SIAM REVIEW
Vol. 44, No. 4, pp. 525–597

Interior Methods for Nonlinear Optimization*

Anders Forsgren†
Philip E. Gill‡
Margaret H. Wright§

Abstract. Interior methods are an omnipresent, conspicuous feature of the constrained optimization landscape today, but it was not always so. Primarily in the form of barrier methods, interior-point techniques were popular during the 1960s for solving nonlinearly constrained problems. However, their use for linear programming was not even contemplated because of the total dominance of the simplex method. Vague but continuing anxiety about barrier methods eventually led to their abandonment in favor of newly emerging, apparently more efficient alternatives such as augmented Lagrangian and sequential quadratic programming methods. By the early 1980s, barrier methods were almost without exception regarded as a closed chapter in the history of optimization.

This picture changed dramatically with Karmarkar's widely publicized announcement in 1984 of a fast polynomial-time interior method for linear programming; in 1985, a formal connection was established between his method and classical barrier methods. Since then, interior methods have advanced so far, so fast, that their influence has transformed both the theory and practice of constrained optimization. This article provides a condensed, selective look at classical material and recent research about interior methods for nonlinearly constrained optimization.

Key words. nonlinear programming, constrained minimization, nonlinear constraints, primal-dual methods, interior methods, penalty methods, barrier methods

AMS subject classifications. 49J20, 49J15, 49M37, 49D37, 65F05, 65K05, 90C30

PII. S0036144502414942

1. Introduction.

It is a truth universally acknowledged that the field of continuous optimization has undergone a dramatic change since 1984. This change, sometimes described as the "interior-point revolution," has featured a continual blending of old and new, with effects far beyond optimization. An especially appealing aspect of the interior-point revolution is its spirit of unification, which has brought together areas of optimization that for many years were treated as firmly disjoint. Prior to 1984, linear and nonlinear programming, one a subset of the other, had evolved for the

*Received by the editors September 22, 2002; accepted for publication (in revised form) October 1, 2002; published electronically October 30, 2002.
http://www.siam.org/journals/sirev/44-4/41494.html

†Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden (anders.forsgren@math.kth.se). The research of this author was supported by the Swedish Research Council (VR).

‡Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112 (pgill@ucsd.edu). The research of this author was supported by National Science Foundation grants DMS-0208449 and ACI-0082100.

§Department of Computer Science, Courant Institute, New York University, New York, NY 10012 (mhw@cs.nyu.edu).

most part along unconnected paths, without even a common terminology. (The use of "programming" to mean "optimization" serves as a persistent reminder of these differences.) Today this separation seems, as it indeed was, artificial, yet it was a fully accepted part of the culture of optimization not so many years ago.

1.1. Roots in Linear and Nonlinear Programming. Although the focus of this article is on nonlinearly constrained problems, understanding the context of the interior-point revolution requires a short digression on linear programming (minimization of a linear function subject to linear constraints). A fundamental property of well-behaved n-variable linear programs with m inequality constraints is that a *vertex* minimizer must exist, i.e., a point where n constraints with linearly independent gradients hold with equality. (See, e.g., [20, 92] for details about linear programming.) The simplex method, invented by Dantzig in 1947, is an iterative procedure that solves linear programs by exploiting this property. A simplex iteration moves from vertex to vertex, changing (one at a time) the set of constraints that hold exactly, decreasing the objective as it goes, until an optimal vertex is found. From the very start, the simplex method dominated the field of linear programming. Although "nonsimplex" strategies for linear programming were suggested and tried from time to time, they could not consistently match the simplex method in overall speed and reliability. Furthermore, a simplex-centric world view had the effect that even "new" techniques mimicked the motivation of the simplex method by always staying on a subset of exactly satisfied constraints.

The preeminence of the simplex method was challenged not because of failures in practice—the simplex method was, and is, used routinely to solve enormous linear programs—but by worries about its computational complexity. One can argue that the simplex method and its progeny are inherently *combinatorial*, in that their performance seems to be bound in the worst case to the maximum number of ways in which n out of m constraints can hold with equality. In fact, with standard pivoting rules specifying the constraint to be dropped and added at each iteration, the simplex method can visit *every vertex* of the feasible region [64]; thus its worst-case complexity is *exponential* in the problem dimension. As a result, there was great interest in finding a polynomial-time linear programming algorithm.[1]

The first success in this direction was achieved in 1979 by Khachian, whose ellipsoid method was derived from approaches proposed originally for nonlinear optimization. (See [92] for details about Khachian's method.) Despite its polynomial complexity bound, however, the ellipsoid method performed poorly in practice compared to the simplex method, and the search continued for a polynomial-time linear programming method that was genuinely fast in running time.

The start of the interior-point revolution was Karmarkar's announcement [63] in 1984 of a polynomial-time linear programming method that was 50 times faster than the simplex method. Amid the frenzy of interest in Karmarkar's method, it was shown in 1985 [51] that there was a formal equivalence between Karmarkar's method and the classical logarithmic barrier method (see sections 1.2 and 3) applied to linear programming, and long-discarded barrier methods were soon rejuvenated as polynomial-time algorithms for linear programming. Furthermore, barrier methods (unlike the simplex method) could be applied not only to linear programming

[1] Assuming various distributions of random inputs, [11, 94] showed that the simplex method converges in *expected* polynomial time. The recent development of "smoothed" complexity analysis [95] has led to new insights about the average behavior of the simplex method.

but also to other optimization problems, such as quadratic programming, linear and nonlinear complementarity, and nonlinear programming. Although the ties between Karmarkar's method and barrier methods were controversial for a few years, these disagreements have (mostly) faded away. Pedagogical and philosophical issues remain about the best way to motivate interior-point methods—perturbing optimality conditions? minimizing a barrier function?—and the multiplicity of viewpoints continues to create new insights and new algorithms.

The interior-point revolution has led to a fundamental shift in thinking about continuous optimization. Linear and nonlinear programming are seen as related parts of a unified whole, and no one would seriously claim today that linear programming is completely different from nonlinear optimization. (Of course, methods for solving linear programs and nonlinear problems vary significantly in detail.)

As we shall see, the signature of interior methods is the existence of continuously parameterized families of approximate solutions that asymptotically converge to the exact solution. These paths trace smooth trajectories with algebraic and geometric properties (such as being "centered" in a precisely defined sense) that can be analyzed and exploited algorithmically. Many interior methods are characterized as "path-following" to signal their dependence on properties of these paths, which provide the foundation for all complexity analyses of interior-point algorithms for linear, quadratic, and convex programming.

The monumental work [79] of Nesterov and Nemirovskii proposed new families of barrier methods and extended polynomial-time complexity results to new convex optimization problems. *Semidefinite programming*—minimization of a convex function in the space of symmetric matrices subject to semidefiniteness constraints—is arguably the most notable of these problems to receive widespread attention as a direct result of the development of interior methods (see, e.g., the surveys [65, 99, 101]). The evident similarity of interior methods to longstanding continuation approaches (see, e.g., [1, 2]) has been recognized since the early days of modern interior methods (see, e.g., [71]), but numerous aspects remain to be explored.

As a remarkable bonus, interior methods are playing a growing role in the study of hard combinatorial problems. Many of the most important problems in discrete optimization (where the variables are required to be integers) are NP-hard, i.e., they cannot be solved in polynomial time unless someone favorably resolves the still-open question of whether $P = NP$. In the meantime, good approximate solutions are being found by *approximation algorithms*—polynomial-time algorithms whose solution is provably within a certain factor of the optimal solution for the hard problem. A main ingredient in a successful approximation algorithm is formulation of a convex relaxation (often a semidefinite program) in which integrality constraints are replaced by definiteness constraints on associated matrices. An exceptionally clear introduction to this subject is given in [108].

Almost twenty years after the beginning of the interior-point revolution, there seems to be no end in sight to new applications of interior methods and new interpretations of the interior-point perspective.

1.2. Classical Barrier Methods. As we have just sketched, classical barrier methods are closely related to modern interior methods, and we briefly summarize their history. During the 1960s, the accepted way to solve constrained problems was to transform them into parameterized *unconstrained* problems via penalty or barrier terms. For inequality constraints, a *barrier* method is motivated by unconstrained

minimization of a function combining f and a positively weighted "barrier" that prevents iterates from leaving the feasible region. *Penalty* methods, in contrast, are based on minimizing a function that includes f and a positive penalty if evaluated at any infeasible point.

A large body of beautiful mathematical theory about barrier and penalty functions was developed during the 1960s by Fiacco and McCormick. They also seem to have introduced the term "interior-point methods" in their seminal book [33, p. 41], which describes in detail the relationships between minimizers of barrier and penalty function and solutions of the original constrained problem.

Despite the good features of barrier methods, they were dogged by several concerns. The worry expressed most often in print involved ill-conditioning, after Lootsma [66] and Murray [73] showed independently in the late 1960s that in general the Hessian of a barrier function becomes increasingly ill-conditioned as the solution is approached and is singular in the limit. Increasing awareness of this property led to serious anxiety about the reliability of barrier methods just as other methods were coming along that seemed to be more efficient in practice without being plagued by unavoidable ill-conditioning. In particular, augmented Lagrangian and sequential quadratic programming (SQP) methods (see, for example, [6, 34, 52, 77, 80]) are based directly on the optimality conditions for constrained optimization. Barrier methods appeared distinctly unappealing by comparison, and almost all researchers in mainstream optimization lost interest in them.

As described in section 1.1, the dormancy of barrier methods ended in high drama near the start of the interior-point revolution. An obvious question then needed to be answered: Are classical barrier methods fundamentally flawed, as once feared? The answer turns out to be "yes," but, surprisingly, not because of ill-conditioning. Classical barrier methods are indeed inefficient—but, by a strange twist of fate, ill-conditioning, their longtime bugbear, has recently been shown not to be harmful under circumstances that almost always hold in practice. We explore several interesting properties, good and bad, of the classical Newton barrier method in section 4.3. An obvious strategy has been to create interior methods that retain the good properties of classical barrier methods, yet do not suffer from their defects. The general opinion today is that *primal-dual* methods, to be discussed in section 5, offer the greatest promise for achieving these ends.

It is impossible to cover interior methods for nonlinear optimization thoroughly in anything less than a large volume. A major goal of this article is thus to show connections between classical and modern ideas and to cover highlights of both theory and practice; readers interested in learning more about interior-point methods will find an abundance of papers and books on the subject. Since linear algebra is a special interest of the authors, we have devoted extra attention to linear algebraic issues associated with interior methods. The linear algebra needs of interior methods are interesting for several reasons. Certain key matrices display increasing ill-conditioning as the solution is approached, but the ill-conditioning is highly structured. In contrast to active-set methods like the simplex method that continually update a set of constraints temporarily treated as equalities, interior methods typically include all constraints at every iteration. Hence the matrices arising in interior methods must somehow reveal, without omitting any constraints, that some constraints are more important than others. Similarly, two subspaces—the range space of the transposed Jacobian of the active constraints and the associated null space—strongly affect all calculations near the solution, but these subspaces are not known explicitly.

1.3. Statement of the Problem. We concentrate on interior methods for continuous nonlinear optimization problems of the following form:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & c_i(x) = 0, \quad i \in \mathcal{E}, \\ & c_i(x) \geq 0, \quad i \in \mathcal{I}, \end{array} \tag{1.1}$$

where $c(x)$ is an m-vector of nonlinear constraint functions with ith component $c_i(x)$, $i = 1, \ldots, m$, and $\mathcal{E}$ and $\mathcal{I}$ are nonintersecting index sets. It is assumed throughout that f and c are twice-continuously differentiable. Any point x satisfying the constraints of (1.1) is called a *feasible point*, and the set of all such points is the *feasible region*. We first consider problems containing only inequality constraints (sections 2 through 5) and then turn in sections 6 and 7 to the general form (1.1).

1.4. A Few Words on Coverage and Notation. Since thousands of scientific papers have been written about interior methods, as already noted we cannot cover more than a tiny fraction of the field, and it would be equally impractical to cite all relevant references. We apologize in advance to all those whose favorite topics or works have not been mentioned here.

Because this is a survey intended for nonexperts, we have included a substantial amount of background material on optimality conditions in section 2. Readers already familiar with optimization should skip directly to section 3. Various useful definitions, lemmas, and miscellaneous results are collected in the appendix.

Finally, because there are not enough letters in the alphabet, especially letters that are free from previous connotations, we confess to straining at times to find notation that is clear and precise without being cluttered. To alleviate this dilemma, we sometimes introduce local abbreviations for the sake of short formulas. For example, when considering a particular point, say x^*, we will sometimes abbreviate quantities evaluated at x^* by adding a superscript "$*$" and omitting the argument, e.g., we denote $c(x^*)$ by c^*. Following common usage in the interior-point literature, if a vector is denoted by a lowercase letter, the same uppercase letter denotes the diagonal matrix whose elements are those of the vector, so that $V \triangleq \text{diag}(v)$. Finally, e denotes the vector of all ones whose dimension is determined by the context.

