program.* **44**, 43–66.

[45] J. J. Moré and S. J. Wright (1990), private communication.

[46] 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.

[47] W. Murray and M. H. Wright (1991), 'Line search procedures for the logarithmic barrier function', Manuscript, AT&T Bell Laboratories, Murray Hill, New Jersey.

[48] Y. Nesterov and A. Nemirovsky (1989), *Self-Concordant Functions and Polynomial-Time Methods in Convex Programming*, USSR Academy of Science (Moscow).

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

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

[51] D. B. Ponceleón (1990), *Barrier Methods for Large-scale Quadratic Programming*, Ph. D. thesis, Stanford University, Stanford, California.

[52] M. J. D. Powell (1972), 'Problems related to unconstrained optimization', in *Numerical Methods for Unconstrained Optimization* (W. Murray, ed.), Academic Press (London and New York), 29–55.

[53] M. J. D. Powell (1990), 'Karmarkar's algorithm: a view from nonlinear programming', *IMA Bulletin* **26**, 165–181.

[54] J. Renegar (1988), 'A polynomial-time algorithm based on Newton's method for linear programming', *Math. Program.* **40**, 59–94.

[55] R. T. Rockafellar (1970), *Convex Analysis*, Princeton University Press (Princeton, NJ).

[56] C. Roos and J.-Ph. Vial (1988), 'A polynomial method of approximate centres for linear programming', Report 88-68, Faculty of Mathematics and Informatics, Delft University of Technology, Delft, Holland. To appear in *Math. Program.*

[57] G. Sonnevend (1986), 'An analytic centre for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming', in *Lecture Notes in Control and Information Science*, Vol. 84, Springer-Verlag (New York), 866–876.

[58] R. J. Vanderbei (1991), 'Symmetric quasi-definite matrices', Report 91-10, Department of Civil Engineering and Operations Research, Princeton University, Princeton, New Jersey.

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

[60] M. H. Wright (1991), 'Determining subspace information from the Hessian of a barrier function', Manuscript, AT&T Bell Laboratories, Murray Hill, New Jersey.