2. Inequality-Constrained Optimization. We begin with problems containing only inequality constraints:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c(x) \geq 0, \tag{2.1}$$

where $c(x)$ is an m-vector of functions $\{c_i(x)\}$, $i = 1, \ldots, m$, and we assume throughout that f and $\{c_i\}$ are twice-continuously differentiable. The gradient of f is denoted by either $\nabla f(x)$ or $g(x)$, and $\nabla^2 f(x)$ denotes the Hessian matrix of second partial derivatives of f. The gradient and Hessian of $c_i(x)$ are denoted by $\nabla c_i(x)$ and $\nabla^2 c_i(x)$. The $m \times n$ Jacobian matrix $c'(x)$ of first derivatives of $c(x)$ has rows $\{\nabla c_i(x)^T\}$, and we sometimes (to avoid clutter) use $J(x)$ to denote this Jacobian.

The topic of optimality conditions for nonlinearly constrained optimization can be complicated and confusing. We present only aspects that will be needed later; detailed discussions may be found in, for example, [6, 88].

2.1. The KKT Conditions. The terms "KKT point" (standing for "Karush–Kuhn–Tucker point") and "KKT conditions" will be used often. In defining these

terms, we use an analogue of the MATLAB componentwise multiplication operator, denoted by "$\cdot$".

DEFINITION 2.1 (the componentwise multiplication operator $\cdot$). *Given two vectors x and y of dimension r, $x \cdot y$ is an r-vector whose ith component is $x_i y_i$.*

DEFINITION 2.2 (feasible region). *Given the constraints $c(x) \geq 0$, the* feasible region *is $\mathcal{F} \triangleq \{x \in \mathbb{R}^n : c(x) \geq 0\}$.*

DEFINITION 2.3 (first-order KKT point). *The* first-order KKT conditions *for the inequality-constrained problem* (2.1) *hold at the point x^*, or, equivalently, x^* is a (first-order)* KKT point*, if there exists an m-vector λ^*, called a* Lagrange multiplier vector*, such that*

$$c(x^*) \geq 0 \quad \text{(feasibility)}, \tag{2.2a}$$

$$g(x^*) = J(x^*)^T \lambda^* \quad \text{(stationarity)}, \tag{2.2b}$$

$$\lambda^* \geq 0 \quad \text{(nonnegativity of the multipliers)}, \quad \textit{and} \tag{2.2c}$$

$$c(x^*) \cdot \lambda^* = 0 \quad \text{(complementarity)}. \tag{2.2d}$$

In the operations research literature, it is common to refer to x as the "primal variables" and to the Lagrange multipliers as the "dual variables." This usage originated with linear programming, where close relationships always hold between the primal and dual variables, and between the original (primal) problem and a closely related (dual) problem. (See the footnote in the proof of Lemma 2.7.) In nonlinear optimization, these connections are local and sometimes tenuous, which is why the divergence of terminology remains. As we shall see later in section 5, however, interior methods involve a mixture of terminology, reflecting their dual origins in linear programming and nonlinear optimization.

The stationarity condition (2.2b) can be written as

$$\nabla_x L(x^*, \lambda^*) = 0, \quad \text{where} \quad L(x, \lambda) \triangleq f(x) - \lambda^T c(x). \tag{2.3}$$

Thus a KKT point is a stationary point with respect to x of the *Lagrangian function* $L(x, \lambda)$ defined in (2.3). (Note that some authors formulate the constraints as $c(x) \leq 0$ and write the Lagrangian function with a "+" sign; in either case, the multipliers are nonnegative.)

The first-order KKT conditions specify only that *some* suitable multiplier exists at x^*. To describe how the nature of the constraints at x^* affects the definition of this multiplier, we need a further definition.

DEFINITION 2.4 (active, inactive, and violated constraints). *For the set of constraints $c(x) \geq 0$, the ith constraint is said to be* active *at a point $\bar{x}$ if $c_i(\bar{x}) = 0$ and* inactive *if $c_i(\bar{x}) > 0$. The* active set *$\mathcal{A}(\bar{x})$ is the set of indices of the active constraints at $\bar{x}$, i.e., $\mathcal{A}(\bar{x}) = \{i : c_i(\bar{x}) = 0\}$; the argument of $\mathcal{A}$ is omitted when it is obvious. The constraint $c_i(x) \geq 0$ is said to be* violated *at $\bar{x}$ if $c_i(\bar{x}) < 0$.*

To satisfy the complementarity condition $c(x^*) \cdot \lambda^* = 0$ (2.2d), the components of λ^* associated with inactive constraints are necessarily zero, which means that the gradient of f at a KKT point x^* must be a linear combination of the *active* constraint gradients:

$$g(x^*) = J_{\mathcal{A}}(x^*)^T \lambda^*_{\mathcal{A}}, \tag{2.4}$$

where $J_{\mathcal{A}}$ denotes the Jacobian of the active constraints and $\lambda^*_{\mathcal{A}}$ the vector of multipliers for the active constraints. Depending on the nature of x^*, there may be an infinite

number of multipliers satisfying (2.4). The set of acceptable multipliers, which we now define, will be of interest throughout.

DEFINITION 2.5 (acceptable Lagrange multipliers). *Given a KKT point x^* for problem* (2.1)*, the set of acceptable multipliers is defined as*

$$\mathcal{M}_\lambda(x^*) \triangleq \{\lambda \in \mathbb{R}^m : g(x^*) = J(x^*)^T\lambda, \ \ \lambda \geq 0, \ \ \textit{and} \ \ c(x^*) \cdot \lambda = 0\}. \tag{2.5}$$

The complementarity condition $c(x^*) \cdot \lambda = 0$ forces λ_i to be zero if constraint i is inactive but allows the possibility that $\lambda_i = 0$ when constraint i is active. The property of *strict complementarity*, which occurs when all the multipliers for active constraints are positive, is so important that we define it formally.

DEFINITION 2.6 (strict complementarity). Strict complementarity *holds at the KKT point x^* if there is a multiplier $\lambda^* \in \mathcal{M}_\lambda$ such that $\lambda_i^* > 0$ for all $i \in \mathcal{A}$.*

A constraint $c_i(x) \geq 0$ is said to be *strongly active* at a KKT point x^* if $i \in \mathcal{A}(x^*)$ and there exists at least one $\lambda^* \in \mathcal{M}_\lambda$ with $\lambda_i^* > 0$. Similarly, the ith constraint is said to be *weakly active* at x^* if $i \in \mathcal{A}(x^*)$ and $\lambda_i = 0$ for all $\lambda \in \mathcal{M}_\lambda$; in this case, we say that constraint i has a *null multiplier*, by analogy with the term "null variable" in linear programming. In the simple problem

$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \quad \tfrac{1}{2}(x_1^2 + x_2^2) \ \text{ subject to } \ x_1 + x_2 \geq 4 \ \text{ and } \ x_2 \geq 2, \tag{2.6}$$

portrayed in Figure 4, the unique solution is $x^* = (2, 2)$ and the (unique) multiplier is $\lambda^* = (2, 0)^T$. Hence the second constraint has a null multiplier.

The next two results summarize the implications of a null multiplier and strict complementarity (when there are no null multipliers).

LEMMA 2.7 (implication of a null multiplier). *Consider the problem of minimizing $f(x)$ subject to $c(x) \geq 0$. If x^* is a KKT point, the active constraint i has a null multiplier at x^* if and only if there exists $p \in \mathbb{R}^n$ such that $J_{\mathcal{A}}^* p \geq 0$, $g^{*T} p = 0$, and $\nabla c_i^{*T} p > 0$, where $J_{\mathcal{A}}^*$ denotes $J_{\mathcal{A}}(x^*)$, g^* denotes $g(x^*)$, and ∇c_i^* denotes the gradient of constraint i at x^*.*

Proof. If the ith active constraint has a null multiplier at x^*, consider the linear program of maximizing $e_i^T\lambda$ subject to $J_{\mathcal{A}}^{*T}\lambda = g^*$ and $\lambda \geq 0$, where e_i is the ith coordinate vector. This linear program is feasible, meaning that there is a vector λ satisfying the constraints, because x^* is a KKT point. By definition of a null multiplier, $\lambda_i = 0$ for all feasible λ, so that this linear program has a bounded optimal objective value of zero. It follows from standard duality theory for linear programming[2] that its dual, minimizing $g^{*T} p$ subject to $J_{\mathcal{A}}^* p \geq e_i$, is also feasible, with a bounded optimal objective value of zero, which gives the desired result. ∎

LEMMA 2.8 (implication 1 of strict complementarity). *Consider the problem of minimizing $f(x)$ subject to $c(x) \geq 0$. If x^* is a KKT point, strict complementarity holds at x^* if and only if every vector p satisfying $J_{\mathcal{A}}^* p \geq 0$ and $g^{*T} p = 0$ also satisfies $J_{\mathcal{A}}^* p = 0$.*

Proof. This result follows immediately from Lemma 2.7. ∎

[2] A tiny summary of the relevant linear programming duality results: for the primal linear program of minimizing $c^T x$ subject to $Ax \geq b$, the dual linear program involves maximizing $b^T\lambda$ subject to $A^T\lambda = c$ and $\lambda \geq 0$; the dual of the dual is the primal. If the primal linear program is feasible (i.e., if there exists x such that $Ax \geq b$) and has a bounded objective value, the same applies for the dual, and vice versa; in this case, the optimal primal and dual objective values are equal. If the primal linear program is feasible but the primal objective is unbounded below, then the dual must be *infeasible*; i.e., there is no λ such that $A^T\lambda = c$ and $\lambda \geq 0$.

2.2. Constraint Qualifications. For problems with linear constraints, the first-order KKT conditions are *necessary* for optimality. Unfortunately, this tidy characterization does not carry over when nonlinear constraints are present. To specify first-order necessary conditions for optimality with nonlinear constraints, we require that the constraints satisfy certain *regularity conditions*, or *constraint qualifications*, at x^*. If these regularity conditions do not hold, a prospective solution x^* may or may not be a KKT point.

Constraint qualifications allow us to deduce information about feasible points in a neighborhood of x^* from local linearizations of the constraints. This is why constraint qualifications always hold when all the problem constraints are linear.

Here we consider the two constraint qualifications most often used in nonlinear optimization.

DEFINITION 2.9 (linear independence constraint qualification). *Consider an inequality-constrained problem with constraints* $c(x) \geq 0$. *The* linear independence constraint qualification (LICQ) *holds at the feasible point* $\bar{x}$ *if* $\bar{x}$ *is strictly feasible (so that there are no active constraints) or if the Jacobian of the active constraints at* $\bar{x}$ *has full row rank, i.e., if the gradients of the active constraints are linearly independent.*

Given a feasible point $\bar{x}$, satisfaction of the LICQ at $\bar{x}$ implies that $\bar{J}_{\mathcal{A}}$ has linearly independent rows, where $\bar{J}_{\mathcal{A}}$ denotes $J_{\mathcal{A}}(\bar{x})$. If $\bar{x}$ is a KKT point, it follows from (2.4) that the linear system $\bar{g} = \bar{J}_{\mathcal{A}}^T \bar{\lambda}$ must be compatible, where $\bar{g}$ denotes $g(\bar{x})$. Full column rank of $\bar{J}_{\mathcal{A}}^T$ then means that $\bar{\lambda}$ is unique, so that $\mathcal{M}_\lambda(\bar{x})$ contains a single multiplier when the LICQ holds at a KKT point.

The second constraint qualification to be considered, the *Mangasarian–Fromovitz constraint qualification* [67], ensures the existence of a direction along which, to first order, all active constraints strictly increase.

DEFINITION 2.10 (Mangasarian–Fromovitz constraint qualification). *Consider an inequality-constrained problem with constraints* $c(x) \geq 0$. *The* Mangasarian–Fromovitz constraint qualification (MFCQ) *holds at the feasible point* $\bar{x}$ *if* $\bar{x}$ *is strictly feasible or if there exists a vector* p *such that* $\nabla c_i(\bar{x})^T p > 0$ *for all* $i \in \mathcal{A}(\bar{x})$, *i.e., if* $J_{\mathcal{A}}(\bar{x})p > 0$.

The MFCQ is a *weaker* condition than the LICQ in the sense that satisfaction of the LICQ implies the MFCQ, but not the reverse. Furthermore, although satisfaction of the MFCQ ensures the existence of a path into the strict interior of the feasible region that is tangent to p at $\bar{x}$, a path into the strict interior of the feasible region can exist even when the MFCQ is not satisfied.

We give two examples to illustrate these situations. First, consider a problem in $\mathbb{R}^2$ with the four constraints

$$c_1(x) = x_1 \geq 0, \qquad c_2(x) = x_2 \geq 0, \tag{2.7a}$$

$$c_3(x) = 4 - x_1^2 - 4x_2^2 \geq 0, \quad \text{and} \tag{2.7b}$$

$$c_4(x) = 5 - (x_1 - 2)^2 - x_2^2 \geq 0, \tag{2.7c}$$

which are depicted on the left in Figure 1, with the darkest shading in the feasible region. At $\bar{x} = (0, 1)^T$, the active set is $\mathcal{A}(\bar{x}) = \{1, 3, 4\}$, and the Jacobian of the active constraints does not have full row rank (so that the LICQ is not satisfied):

$$J_{\mathcal{A}}(\bar{x}) = \begin{pmatrix} 1 & 0 \\ -2x_1 & -8x_2 \\ -2(x_1 - 2) & -2x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -8 \\ 4 & -2 \end{pmatrix}.$$

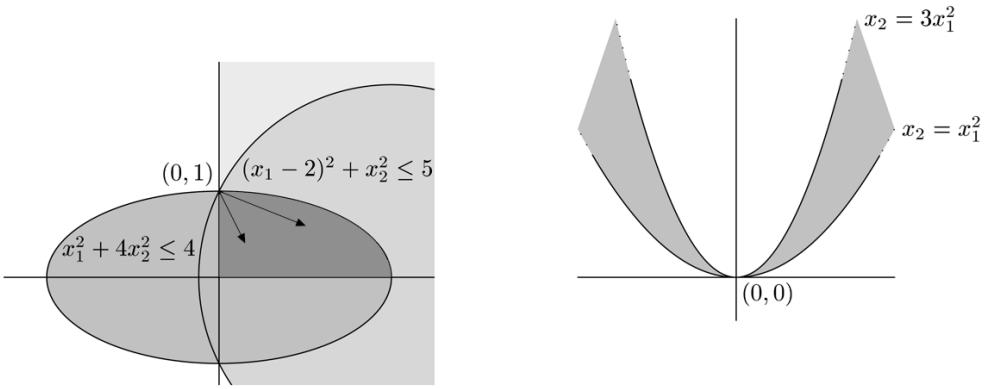

Fig. 1 *The left figure depicts the four constraints of* (2.7)*, with the darkest shading inside the feasible region. At the point* $(0,1)^T$*, constraints* 1*,* 3*, and* 4 *are active, and* $J_{\mathcal{A}}$ *is rank-deficient. Nonetheless, the MFCQ is satisfied, as can be seen by the vectors pointing into the strict interior of the feasible region. The right figure shows the two constraints of* (2.8)*, with the feasible region (between the two parabolas) shaded. At the origin, the MFCQ is not satisfied, since there are no directions pointing strictly into the feasible region. Yet every neighborhood of the origin contains strictly feasible points.*

Even so, the MFCQ holds, since (for example) $p = (1, -1)^T$ satisfies $J_{\mathcal{A}}(\bar{x})p > 0$. Two vectors pointing into the strictly feasible region are shown in the figure.

In contrast, consider the two constraints

$$\begin{aligned} c_1(x) &= -x_1^2 + x_2 \geq 0, \\ c_2(x) &= \ 3x_1^2 - x_2 \geq 0, \end{aligned} \tag{2.8}$$

which are depicted on the right in Figure 1. The first constraint is satisfied by points lying above or on the parabola $x_2 = x_1^2$; the second is satisfied by points below or on the parabola $x_2 = 3x_1^2$. If we take $\bar{x}$ as the origin, both constraints are active, with

$$J_{\mathcal{A}}(\bar{x}) = \begin{pmatrix} -2x_1 & 1 \\ 6x_1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}.$$

The MFCQ clearly does not hold, since $J_{\mathcal{A}}(\bar{x})p > 0$ would imply the impossible condition that $p_2 > 0$ and $-p_2 > 0$. Yet every neighborhood of the origin contains strictly feasible points, and there are infinitely many strictly feasible, but nonlinear, paths emanating from the origin.

An important consequence of the MFCQ is that its satisfaction at a first-order KKT point implies boundedness of the set of multipliers [45].

LEMMA 2.11 (implication 1 of MFCQ: a bounded multiplier set). *If* $\bar{x}$ *is a first-order KKT point at which the MFCQ is satisfied, then the set of multipliers* $\mathcal{M}_\lambda$ *defined in* (2.5) *is bounded.*

Proof. First we consider the nature of $\mathcal{M}_\lambda$ at $\bar{x}$, which consists of $\lambda \in \mathbb{R}^m$ satisfying $\bar{c} \cdot \lambda = 0$, $\bar{g} = \bar{J}^T\lambda$, and $\lambda \geq 0$, where $\bar{c}$ denotes $c(\bar{x})$ and $\bar{J}$ denotes $J(\bar{x})$. It is easy to see that $\mathcal{M}_\lambda$ is convex. Given any $\bar{\lambda} \in \mathcal{M}_\lambda$, $\mathcal{M}_\lambda$ can be unbounded only if there is a nonzero ray u emanating from $\bar{\lambda}$ such that $\bar{\lambda} + \alpha u \in \mathcal{M}_\lambda$ for all $\alpha \geq 0$. If such a ray exists, the complementarity condition will be satisfied only if components of u corresponding to inactive constraints are zero. Thus, in order for both $\bar{\lambda}$ and

$\bar{\lambda} + \alpha u$ to lie in $\mathcal{M}_\lambda$, it must be true that

$$\bar{g} = \bar{J}_{\mathcal{A}}^T \bar{\lambda}_{\mathcal{A}} = \bar{J}_{\mathcal{A}}^T (\bar{\lambda}_{\mathcal{A}} + \alpha u_{\mathcal{A}}),$$

where $\bar{\lambda}_{\mathcal{A}}$ and $u_{\mathcal{A}}$ denote the subvectors of $\bar{\lambda}$ and u corresponding to active constraints. It follows that $\bar{J}_{\mathcal{A}}^T u_{\mathcal{A}} = 0$. Finally, $\bar{\lambda}_{\mathcal{A}} + \alpha u_{\mathcal{A}}$ will remain nonnegative for arbitrarily large positive α only if $u_{\mathcal{A}} \geq 0$.

Turning now to the implications of the MFCQ, we know that a vector p exists such that $\bar{J}_{\mathcal{A}} p > 0$, which means that $\alpha \bar{J}_{\mathcal{A}} p > 0$ for any positive α. Thus for any positive θ there is a vector p satisfying $\bar{J}_{\mathcal{A}} p \geq \theta e$, where e denotes the vector of all ones. As a result, the linear program

$$\underset{p,\theta}{\text{minimize}} \; -\theta \quad \text{subject to} \quad \bar{J}_{\mathcal{A}} p - \theta e \geq 0, \;\; \theta \geq 0, \tag{2.9}$$

is feasible, but its objective function is unbounded below. Using standard duality theory for linear programming (summarized in the footnote to the proof of Lemma 2.7), unboundedness of the primal objective implies infeasibility of the dual. The constraints of the dual corresponding to (2.9) are $\bar{J}_{\mathcal{A}}^T u_{\mathcal{A}} = 0$, $e^T u_{\mathcal{A}} = 1$, and $u_{\mathcal{A}} \geq 0$, and so we know that there is no vector $u_{\mathcal{A}}$ satisfying these conditions. But, as shown in the first part of the proof, these are precisely the properties that $u_{\mathcal{A}}$ must have in order for $\mathcal{M}_\lambda$ to be unbounded. (The condition $e^T u_{\mathcal{A}} = 1$ is simply a scaling restriction to ensure that $u_{\mathcal{A}} \neq 0$.) Consequently no ray u exists, and $\mathcal{M}_\lambda$ is bounded. ∎

The next two lemmas summarize results in linear algebra associated with the MFCQ.

LEMMA 2.12 (implication 2 of MFCQ). *Given a matrix A and a vector p such that $Ap > 0$, the only nonnegative vector λ satisfying $A^T \lambda = 0$ is $\lambda = 0$.*

Proof. If $A^T \lambda = 0$, then $v^T A^T \lambda = 0$ for any vector v. Considering the specific vector p for which $Ap > 0$ and noting that λ is required to be nonnegative, $p^T A^T \lambda = (Ap)^T \lambda$ can be zero only if every component of λ is zero. ∎

LEMMA 2.13 (implication 3 of MFCQ and a nonzero multiplier). *Given a matrix A and a vector g such that the set*

$$\mathcal{M}_\lambda = \{\lambda : g = A^T \lambda, \;\; \lambda \geq 0\}$$

is nonempty, suppose that there is a vector p such that $Ap > 0$. If there is a nonzero vector $\widehat{\lambda} \in \mathcal{M}_\lambda$, then $\lambda = 0$ is not in $\mathcal{M}_\lambda$.

Proof. The proof is by contradiction. If $\lambda = 0$ is in $\mathcal{M}_\lambda$, then $g = 0$, which means that $A^T \lambda = 0$. Using Lemma 2.12, it follows that $\lambda = 0$ is the only point in $\mathcal{M}_\lambda$. But this means that $\mathcal{M}_\lambda$ cannot contain a nonzero multiplier. ∎

A practical disadvantage of the MFCQ compared to the LICQ is that verifying satisfaction of the MFCQ is more difficult—in fact, determining whether or not the MFCQ holds requires solving a linear program. The argument developed in the proof of Lemma 2.11 shows that the MFCQ holds at the KKT point $\bar{x}$ if the optimal solution of the linear program

$$\underset{p,\theta}{\text{minimize}} \; -\theta \quad \text{subject to} \quad \bar{J}_{\mathcal{A}} p - \theta e \geq 0, \;\; 0 \leq \theta \leq 1, \tag{2.10}$$

occurs at the maximum possible value of θ, namely, $\theta = 1$. Note that this linear program is feasible since its constraints are satisfied by $\theta = 0$ and $p = 0$.

2.3. Necessary Optimality Conditions. Before stating necessary optimality conditions, we need a precise definition of a local constrained minimizer.

DEFINITION 2.14 (local constrained minimizer). *The point x^* is a* local constrained minimizer *for the problem of minimizing $f(x)$ subject to $c(x) \geq 0$ if $x^* \in \mathcal{F}$ and there exists a compact set S such that*

$$x^* \in \operatorname{int}(S) \quad \textit{and} \quad f(x^*) = \min f(x) \quad \textit{for all } x \in S \cap \mathcal{F},$$

where $\operatorname{int}(S)$ *denotes the interior of the set S (see Definition* A.2*).*

Taking S as a closed neighborhood of x^* (see Definition A.1), we obtain the usual form of this definition.

Our main interest is in sufficient conditions for optimality, which we address in section 2.4. But first we show that, when a minimizer x^* satisfies the MFCQ, x^* must be a KKT point; see, for example, [33, p. 21]. The proof uses the ubiquitous, incredibly useful lemma of Farkas, which we state for completeness.

LEMMA 2.15 (Farkas' lemma). *Given an $r \times n$ matrix A and an n-vector b, $b^T p \geq 0$ for all p such that $Ap \geq 0$ if and only if $b = A^T y$ for some $y \geq 0$.* ∎

LEMMA 2.16 (first-order necessary conditions for a local constrained minimizer). *Consider the problem of minimizing $f(x)$ subject to $c(x) \geq 0$, and suppose that x^* is a local minimizer where the MFCQ holds. Then x^* must be a KKT point.*

Proof. Suppose that x^* is a minimizer where the MFCQ holds. By definition, x^* is feasible. If no constraints are active at x^*, x^* is a local *unconstrained* minimizer of $f(x)$, and hence $g(x^*) = 0$ (see Lemma A.7), implying that x^* is (trivially) a KKT point. If at least one constraint is active, satisfaction of the MFCQ implies existence of a vector $\widehat{p}$ such that $J_{\mathcal{A}}(x^*)\widehat{p} > 0$. Using the smoothness of c, we have

$$c_i(x^* + \alpha\widehat{p}) = c_i(x^*) + \alpha \nabla c_i(x^*)^T \widehat{p} + \alpha^2 O(\|\widehat{p}\|^2).$$

For all sufficiently small $\alpha > 0$ and $i \in \mathcal{A}(x^*)$, it follows that $c_i(x^* + \alpha\widehat{p}) > 0$. If, for this same vector $\widehat{p}$, $g(x^*)^T\widehat{p} < 0$, then smoothness of f implies that $f(x^* + \alpha\widehat{p}) < f(x^*)$ for all sufficiently small α. Thus, since every neighborhood of x^* contains feasible points with strictly smaller values of f, x^* violates Definition 2.14 and is not a local constrained minimizer.

It follows that x^* can be a local constrained minimizer only if $g(x^*)^T p \geq 0$ for every p satisfying $J_{\mathcal{A}}(x^*)p > 0$. Invoking Farkas' lemma, this implies that $g(x^*)$ is a nonnegative linear combination of the columns of $J_{\mathcal{A}}(x^*)^T$, say $g(x^*) = J_{\mathcal{A}}(x^*)^T \lambda_{\mathcal{A}}$ with $\lambda_{\mathcal{A}} \geq 0$. Taking $\lambda_i = 0$ if $i \neq \mathcal{A}(x^*)$, complementarity is satisfied, and it follows that x^* is a KKT point. ∎

First-order optimality conditions such as the KKT conditions (2.2a)–(2.2d) are not enough to ensure optimality unless problem (2.1) is a convex program (Definition A.8). (For discussions of convex programming and convex analysis, see, for example, [12, 89].) Consequently, it is common to consider *second-order optimality conditions* as well. In this regard, the Hessian of the Lagrangian $L(x, \lambda)$ (2.3) with respect to x will be crucial:

$$H(x, \lambda) \triangleq \nabla^2_{xx} L(x, \lambda) = \nabla^2 f(x) - \sum_{i=1}^{m} \lambda_i \nabla^2 c_i(x). \tag{2.11}$$

Second-order conditions typically involve curvature of the Lagrangian function along directions that belong to a subspace or a cone, meaning that we are interested in quantities of the form $p^T H(x, \lambda) p$ for some direction p; see Definition A.13 for the definition of "curvature" in this context.

We now give a second-order necessary optimality condition which states that if the LICQ holds, a local minimizer must be a KKT point where the Hessian of

the Lagrangian has nonnegative curvature when projected into the null space of the Jacobian of the active constraints.

LEMMA 2.17 (second-order necessary conditions). *Consider the problem of minimizing $f(x)$ subject to $c(x) \geq 0$, and suppose that x^* is a local minimizer at which the LICQ holds. Then there is a vector λ^* that satisfies $\lambda^* \geq 0$, $c^* \cdot \lambda^* = 0$, and $g^* = J^{*T}\lambda^*$, and, further, it must be true that*

$$p^T H(x^*, \lambda^*)p \geq 0 \quad \text{for all } p \text{ satisfying } J_{\mathcal{A}}^* p = 0. \tag{2.12}$$

Let $N_{\mathcal{A}}$ denote a matrix whose columns form a basis for the null space of $J_{\mathcal{A}}^*$. The curvature condition (2.12) of Lemma 2.17 may also be expressed in terms of the matrix $N_{\mathcal{A}}^T H(x^*, \lambda^*)N_{\mathcal{A}}$, which is called the *reduced Hessian* of the Lagrangian. Condition (2.12) is equivalent to a requirement that $N_{\mathcal{A}}^T H(x^*, \lambda^*)N_{\mathcal{A}}$ be positive semidefinite.

2.4. Sufficient Optimality Conditions. During the past twenty years, a heightened awareness has developed of the importance of conditions that ensure an *isolated* local minimizer—a minimizer with the property that it lies in a neighborhood containing no other minimizers. Being an isolated minimizer is a stronger property than being a *strict* local minimizer—a point whose function value is strictly greater than at all points in a neighborhood. To understand the somewhat counterintuitive idea that a point can be a strict minimizer without being isolated, consider minimizing (without constraints) the function $f(x) = x^4\big(2 + \cos(1/x)\big)$, with $f(0) = 0$ (see, e.g., [34, p. 13] or [88, p. 206]). The origin is obviously a strict local minimizer: since $2 + \cos(1/x) > 0$ and $x^4 > 0$ if $x \neq 0$, it follows that $f(x) > 0$ when $x \neq 0$. Yet the infinitely many oscillations in $\cos(1/x)$ mean that every neighborhood of the origin contains infinitely many local minimizers.

DEFINITION 2.18 (isolated constrained minimizer). *A local constrained minimizer x^* is* isolated *if there is a neighborhood of x^* containing no other local constrained minimizers.*

To guarantee that a point is an isolated local constrained minimizer, we need conditions that involve second derivatives. Several forms of *second-order sufficient conditions* exist in the literature. We discuss these conditions in some detail to alert readers to the easy-to-miss nuances that may occur in their statement.

A common second-order condition applies to a KKT point x^* and a *specific* Lagrange multiplier λ^*. Let g^*, c^*, and $J_{\mathcal{A}}^*$ denote $g(x^*)$, $c(x^*)$, and $J_{\mathcal{A}}(x^*)$, respectively.

DEFINITION 2.19 (first second-order sufficient condition (SSC1)). *Consider the problem of minimizing $f(x)$ subject to $c(x) \geq 0$. Let x^* denote a KKT point, so that $c^* \geq 0$, and let the vector λ^* satisfy $\lambda^* \geq 0$, $c^* \cdot \lambda^* = 0$, and $g^* = J^{*T}\lambda^*$. We say that SSC1 holds for (x^*, λ^*) if there exists $\omega > 0$ such that*

$$p^T H(x^*, \lambda^*)p \geq \omega\|p\|^2 \text{ for all nonzero } p \text{ satisfying } g^{*T}p = 0 \text{ and } J_{\mathcal{A}}^* p \geq 0. \tag{2.13}$$

While SSC1 may hold for one multiplier λ^*, it may not hold for all acceptable $\lambda \in \mathcal{M}_\lambda$. For example, consider the problem

$$\underset{x \in \mathbb{R}^2}{\text{minimize}}\; x_1 \quad \text{subject to} \quad x_1 \geq 0 \quad \text{and} \quad 1 - (x_1 - 1)^2 - x_2^2 \geq 0,$$

in which only the second constraint has a nonzero Hessian. The origin is a KKT point where both constraints are active, with

$$g^* = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad J_{\mathcal{A}}^* = \begin{pmatrix} 1 & 0 \\ 2 & 0 \end{pmatrix}, \quad \text{and} \quad \mathcal{M}_\lambda(x^*) = \{\lambda : \lambda_1 + 2\lambda_2 = 1, \;\; \lambda \geq 0\}.$$

To determine whether condition SSC1 holds, we need to characterize all vectors p satisfying $g^{*T}p = 0$ and $J_{\mathcal{A}}^{*}p \geq 0$. Every such vector has the form $\alpha(0,1)^T$ for some scalar α, so that the inner product appearing in (2.13) of SSC1 is given by

$$p^T H(x^*, \lambda)p = 2p^T \begin{pmatrix} \lambda_2 & 0 \\ 0 & \lambda_2 \end{pmatrix} p = 2\alpha^2 \lambda_2.$$

It follows that SSC1 will hold only if $\lambda_2 > 0$, which means that SSC1 holds for x^* and (say) $\lambda = (0, \frac{1}{2})^T$. However, although $\bar{\lambda} = (1,0)^T$ is an acceptable multiplier, SSC1 does not hold for $(x^*, \bar{\lambda})$.

A more demanding second-order condition imposes a requirement analogous to that in SSC1 on *every* acceptable multiplier.

DEFINITION 2.20 (second second-order sufficient condition (SSC2)). *Consider the problem of minimizing $f(x)$ subject to $c(x) \geq 0$, and let x^* denote a KKT point. We say that SSC2 holds at x^* if for* every *Lagrange multiplier λ satisfying $\lambda \geq 0$, $c^* \cdot \lambda = 0$, and $g^* = J^{*T}\lambda$, there exists $\omega > 0$ such that $p^T H(x^*, \lambda)p \geq \omega \|p\|^2$ for all nonzero p satisfying $g^{*T}p = 0$ and $J_{\mathcal{A}}^{*}p \geq 0$.*

The purpose of defining second-order sufficient conditions is to allow us to verify conclusively that a certain point is a local constrained minimizer. It turns out that, strong as SSC1 and SSC2 may seem, they allow us to prove only that x^* is a *strict* local constrained minimizer of the inequality-constrained problem. The proof of the next result is widely available in the literature; see, for example, [33, 34, 80, 88].

THEOREM 2.21 (sufficient conditions for a strict constrained minimizer). *Consider the problem of minimizing $f(x)$ subject to $c(x) \geq 0$. The point x^* is a strict local constrained minimizer if*

(i) *x^* is a KKT point, i.e., $c^* \geq 0$ and there exists a nonempty set $\mathcal{M}_\lambda$ of multipliers λ satisfying $\lambda \geq 0$, $c^* \cdot \lambda = 0$, and $g^* = J^{*T}\lambda$;*

(ii) *for all $\lambda \in \mathcal{M}_\lambda$ and all nonzero p satisfying $g^{*T}p = 0$ and $J_{\mathcal{A}}^{*}p \geq 0$, there exists $\omega > 0$ such that $p^T H(x^*, \lambda)p \geq \omega \|p\|^2$.* ∎

Theorem 2.21 includes no mention of a constraint qualification, but such a condition is needed to prove the stronger result that x^* is an *isolated* local constrained minimizer, as we are about to do in Theorem 2.23. The next lemma verifies boundedness of a *sequence* of multiplier sets; it also verifies boundedness of the multipliers when the MFCQ is satisfied at a point (see Lemma 2.11).

LEMMA 2.22 (locally uniformly bounded multipliers). *Consider the problem of minimizing $f(x)$ subject to $c(x) \geq 0$. Let x^* be a KKT point, and assume that the MFCQ holds at x^*. Consider a sequence $\{x_k\}$ converging to x^* such that each x_k is a KKT point, and let $\mathcal{M}_\lambda^k$ denote the set of acceptable multipliers at x_k. Then $\mathcal{M}_\lambda^k$ is uniformly bounded for sufficiently large k.*

Proof. The proof is by contradiction. Suppose that, for the given sequence $\{x_k\}$ of KKT points converging to x^*, there is a corresponding sequence $\{\lambda^k\}$ of acceptable multipliers such that $\lim_{k\to\infty} \|\lambda^k\| = \infty$. Since the sequence $\{\lambda^k/\|\lambda^k\|\}$ of unit-norm vectors lies in a compact set, we can assume without loss of generality that $\lim_{k\to\infty} \lambda^k/\|\lambda^k\| = \ell^*$ with $\|\ell^*\| = 1$.

For sufficiently large k, the constraints inactive at x^* are also inactive at x_k, and thus $\mathcal{A}(x_k) \subseteq \mathcal{A}(x^*)$. In the rest of the proof, $\mathcal{A}$ without an argument denotes $\mathcal{A}(x^*)$. To satisfy complementarity at each x_k, we must have $\lambda_i^k = 0$ for $i \notin \mathcal{A}(x_k)$, which means that $\lambda_i^k = 0$ for $i \notin \mathcal{A}$. Consequently, $\|\lambda^k\| = \|\lambda_{\mathcal{A}}^k\|$, $\ell_i^* = 0$ for $i \notin \mathcal{A}$, and $\|\ell_{\mathcal{A}}^*\| = 1$.

For each k, $\lambda_{\mathcal{A}}^k \geq 0$, and it follows that $\ell_{\mathcal{A}}^* \geq 0$. Using the relation $g_k = J_{\mathcal{A}}(x_k)^T \lambda_{\mathcal{A}}^k$, dividing by $\|\lambda_{\mathcal{A}}^k\|$, and letting $k \to \infty$, our assumption that $\|\lambda_{\mathcal{A}}^k\| \to \infty$ implies that $J_{\mathcal{A}}^{*T} \ell_{\mathcal{A}}^* = 0$.

Combining our results, $\ell_{\mathcal{A}}^* \geq 0$, $\|\ell_{\mathcal{A}}^*\| = 1$, and $J_{\mathcal{A}}^{*T} \ell_{\mathcal{A}}^* = 0$. However, because the MFCQ holds at x^*, Lemma 2.12 shows that no such $\ell_{\mathcal{A}}^*$ can exist. We conclude that there can be no unbounded sequence $\{\lambda^k\}$, proving the result. ∎

The proof here of Theorem 2.23 specializes that of [88, Theorem 2.4].

THEOREM 2.23 (sufficient conditions for an isolated constrained minimizer). *Consider the problem of minimizing $f(x)$ subject to $c(x) \geq 0$. The point x^* is an isolated local constrained minimizer if*

(i) *x^* is a KKT point, i.e., $c^* \geq 0$ and there exists a nonempty set $\mathcal{M}_\lambda$ of multipliers λ satisfying $\lambda \geq 0$, $c^* \cdot \lambda = 0$, and $g^* = J^{*T}\lambda$;*

(ii) *the MFCQ holds at x^*, i.e., there is a vector p such that $J_{\mathcal{A}}^* p > 0$;*

(iii) *for all $\lambda \in \mathcal{M}_\lambda$ and all nonzero p satisfying $g^{*T} p = 0$ and $J_{\mathcal{A}}^* p \geq 0$, there exists $\omega > 0$ such that $p^T H(x^*, \lambda) p \geq \omega \|p\|^2$.*

Proof. The result is proved by contradiction, starting by showing that x^* is an isolated KKT point. Assume the contrary, that there is a sequence of points $\{x_k\}$, $x_k \neq x^*$, converging to x^* such that x_k is a KKT point for every k. Consequently, $c(x_k) \geq 0$, and there exists a set of multipliers $\mathcal{M}_\lambda(x_k)$ containing vectors λ_k satisfying $\lambda_k \geq 0$, $c(x_k) \cdot \lambda_k = 0$, and $g(x_k) = J_{\mathcal{A}}(x_k)^T \lambda_k$.

The fact that c is twice-continuously differentiable means that, for sufficiently large k, the only constraints that can be active at x_k are those already active at x^*, so that $\mathcal{A}(x_k) \subseteq \mathcal{A}(x^*)$. Because the MFCQ holds at x^*, we know that there is a vector p^* such that $\nabla c_i(x^*)^T p^* > 0$ for $i \in \mathcal{A}(x^*)$. Our assumptions about differentiability of c and convergence of x_k to x^* mean that, for sufficiently large k, $\nabla c_i(x_k)^T p^* > 0$ for $i \in \mathcal{A}(x^*)$. Since $\mathcal{A}(x_k) \subseteq \mathcal{A}(x^*)$, it follows that $\nabla c_i(x_k)^T p^* > 0$ for $i \in \mathcal{A}(x_k)$, and hence that the MFCQ also holds at x_k.

From Lemma 2.22, the sets of multipliers $\mathcal{M}_\lambda(x_k)$ are uniformly bounded. Thus, moving to a subsequence if necessary and noting that $\mathcal{M}_\lambda(x_k)$ and $\mathcal{M}_\lambda(x^*)$ are closed, we conclude that there is a sequence of multipliers $\{\lambda_k\}$ that converges to some $\lambda^* \in \mathcal{M}_\lambda(x^*)$.

By again moving to a subsequence if needed, we know that the unit vectors $(x_k - x^*)/\|x_k - x^*\|$ converge to a limit, denoted by q:

$$\lim_{k \to \infty} \frac{x_k - x^*}{\|x_k - x^*\|} = q, \quad \text{with} \quad \|q\| = 1. \tag{2.14}$$

We will derive a contradiction by applying assumption (iii) of the theorem with q playing the role of p. Thus we need to show that $J_{\mathcal{A}}^* q \geq 0$ and $g^{*T} q = 0$.

As in the proof of Lemma 2.16, the first step is to expand c_i around x^*, which gives

$$c_i(x_k) = c_i^* + \nabla c_i(x^*)^T (x_k - x^*) + o(\|x_k - x^*\|).$$

Feasibility of $c(x_k)$ implies that, for all $i \in \mathcal{A}(x^*)$,

$$\nabla c_i(x^*)^T (x_k - x^*) + o(\|x_k - x^*\|) \geq 0,$$

and so, dividing by $\|x_k - x^*\|$ and letting $k \to \infty$, we conclude that $J_{\mathcal{A}}^* q \geq 0$. Since $g^* = J_{\mathcal{A}}^{*T} \lambda_{\mathcal{A}}$ for some $\lambda_{\mathcal{A}} \geq 0$, it follows that

$$g^{*T} q = \lambda_{\mathcal{A}}^T J_{\mathcal{A}}^* q \geq 0. \tag{2.15}$$

Now we wish to show that $g^{*T}q \leq 0$, which will verify, in combination with (2.15), that $g^{*T}q = 0$. For this, we need the hypothesis that $\{x_k\}$ is a sequence of KKT points. For each k, it follows from the complementarity condition $c(x_k) \cdot \lambda_k = 0$, boundedness of $\mathcal{M}_\lambda(x_k)$, and smoothness of c that

$$0 = \lambda_k^T c(x_k) = \lambda_k^T c^* + \lambda_k^T J^*(x_k - x^*) + o(\|x_k - x^*\|). \tag{2.16}$$

Note that boundedness of $\mathcal{M}_\lambda(x_k)$, a consequence of the MFCQ, is necessary to justify the size of the remainder term. Since $c^* \geq 0$ and $\lambda_k \geq 0$, the quantity $\lambda_k^T c^*$ is nonnegative, which means, rearranging (2.16), that

$$\lambda_k^T J^*(x_k - x^*) + o(\|x_k - x^*\|) = -\lambda_k^T c^* \leq 0.$$

If we now divide by $\|x_k - x^*\|$ and let $k \to \infty$, the result is

$$\lambda^{*T} J^* q \leq 0. \tag{2.17}$$

Again invoking the hypothesis that x_k is a KKT point, we have $g(x_k) = J(x_k)^T \lambda_k$, so that

$$g(x_k)^T(x_k - x^*) = \lambda_k^T J(x_k)(x_k - x^*).$$

Dividing by $\|x_k - x^*\|$, letting $k \to \infty$, and using (2.17), the result is $g^{*T}q \leq 0$. Combining this with (2.15), it follows that $g^{*T}q = 0$. The vector q defined in (2.14) thus satisfies $g^{*T}q = 0$ and $J_A^* q \geq 0$, qualifying for assumption (iii) of the theorem.

Let δ_k denote $x_k - x^*$ and ξ_k denote $\lambda_k - \lambda^*$. For $\theta \in [0, 1]$, define

$$x(\theta) = x^* + \theta\delta_k \quad \text{and} \quad \lambda(\theta) = \lambda^* + \theta\xi_k,$$

so that $x(0) = x^*$, $x(1) = x_k$, $\lambda(0) = \lambda^*$, and $\lambda(1) = \lambda_k$. Now consider, for each k, the scalar function $s_k(\theta)$, defined for $0 \leq \theta \leq 1$ as

$$\begin{aligned} s_k(\theta) &= \delta_k^T \Big(g\big(x(\theta)\big) - J\big(x(\theta)\big)^T \lambda(\theta) \Big) + \xi_k^T c\big(x(\theta)\big) \\ &= \delta_k^T \nabla_x L\big(x(\theta), \lambda(\theta)\big) + \xi_k^T c\big(x(\theta)\big), \end{aligned}$$

where L is the Lagrangian function (2.3). Because of the assumed smoothness properties of f and c, this function is continuous for $\theta \in [0, 1]$ and continuously differentiable for $\theta \in (0, 1)$.

Observe that, from the definition of ξ_k and complementarity at x^* and x_k,

$$\begin{aligned} s_k(0) &= \delta_k^T \nabla_x L(x^*, \lambda^*) + \lambda_k^T c^* - \lambda^{*T} c^* = \lambda_k^T c^* \geq 0 \quad \text{and} \\ s_k(1) &= \delta_k^T \nabla_x L(x_k, \lambda_k) + \lambda_k^T c(x_k) - \lambda^{*T} c(x_k) = -\lambda^{*T} c(x_k) \leq 0. \end{aligned}$$

Applying the mean-value theorem, for each k there must be some $\theta_k \in (0, 1)$ such that $s_k'(\theta_k) \leq 0$. Writing out $s'(\theta_k)$, we have

$$\begin{aligned} 0 \geq s_k'(\theta_k) &= \delta_k^T \nabla^2_{xx} L\big(x(\theta_k), \lambda(\theta_k)\big)\delta_k - \delta_k^T J\big(x(\theta_k)\big)^T \xi_k + \xi_k^T J\big(x(\theta_k)\big)\delta_k \\ &= \delta_k^T \nabla^2_{xx} L\big(x(\theta_k), \lambda(\theta_k)\big)\delta_k. \end{aligned}$$

Dividing by $\|\delta_k\|^2$ and letting $k \to \infty$ gives $q^T \nabla^2_{xx} L(x^*, \lambda^*) q \leq 0$, which violates assumption (iii) of the theorem. We conclude that x^* is an isolated KKT point.

Since x^* satisfies all the conditions of Theorem 2.21, we know that x^* is a strict local constrained minimizer.

Finally, to show that x^* is an *isolated* local minimizer, we use Lemma 2.16, which asserts that any local minimizer where the MFCQ holds must be a KKT point. Since the MFCQ holds in a neighborhood of x^*, any minimizer in that neighborhood must be a KKT point—but we have just shown that a neighborhood of x^* exists in which there are no other KKT points. Hence the neighborhood is also devoid of local minimizers (apart from x^*), and x^* is an isolated local constrained minimizer. ∎

Although Theorem 2.23 is very nice, its conditions are not easy to check in their full generality. Confirming that the MFCQ holds when the LICQ does not requires solution of a linear program (2.10); verifying assumption (iii) for all p such that $J_{\mathcal{A}}^* p \geq 0$ requires finding the global minimizer of a possibly indefinite quadratic form over a cone, an NP-hard problem [75, 82], not to mention the issue of how to check that (iii) holds for all $\lambda \in \mathcal{M}_\lambda$. If, however, the gradients of the active constraints at x^* are linearly independent and strict complementarity holds, Theorem 2.23 leads immediately to the following result, which we state separately for future reference.

THEOREM 2.24 (strong sufficient conditions for an isolated constrained minimizer). *Consider the problem of minimizing $f(x)$ subject to $c(x) \geq 0$. The point x^* is an isolated local constrained minimizer if*

(i) *x^* is feasible and the LICQ holds at x^*, i.e., $J_{\mathcal{A}}(x^*)$ has full row rank;*

(ii) *x^* is a KKT point and strict complementarity holds, i.e., the (necessarily unique) multiplier λ^* has the property that $\lambda_i^* > 0$ for all $i \in \mathcal{A}(x^*)$;*

(iii) *for all nonzero vectors p satisfying $J_{\mathcal{A}}(x^*)p = 0$, there exists $\omega > 0$ such that $p^T H(x^*, \lambda^*)p \geq \omega \|p\|^2$.*

Proof. The conditions of this theorem imply those of Theorem 2.23. In both cases, x^* is a KKT point. Satisfaction of the LICQ at x^* (the second part of condition (i) in this theorem) implies that the MFCQ holds at x^*. Using Lemma 2.8, it follows from strict complementarity (the second part of (ii) in this theorem) that every nonzero vector satisfying $g^{*T} p = 0$ and $J_{\mathcal{A}}^* p \geq 0$ satisfies $J_{\mathcal{A}}^* p = 0$. Because of this and the uniqueness of λ^*, which follows from the full row rank of $J_{\mathcal{A}}^*$, condition (iii) of this theorem is equivalent to condition (iii) of Theorem 2.23. ∎

Although the assumptions of Theorem 2.24 are more restrictive than those of Theorem 2.23, they have the practical advantage that they can be confirmed (or not) in a straightforward fashion. Linear independence of the active constraint gradients can be verified, or shown to fail, by computing either the QR or LU factorization of $J_{\mathcal{A}}^*$. The (unique) multiplier λ^* is obtained by solving a single compatible linear system for $\lambda_{\mathcal{A}}^*$, and strict complementarity can be checked by examining the components of $\lambda_{\mathcal{A}}^*$ to see whether they are sufficiently positive. Finally, the vectors of interest in condition (iii) lie in the null space of $J_{\mathcal{A}}^*$, for which an explicit basis, say $N_{\mathcal{A}}$, can be computed using either the QR or LU factorization. If this calculation is too expensive, the inertia of $N_{\mathcal{A}}^T H(x^*, \lambda^*) N_{\mathcal{A}}$ can be computed implicitly through the associated augmented system; see Lemma A.15.

3. Barrier Methods.

As described in section 1, classical barrier methods provide the foundation for modern interior methods, and we accordingly devote this section to a survey of the associated theory.

3.1. Overview of Barrier Methods.

As noted in section 1.2, in the 1960s the most popular approaches to solving constrained nonlinear optimization problems were based on penalty and barrier methods, which have a common motivation: finding an *unconstrained minimizer of a composite function* that reflects the original objective function as well as the presence of constraints.

The influence of the constraints can be reflected in the composite function in at least two distinct ways: the constraints can be ignored as long as they are satisfied (i.e., the composite function matches the objective function at all feasible points), or the composite function is defined only at feasible points. An extreme way to take the second approach is to define the composite function as $f(x)$ when x is feasible and as $+\infty$ otherwise, but the wildly discontinuous result would be effectively impossible to minimize. Adopting the approach of remaining feasible while preserving any nice properties (such as smoothness) of $c(x)$ leads to the concept of a well-behaved *interior* function $I(x)$ that is not allowed to leave the interior of the feasible region. By appropriately combining $I(x)$ with $f(x)$ as well as systematically reducing the effect of the constraints, we should be able to create a composite function whose unconstrained minimizers will converge to a local constrained minimizer of the original problem.

This motivation suggests the following desirable properties for $I(x)$:

I1. $I(x)$ depends only on the constraint functions.

I2. $I(x)$ preserves the continuity properties of $c(x)$ at all points in $\mathrm{int}(\mathcal{F})$.

I3. For any sequence of points in $\mathrm{int}(\mathcal{F})$ converging to a point on the boundary of the feasible region, $I(x) \to \infty$.

If these desiderata can be achieved (and, as we shall see in section 3.2, they almost can be), there is an obvious rationale for applying the descriptors "interior" and "barrier" to the associated methods, since the effect of $I(x)$ is to prevent unconstrained minimizers of the composite function from leaving the feasible region.

The properties of generic interior functions are analyzed in [32, 33]. In 1961, based on his 1959 Ph.D. thesis, Carroll [19] described the *inverse* interior function,

$$I_{\mathrm{inv}}(x) \triangleq \sum_{i=1}^{m} \frac{1}{c_i(x)}. \tag{3.1}$$

Alternatively, since $-\ln c_i(x) \to \infty$ as $c_i(x) \to 0+$ and $\ln c_i(x)$ is twice-continuously differentiable when $c_i(x) > 0$, the *logarithmic interior function* $I_{\log}(x)$, defined as

$$I_{\log}(x) \triangleq -\sum_{i=1}^{m} \ln c_i(x), \tag{3.2}$$

also appears to satisfy the desired properties. The earliest history of the logarithmic interior function is worth noting. Frisch's 1955 "logarithmic potential method" [44] is based on using the gradient of $f(x) + \sum_{i=1}^{m} \alpha_i \ln c_i(x)$ to retain feasibility and accelerate convergence; however, Frisch's approach did not involve unconstrained minimization of this function. See [33, pp. 6–15] for a thorough and fascinating chronological survey of ideas for solving constrained problems via a sequence of unconstrained problems.

Because of its connection with Karmarkar's linear programming method as well as other reasons, the overwhelmingly predominant barrier function today is the *logarithmic barrier function*, a composite based on the logarithmic interior function:

$$B(x, \mu) = f(x) - \mu I_{\log}(x) = f(x) - \mu \sum_{i=1}^{m} \ln c_i(x), \tag{3.3}$$

where μ is a positive scalar, the *barrier parameter*. An important feature of $B(x, \mu)$ is that it retains the smoothness properties of $f(x)$ and $c(x)$ as long as $c(x) > 0$. Accordingly, for very small $\mu > 0$, $B(x, \mu)$ "acts like" $f(x)$ except close to points where any constraint is zero. Intuition then suggests that minimizing $B(x, \mu)$ for a sequence

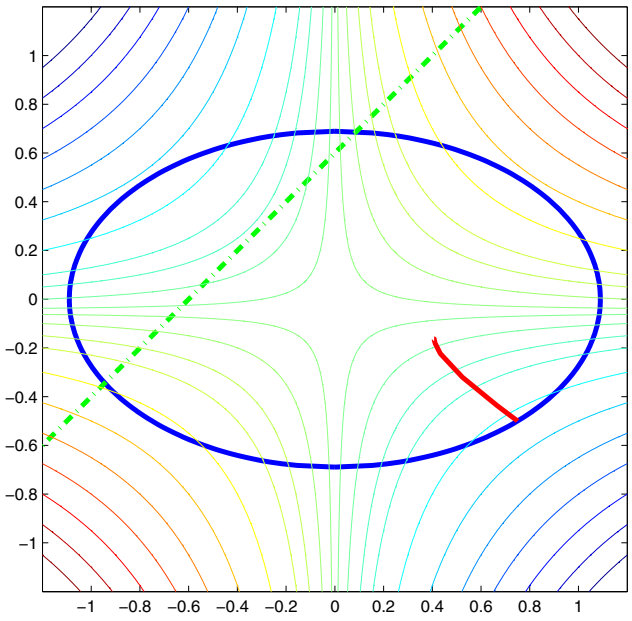

Fig. 2 *The contours of the nonconvex objective function* (3.4a) *are shown, along with the boundaries of the ellipsoidal and linear constraints* (3.4b). *A trajectory of local unconstrained minimizers of the logarithmic barrier function, shown in red, begins at the strictly feasible analytic center of the feasible region, corresponding to* $\mu = \infty$, *and converges to the boundary as* $\mu \to 0$.

of positive μ values converging to zero will cause the unconstrained minimizers of $B(x, \mu)$ to converge to a local constrained minimizer of the original problem.

To illustrate the behavior of the log barrier function, we consider the two-variable inequality-constrained problem:

$$\text{(3.4a)} \qquad \text{minimize} \quad \tfrac{10}{3}x_1x_2 + \tfrac{1}{6}x_1$$

$$\text{(3.4b)} \qquad \text{subject to} \quad \tfrac{19}{16} - x_1^2 - \tfrac{5}{2}x_2^2 \geq 0 \quad \text{and} \quad x_1 - x_2 + \tfrac{3}{5} \geq 0.$$

The first (nonlinear) constraint is satisfied inside an ellipse centered at the origin; the second (linear) constraint cuts off part of the ellipse. Figure 2 shows the contours of f, which is unbounded below, and the boundaries of these two constraints; the feasible region lies inside the ellipse, to the right of the line.

The figure makes clear that there are two local minimizers of f in the feasible region. At the isolated constrained minimizer $x^* = (\frac{3}{4}, \frac{-1}{2})$, the first constraint is active. The path of barrier minimizers converging to x^* is shown in red. The strictly feasible starting point of the path of barrier minimizers corresponds to the minimizer of $-\sum \ln c_i(x)$—in effect, to an infinite value of μ, so that the objective function has no effect.[3]

[3] In general, the *analytic center* corresponding to the constraints $c(x) \geq 0$ is the maximizer of $\ln(\prod_{i=1}^{m} c_i(x))$. The analytic center depends on the forms of the constraints rather than on the nature of the feasible region itself; see the discussion at the beginning of section 3.2.

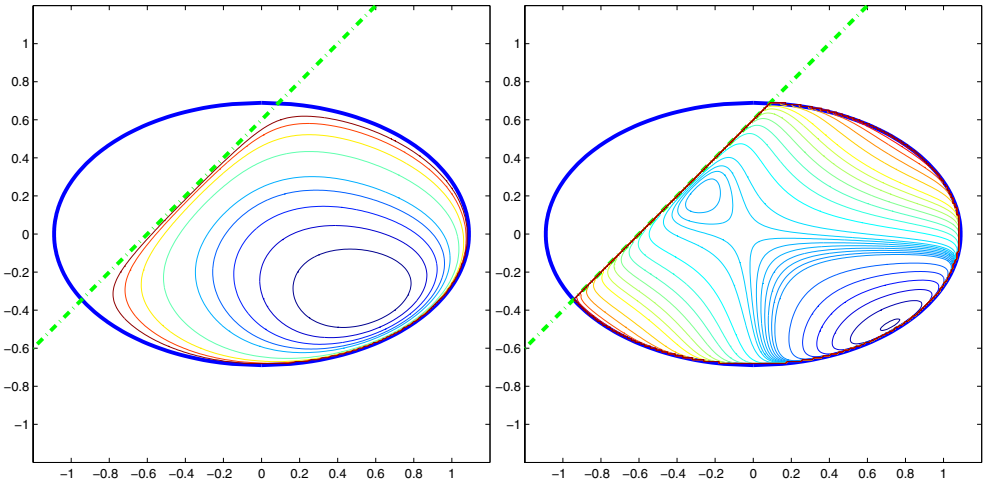

Fig. 3 *The contours of the logarithmic barrier function* (3.3) *for the problem depicted in Figure* 2 *are shown for* $\mu = 1$ *and* $\mu = 0.1$.

The effect of μ on the contours of the barrier function is depicted in Figure 3. As μ becomes smaller, the barrier function becomes more and more like the objective function, except very close to the boundary where the logarithmic singularity has an effect. When $\mu = 0.1$, the barrier function obviously has two local minimizers, reflecting the same property of the original problem.

In sections 3.3 and 3.4 we shall probe the connection between minimizers of the barrier function and local constrained minimizers of the original problem. An initial hint of those relationships can be seen algebraically by writing down the gradient of the barrier function (3.3), denoted by $\nabla B(x, \mu)$, which can be expressed in various equivalent forms:

$$\nabla B(x,\mu) = g(x) - \sum_{i=1}^{m} \frac{\mu}{c_i(x)} \nabla c_i(x), \tag{3.5a}$$

$$= g(x) - \mu J(x)^T C(x)^{-1} e, \quad \text{and} \tag{3.5b}$$

$$= g(x) - J(x)^T \big(\mu \mathbin{./} c(x)\big). \tag{3.5c}$$

In the form (3.5b), $C(x)$ denotes the $m \times m$ diagonal matrix of constraint values and e the m-vector of all ones. The third form, (3.5c), uses a MATLAB-like notation for a componentwise quotient, which we now formally define.

DEFINITION 3.1 (the componentwise division operator $\mathbin{./}$). *Given a scalar* α *and an* r*-vector* v *whose components are all nonzero,* $\alpha \mathbin{./} v$ *is an* r*-vector whose* i*th component is* α/v_i*. Given two vectors* u *and* v *of dimension* r*, where each component of* v *is nonzero,* $u \mathbin{./} v$ *is an* r*-vector whose* i*th component is* u_i/v_i*.*

An unconstrained minimizer of $B(x, \mu)$ will be denoted by either x_μ or $x(\mu)$, and we assume for the moment that $c(x_\mu) > 0$ (which we shall prove later). Because $\nabla B(x, \mu)$ is twice-continuously differentiable, it must hold that $\nabla B(x_\mu, \mu) = 0$ (see

Lemma A.7), which means, using (3.5a) and (3.5b), that

$$g(x_\mu) = \sum_{i=1}^{m} \frac{\mu}{c_i(x_\mu)} \nabla c_i(x_\mu) = \mu J(x_\mu)^T C(x_\mu)^{-1} e.$$

Hence the objective gradient at x_μ is a positive linear combination of the constraint gradients. The coefficients in that linear combination, which will be important later, are called the *barrier multipliers* (by analogy with Lagrange multipliers) and denoted by λ_μ. Formally, λ_μ is defined as

$$\lambda_\mu \triangleq \mu C^{-1}(x_\mu) e = \mu \cdot/ c(x_\mu). \tag{3.6}$$

Thus at x_μ we have

$$g(x_\mu) = \sum_{i=1}^{m} \frac{\mu}{c_i(x_\mu)} \nabla c_i(x_\mu) = J(x_\mu)^T \lambda_\mu, \quad \text{with} \quad \lambda_\mu > 0. \tag{3.7}$$

This should vividly remind the reader of the stationarity and nonnegativity properties (2.2b) and (2.2c), $g(x^*) = J(x^*)^T \lambda^*$ and $\lambda^* \geq 0$, that hold at a KKT point.

The resemblance is even more evocative when we express the definition of λ_μ as

$$c(x_\mu) \cdot \lambda_\mu = \mu, \quad \text{or} \quad c_i(x_\mu)(\lambda_\mu)_i = \mu, \ \ i = 1, \ldots, m. \tag{3.8}$$

This componentwise relationship between the barrier multipliers, constraint values, and the barrier parameter, called *perturbed complementarity*, is analogous as $\mu \to 0$ to the complementarity condition $c(x^*) \cdot \lambda^* = 0$ (2.2d) that holds at a KKT point.

We shall explore these relationships in detail after presenting background that exposes some messy nuances.

3.2. Background Results. Since our analysis of barrier methods hinges on the properties of unconstrained minimizers, we need to define them formally.

DEFINITION 3.2 (local unconstrained minimizer). *The point x^* is a* local unconstrained minimizer *of $f(x)$ if there exists a compact set S such that $x^* \in \text{int}(S)$ and $f(x^*) = \min f(x)$ for $x \in S$.*

As with Definition 2.14, the set S is usually taken as a neighborhood of x^*.

Our discussion of interior methods thus far has glossed over a few points that need further examination—in particular, the relationship between topological properties of the feasible region $\mathcal{F}$ viewed as a set and the representation of $\mathcal{F}$ via inequality constraints. Unfortunately, the two representations are not always equivalent. This affects our discussion of barrier functions because the desired property I1 of an interior function specifies that it should depend on the constraint functions, whereas properties I2 and I3 are couched in terms of the interior and boundary of $\mathcal{F}$.

For the set $\mathcal{F}$ of points $x \in \mathbb{R}^n$ satisfying $c(x) \geq 0$, Definition A.2 characterizes its interior, or *topological interior*, independently of how $\mathcal{F}$ happens to be represented. Similarly, the boundary of a set (Definition A.3) is a topological property. On the other hand, the set of *strictly feasible points* is defined in terms of the constraint functions $\{c_i\}$.

DEFINITION 3.3 (strictly feasible points). *The subset of points in $\mathcal{F}$ for which all the constraint functions are* strictly positive *is denoted by* strict($\mathcal{F}$) *and defined as* strict($\mathcal{F}$) $\triangleq \{x : c_i(x) > 0, i = 1, \ldots, m\}$. *A point x in* strict($\mathcal{F}$) *is said to be* strictly feasible.

The topological interior of $\mathcal{F}$ is not always the same as the set of strictly feasible points. Consider the two constraints

$$c_1(x) = x^2 \geq 0 \quad \text{and} \quad c_2(x) = x + \gamma \geq 0, \tag{3.9}$$

where $\gamma > 0$, with associated feasible set

$$\mathcal{F} = \{x : x^2 \geq 0 \quad \text{and} \quad x \geq -\gamma\}.$$

Since all real x satisfy $x^2 \geq 0$, the first constraint does not actually constrain anything, and $x = 0$ is in fact an interior point of $\mathcal{F}$, even though the first constraint is equal to zero; thus $\text{int}(\mathcal{F}) = \{x : x > -\gamma\}$. In contrast, the origin is excluded from $\text{strict}(\mathcal{F})$, so that

$$\text{strict}(\mathcal{F}) = \{x : 0 > x > -\gamma\} \cup \{x : x > 0\} = \text{int}(\mathcal{F}) \backslash \{0\},$$

and $\text{strict}(\mathcal{F}) \neq \text{int}(\mathcal{F})$. If $\mathcal{F}$ is convex (Definition A.4), every point lying between two strictly feasible points is also strictly feasible, so that these difficulties do not occur. We briefly consider how to formalize such anomalies.

DEFINITION 3.4 (topological inconsistency). *A constraint $c_i(x) \geq 0$ is said to be* topologically inconsistent *at $\hat{x}$ if $\hat{x} \in \text{int}(\mathcal{F})$ and $c_i(\hat{x}) = 0$. The constraint $c_i(x) \geq 0$ is* topologically consistent *if, for all $x \in \mathcal{F}$, $c_i(x) = 0$ only if $x \notin \text{int}(\mathcal{F})$.*

LEMMA 3.5. *If a constraint $c_i(x) \geq 0$ is topologically inconsistent at $\hat{x}$, then $\hat{x}$ is a local unconstrained minimizer of $c_i(x)$, with $\nabla c_i(\hat{x}) = 0$ and $\nabla^2 c_i(\hat{x})$ positive semidefinite.*

Proof. From the definition of topological inconsistency, $c_i(\hat{x}) = 0$ and $\hat{x} \in \text{int}(\mathcal{F})$. By definition of $\text{int}(\mathcal{F})$ (Definition A.2), there is a neighborhood of $\hat{x}$ contained entirely in $\mathcal{F}$. Hence there exists $\epsilon > 0$ such that if $\|x - \hat{x}\| \leq \epsilon$, then $c_i(x) \geq 0$, so that $c_i(\hat{x}) \leq c_i(x)$. Applying Definition 3.2, $\hat{x}$ is a local unconstrained minimizer of c_i. Since c_i is smooth, $\hat{x}$ must satisfy standard necessary conditions for an unconstrained minimizer, namely, that $\nabla c_i(\hat{x}) = 0$ and $\nabla^2 c_i(\hat{x})$ is positive semidefinite (see Lemma A.7). ∎

In practice it is useful to determine when a constraint is topologically inconsistent. We next summarize conditions under which topologically inconsistent constraints are locally redundant, meaning that all points in a neighborhood are feasible. In particular, if $c_i(x)$ is *convex* and topologically inconsistent at a point, it is completely redundant, meaning that all points are feasible.

LEMMA 3.6 (redundant topologically inconsistent constraints).

(i) *Assume that the constraint $c_i(x) \geq 0$ is topologically inconsistent at $\bar{x}$ and that $\nabla^2 c_i(\bar{x})$ is positive definite. Then there is a neighborhood of $\bar{x}$ in which all points are strictly feasible with respect to this constraint.*

(ii) *If $c_i(x)$ is convex on $\mathbb{R}^n$ and topologically inconsistent at $\bar{x}$, all points x in $\mathbb{R}^n$ satisfy $c_i(x) \geq 0$, so that this constraint is redundant.*

Proof. If $c_i(x)$ is topologically inconsistent at $\bar{x}$, then Lemma 3.5 shows that $\bar{x}$ is a local unconstrained minimizer of $c_i(x)$. To verify (i), note that if $\nabla^2 c_i(\bar{x})$ is positive definite, $\bar{x}$ must be an isolated, and hence strict, local unconstrained minimizer of $c_i(x)$. Thus there is a neighborhood of $\bar{x}$ in which $c_i(\bar{x}) < c_i(x)$ for all $x \neq \bar{x}$ in the neighborhood, which means that all points in the neighborhood except $\bar{x}$ are strictly feasible.

To show (ii), we observe that if $c_i(x)$ is convex and topologically inconsistent at $\bar{x}$, we know from Lemma 3.5 and Definition A.5 that $\bar{x}$ must also be a global unconstrained minimizer of $c_i(x)$. Thus for all $x \in \mathbb{R}^n$, $c_i(x) \geq c_i(\bar{x}) = 0$, all points are feasible, and the constraint is redundant. ∎

The constraint $c_1(x) = x^2 \geq 0$ of example (3.9) is convex and topologically inconsistent at the origin, since $0 \in \text{int}(\mathcal{F})$ but $c_1(0) = 0$. Hence $c_1(x) \geq 0$ is redundant, as already observed for this example.

Although topologically inconsistent constraints tend to occur almost exclusively in carefully constructed examples, we need to allow for them in our analysis. In practice, Lemma 3.5 allows us to assert that all constraints are topologically consistent at $\bar{x}$ if, for all $i \in \mathcal{A}(\bar{x})$, either $\nabla c_i(\bar{x}) \neq 0$ or $\nabla^2 c_i(\bar{x})$ has at least one strictly negative eigenvalue.

The effect of these anomalies is that we need to lower our expectations relative to the three original goals I1–I3 for interior functions. Rather than being well behaved for all points in $\text{int}(\mathcal{F})$, the logarithmic interior function $-\sum_{i=1}^m \ln c_i(x)$ is well behaved only for points in $\text{strict}(\mathcal{F})$. Similarly, the logarithmic interior function is unbounded above for any sequence of points in $\text{strict}(\mathcal{F})$ that converges to a point on the boundary of $\text{strict}(\mathcal{F})$, which is not necessarily the same as the boundary of $\mathcal{F}$. These fine points are irritating but inescapable since our only means of defining $\mathcal{F}$ for optimization algorithms requires calculation of $c(x)$.

We note three final details that are peculiar to the logarithmic interior function:

1. Because $\sum_{i=1}^m \ln c_i(x) = \ln\bigl(\prod_{i=1}^m c_i(x)\bigr)$, at points in $\text{strict}(\mathcal{F})$ the logarithmic interior function corresponding to the m individual constraints $c_i(x) \geq 0$ has the same value, gradient, and Hessian as if the problem contained the single constraint $\prod_{i=1}^m c_i(x) \geq 0$, for which $\mathcal{F}$ is obviously different. See example (3.16) for a related discussion.

2. The logarithmic interior function can be negative in $\text{strict}(\mathcal{F})$. In contrast, the inverse interior function $1/c_i(x)$ (3.1) is positive and bounded below if $c_i(x) > 0$.

3. The logarithmic interior function can be unbounded below when the feasible region is unbounded. In such a case, the downward pull of the logarithmic interior function may overpower the objective function and lead to an unbounded sequence of minimizers of the barrier function. An example from Powell [87] illustrates what can go wrong:

$$\underset{x}{\text{minimize}} \quad -1/(x^2+1) \quad \text{subject to} \quad x \geq 1.$$

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

$$B(x,\mu) = -1/(x^2+1) - \mu \ln(x-1)$$

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

Two important background results, which we state without proof, are intuitively clear. These results are needed to prove Theorem 3.10, which shows under relatively weak assumptions that there is a sequence of unconstrained minimizers of the logarithmic barrier function that converges to a local constrained minimizer of the original problem. Although in general we consider problems in which the objective function and constraint functions are twice-continuously differentiable, Lemma 3.7, Theorem 3.9, and Theorem 3.10 assume *only continuity* of f and $\{c_i\}$. Since the definitions of both unconstrained and constrained minimizers (Definitions 3.2 and 2.14) involve finding an appropriate compact set, much of the effort boils down to ensuring existence of, and characterizing, this compact set.

The following lemma confirms our intuition that the barrier function achieves a finite minimum value at a point in $\text{strict}(\mathcal{F})$.

LEMMA 3.7. *Given the constraints $c(x) \geq 0$, where each $c_i(x)$ is continuous for $x \in \mathcal{F}$, let* strict$(\mathcal{F})$ *denote the set of points such that $c(x) > 0$, and let S be a compact set such that* strict$(\mathcal{F}) \cap S$ *is nonempty. Consider a convergent sequence* $\{y_k\} \in$ strict$(\mathcal{F}) \cap S$ *whose limit point $\bar{y}$ lies in S and on the boundary of* strict$(\mathcal{F})$*:*

$$\lim_{k\to\infty} y_k = \bar{y}, \quad \textit{where} \quad \bar{y} \in \mathrm{bnd}\big(\mathrm{strict}(\mathcal{F})\big) \cap S. \tag{3.10}$$

Suppose that φ is a continuous function on 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* (3.10)*. Then φ achieves its (finite) smallest value φ^* for $x \in$* strict$(\mathcal{F}) \cap S$ *at some point x^* in* strict$(\mathcal{F}) \cap S$*.*

Proof. See [33, Corollary 8] and [110, Lemma 1]. □

Our basic convergence theorem (Theorem 3.10) depends on the existence of a subset of local constrained minimizers that is "isolated" within the full set of local constrained minimizers. To ensure existence of that subset, we need the following definition and theorem.

DEFINITION 3.8 (isolated subset). *Let $\mathcal{N}$ and $\mathcal{N}^*$ be sets in $\mathbb{R}^n$ such that $\mathcal{N}^* \subseteq \mathcal{N}$. The set $\mathcal{N}^*$ is called an* isolated *subset of $\mathcal{N}$ if there exists a closed set E such that $\mathcal{N}^* \subset \mathrm{int}(E)$ and $E \cap \mathcal{N} = \mathcal{N}^*$.*

The key element in this definition is the closed set E, which, in effect, "separates" $\mathcal{N}^*$ from the rest of $\mathcal{N}$. The definition is satisfied if $\mathcal{N}$ is compact and $\mathcal{N}^* = \mathcal{N}$, or if $\mathcal{N}^*$ consists of an isolated point in $\mathcal{N}$, i.e., a point in a neighborhood containing no other points of $\mathcal{N}$. In most situations of interest, the local constrained minimizer of interest is itself an isolated point within the set of local constrained minimizers, so this definition applies automatically.

The purpose of the next theorem is to vouch for the existence of a compact set S with two key properties: a set $\mathcal{N}^*$ of local constrained minimizers lies entirely in the interior of S, and all minimizers in $\mathcal{N}^*$ display the strictly smallest value of $f(x)$ for all feasible points in S.

THEOREM 3.9 (existence of compact enclosing set). *Consider minimizing $f(x)$ subject to $c(x) \geq 0$. Let $\mathcal{N}$ denote the set of all local constrained minimizers with objective function value f^*, and assume that f^* has been chosen so that $\mathcal{N}$ is nonempty. Assume further that the set $\mathcal{N}^* \subseteq \mathcal{N}$ is a nonempty compact isolated subset of $\mathcal{N}$. Then there exists a compact set S such that $\mathcal{N}^*$ lies in* int$(S) \cap \mathcal{F}$ *and $f(y) > f^*$ for any feasible point y in S but not in $\mathcal{N}^*$. Every point x^* in $\mathcal{N}^*$ thus has the property that $f(x^*) = f^* = \min f(x)$ for all $x \in S \cap \mathcal{F}$.*

Proof. See [33, Theorem 7] and [110, Theorem 6]. □

The conditions of this theorem are relatively mild, but they do not apply in all cases. Consider the problem of minimizing $f(x) = x^6 \sin^2(1/x)$ with no constraints, where $f(0) \triangleq 0$. If we take $f^* = 0$, $\mathcal{N}$ contains the origin and all points where $\sin^2(1/x) = 0$. However, we cannot take $\mathcal{N}^*$ as the origin since every neighborhood of the origin includes minimizers for which $f(x) = f^* = 0$.

3.3. A General Convergence Theorem for the Logarithmic Barrier Function.

We will (momentarily) state and prove a fundamental theorem about local convergence of a sequence of minimizers of the logarithmic barrier function. Two ingredients in this theorem have already been discussed—existence of a nonempty set $\mathcal{N}$ of local constrained minimizers with objective function value f^*, and, as in Theorem 3.9, existence of $\mathcal{N}^* \subseteq \mathcal{N}$, a nonempty compact isolated subset of $\mathcal{N}$. These assumptions are mild and easy to satisfy.

A new, not so mild requirement is a prescribed relationship between $\mathcal{N}^*$ and the strictly feasible region: that at least one of the points in $\mathcal{N}^*$ must lie in the closure of strict($\mathcal{F}$), i.e., $\mathcal{N}^*$ contains either a strictly feasible point or a limit point of strict($\mathcal{F}$). This assumption is needed because barrier methods can be viewed as finding the infimum of f subject to $c(x) > 0$, so that a sequence of barrier minimizers, each of which lies in strict($\mathcal{F}$), cannot converge to a feasible point in a neighborhood that contains no other feasible points. For example, consider minimizing $f(x) = (x+1)^2$ subject to $c_1(x) = x(x-1) \geq 0$ and $c_2(x) = x \geq 0$. The constraint function $x(x-1)$ is nonnegative if $x \leq 0$ or if $x \geq 1$, so that the feasible points lie in two separated regions. The second constraint, $x \geq 0$, eliminates all of the region $x \leq 0$ except for the origin. It follows that the feasible region $\mathcal{F}$ for *both* constraints consists of the origin and points satisfying $x \geq 1$. The smallest value of f in $\mathcal{F}$ is $f^* = 1$, which occurs only at $x^* = 0$, so that $\mathcal{N}^*$ contains a single point, the origin. But since strict($\mathcal{F}$) is the open interval $(1, \infty)$, the origin does not lie in the closure of strict($\mathcal{F}$), and $\mathcal{N}^*$ does not satisfy the needed condition.

The proof of this theorem follows those of [33, Theorems 8 and 10] and [110, Theorem 7]. We include a complete proof for two reasons: first, the proofs in [33] are not completely correct because the logarithmic interior function can be negative in strict($\mathcal{F}$), as noted in the second observation in section 3.2; second, the proofs in [110] treat the convex case separately, which we have not done here. An analogous convergence theorem [32, Theorem 1] with mild assumptions treats only the inverse barrier function.

THEOREM 3.10 (local convergence for barrier methods). *Consider the problem of minimizing $f(x)$ subject to $c(x) \geq 0$, where f and c are continuous. Let $\mathcal{F}$ denote the feasible region, let $\mathcal{N}$ denote the set of minimizers with objective function value f^*, and assume that $\mathcal{N}$ is nonempty. Let $\{\mu_k\}$ be a strictly decreasing sequence of positive barrier parameters such that $\lim_{k\to\infty} \mu_k = 0$. Assume that*

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

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

Then the following results hold:

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

(ii) *for all sufficiently small μ_k, there is an* unconstrained *minimizer y_k of the barrier function $B(x, \mu_k)$ in* strict($\mathcal{F}$) $\cap$ int(S), *with*

$$B(y_k, \mu_k) = \min \{ B(x, \mu_k) : x \in \text{strict}(\mathcal{F}) \cap S \}.$$

Thus $B(y_k, \mu_k)$ is the smallest value of $B(x, \mu_k)$ for any $x \in$ strict($\mathcal{F}$) $\cap S$;

(iii) *any sequence of these unconstrained minimizers $\{y_k\}$ of $B(x, \mu_k)$ has at least one convergent subsequence;*

(iv) *the limit point x_∞ of any convergent subsequence $\{x_k\}$ of the unconstrained minimizers $\{y_k\}$ defined in* (ii) *lies in $\mathcal{N}^*$;*

(v) *for the convergent subsequences $\{x_k\}$ of part* (iv),

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

Proof. Assumption (a) satisfies the conditions of Theorem 3.9, which immediately implies (i). Thus there is a compact set S such that all points in $\mathcal{N}^*$ are in int(S) $\cap\, \mathcal{F}$ and, further, the value of f at points in $\mathcal{N}^*$ is the smallest for any $x \in S \cap \mathcal{F}$. This

conclusion is independent of the barrier function, but sets the stage for subsequent results.

We now turn to the behavior of the barrier function $B(x, \mu_k)$ in the bounded set strict$(\mathcal{F}) \cap S$. It follows from continuity of f and $\{c_i\}$ that $B(x, \mu_k)$ is continuous in strict$(\mathcal{F}) \cap S$. Applying Lemma 3.7, for all k, $B(x, \mu_k)$ achieves its finite smallest value at some point in strict$(\mathcal{F}) \cap S$. (This result is close to but not the same as (ii), which states that the minimizing point is an unconstrained minimizer of $B(x, \mu_k)$; this will be shown later.)

Let y_k be any point in strict$(\mathcal{F}) \cap S$ for which the smallest value of $B(x, \mu_k)$ is achieved. Boundedness of $\{y_k\}$ means that it has at least one limit point, which we denote by x_∞. Let $\{x_k\}$ denote any convergent subsequence of $\{y_k\}$ whose limit is x_∞. Each x_k is strictly feasible and the set S is compact; hence x_∞ must lie in $\mathcal{F} \cap S$, so that x_∞ is feasible.

Now we wish to show that the limit point x_∞ lies in the set $\mathcal{N}^*$ of constrained minimizers, with $f(x_\infty) = f^*$. The result will be proved by contradiction, and we accordingly assume the contrary, that $x_\infty \notin \mathcal{N}^*$. In this case, since x_∞ is feasible and in S, result (i) implies that $f(x_\infty) > f^*$.

We next prove that this inequality implies the existence of a strictly feasible point x_{int} in S such that $f(x_\infty) > f(x_{\text{int}})$, and we find x_{int} by exploiting the properties of $\mathcal{N}^*$. Assumption (b) shows that at least one point in $\mathcal{N}^*$, say x^*, is in the closure of strict$(\mathcal{F})$, and consequently x^* must be either strictly feasible or else an accumulation point of strict$(\mathcal{F})$. The specification of x_{int} depends on which of these two situations applies.

Case 1. If x^* is strictly feasible, we take $x_{\text{int}} = x^*$.

Case 2. Otherwise, x^* is an accumulation point of strict$(\mathcal{F})$, which means that every neighborhood of x^* contains strictly feasible points. Since result (i) implies that $\mathcal{N}^*$ is in int(S), x^* is also in the interior of S, and every neighborhood of x^* thus contains points in S. Because f is continuous, our assumption that $f(x_\infty) > f^*$ means that there must be a strictly feasible point x_{int} in a neighborhood of x^* for which $f(x_{\text{int}}) < f(x_\infty)$.

With x_{int} in hand, the hypothesis that $f(x_\infty) > f(x_{\text{int}})$ implies that, for the convergent sequence $\{x_k\}$ whose limit is x_∞ and for sufficiently large k,

$$f(x_k) > f(x_{\text{int}}). \tag{3.11}$$

Since x_{int} is in strict$(\mathcal{F}) \cap S$, the definition of x_k as a point with the smallest value of $B(x, \mu_k)$ in strict$(\mathcal{F}) \cap S$ implies that

$$f(x_k) - \mu_k \sum_{i=1}^{m} \ln c_i(x_k) \;\le\; f(x_{\text{int}}) - \mu_k \sum_{i=1}^{m} \ln c_i(x_{\text{int}}). \tag{3.12}$$

It follows from strict feasibility of x_{int} that the sum of logarithms involving x_{int} in (3.12) is *finite*, and thus, letting $\mu_k \to 0$, $\lim_{k\to\infty} B(x_{\text{int}}, \mu_k) = f(x_{\text{int}})$.

Now we consider the nature of x_∞. If x_∞ is strictly feasible, then the sum of logarithms in (3.12) involving x_k is finite as $k \to \infty$ and $\lim_{k\to\infty} B(x_k, \mu_k) = f(x_\infty)$. Letting $k \to \infty$ in (3.12), the result is $f(x_\infty) \le f(x_{\text{int}})$, which contradicts the hypothesis of (3.11) that $f(x_k) > f(x_{\text{int}})$.

Alternatively, we consider the consequences if x_∞ is not strictly feasible. Adding $-\mu_k \sum_{k=1}^{m} \ln c_i(x_{\text{int}})$ to both sides of $f(x_k) > f(x_{\text{int}})$, combining with (3.12), rear-

ranging, and dividing by μ_k, we obtain the following result:

$$-\sum_{k=1}^{m} \ln c_i(x_k) < -\sum_{k=1}^{m} \ln c_i(x_{\text{int}}). \tag{3.13}$$

As before, strict feasibility of x_{int} guarantees that the sum on the right-hand side is fixed and finite. However, since by assumption x_∞ is not strictly feasible, $-\ln c_i(x_k)$ approaches infinity for at least one i, and the left-hand side of (3.13) is unbounded above. Once again, we have derived a contradiction.

In either case, we have shown that $f(x_\infty)$ cannot exceed f^*, which implies that $f(x_\infty) = f^*$ and hence that $x_\infty \in \mathcal{N}^*$. Since x_∞ was taken as any limit point of $\{y_k\}$, we conclude that every limit point of a convergent subsequence of the points for which the barrier function achieves its smallest value in $\text{strict}(\mathcal{F}) \cap S$ must be a constrained minimizer with objective value f^*.

To complete the proof of (ii)–(iv), we need to show that the strictly feasible point x_k, which corresponds to the smallest value of $B(x, \mu_k)$ for $x \in \text{strict}(\mathcal{F}) \cap S$, is an *unconstrained* minimizer of $B(x, \mu_k)$. We have just shown that $x_\infty \in \mathcal{N}^*$, which means that $x_\infty \in \text{int}(S)$. Because $\lim_{k\to\infty} x_k = x_\infty$, x_k must also lie in $\text{int}(S)$ for sufficiently large k. Hence, because $x_k \in \text{strict}(\mathcal{F}) \cap \text{int}(S)$, there is a closed neighborhood $N \subseteq S$ of x_k with the property that, for all $x \in N$, $B(x_k, \mu_k) \leq B(x, \mu_k)$. As a result, we can apply Definition 3.2, using N as the needed compact set, to confirm that x_k is an *unconstrained* minimizer of $B(x, \mu_k)$. This proves results (ii), (iii), and (iv).

The first relation in (v), that $\lim_{k\to\infty} f(x_k) = f^*$, follows because $f(x_\infty) = f^*$. To verify the second part of (v), that $\lim_{k\to\infty} B(x_k, \mu_k) = f^*$, we recall that x_k is a point where the barrier function $B(x, \mu_k)$ achieves its smallest value in $\text{strict}(\mathcal{F}) \cap S$. Thus we have, for sufficiently large k, that x_k and x_{k+1} lie in $\text{strict}(\mathcal{F}) \cap S$, and consequently

$$B(x_k, \mu_k) \;\leq\; B(x_{k+1}, \mu_k) \quad \text{and} \quad B(x_{k+1}, \mu_{k+1}) \;\leq\; B(x_k, \mu_{k+1}).$$

Multiplying the first inequality by μ_{k+1}/μ_k (a number strictly between 0 and 1), adding the result to the second inequality, and canceling the terms containing logarithms gives the result

$$f(x_{k+1})\Big(1 - \frac{\mu_{k+1}}{\mu_k}\Big) \;\leq\; f(x_k)\Big(1 - \frac{\mu_{k+1}}{\mu_k}\Big).$$

Since $0 < \mu_{k+1} < \mu_k$, it follows that $f(x_{k+1}) \leq f(x_k)$, so that $f(x_k)$ converges monotonically from above to $f(x_\infty)$.

There are two cases in treating the logarithm terms, depending on whether or not x_∞, the limit point of $\{x_k\}$, is strictly feasible.

Case 1. 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 $\lim_{k\to\infty} B(x_k, \mu_k) = f(x_\infty) = f^*$, where the final equality was derived earlier in the proof, and result (v) is proved.

Case 2. If x_∞ is not strictly feasible, at least one constraint is converging to zero. Hence $-\sum_{i=1}^{m} \ln c_i(x_k)$ must be *positive* for all sufficiently large k, which immediately implies that $B(x_k, \mu_k) > f(x_k)$. Since $f(x_k) \geq f(x_{k+1})$ and $\lim_{k\to\infty} f(x_k) = f_\infty$, the values $B(x_k, \mu_k)$ are thus bounded from below.

Because $0 < \mu_{k+1} < \mu_k$ and $-\sum_{i=1}^m \ln c_i(x_k) > 0$ for k large enough, we have

$$-\mu_{k+1} \sum_{i=1}^m \ln c_i(x_k) < -\mu_k \sum_{i=1}^m \ln c_i(x_k), \quad \text{so that} \quad B(x_k, \mu_{k+1}) < B(x_k, \mu_k).$$

Using this relation and the minimizing property of x_{k+1} gives

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

Letting B_k denote $B(x_k, \mu_k)$, we see that the sequence $\{B_k\}$ is monotonically decreasing and bounded from below, and consequently must converge monotonically from above to a limit, say B_∞, where $B_\infty \ge f_\infty$ and f_∞ denotes $f(x_\infty)$.

We wish, finally, to show that $B_\infty = f_\infty$. Suppose instead that $B_\infty > f_\infty$ and define $\zeta = \frac{1}{2}(B_\infty - f_\infty)$, so that $\zeta > 0$. Because f is continuous, there would then be a neighborhood of x_∞ in which, for all x in the neighborhood,

$$f(x) \;\le\; f_\infty + \zeta \;=\; B_\infty - \zeta. \tag{3.14}$$

Consider a specific strictly feasible point $\bar{x}$ in this neighborhood; such a point must exist because x_∞ is either strictly feasible or lies in the closure of strict($\mathcal{F}$). By definition of x_k, $B(x_k, \mu_k)$ is the smallest value of $B(x, \mu_k)$ for all $x \in \text{strict}(\mathcal{F}) \cap S$, which means that

$$B(x_k, \mu_k) \;\le\; B(\bar{x}, \mu_k) \;=\; f(\bar{x}) - \mu_k \sum_{i=1}^m \ln c_i(\bar{x}). \tag{3.15}$$

Further, strict feasibility of $\bar{x}$ implies that $\sum_{i=1}^m \ln c_i(\bar{x})$ is finite. Because $\mu_k > 0$ and $\mu_k \to 0$, there must be an integer K such that, for $k \ge K$,

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

Combining this relation with the fact that $f(\bar{x}) \le B_\infty - \zeta$ (from (3.14)), it follows from (3.15) that

$$B(x_k, \mu_k) \;<\; B_\infty - \zeta + \tfrac{1}{2}\zeta \;=\; B_\infty - \tfrac{1}{2}\zeta,$$

which contradicts the monotonic convergence of $\{B_k\}$ to B_∞ from above. We conclude that $B_\infty = f_\infty$, verifying the second equality in result (v) when x_∞ is not strictly feasible. □

On the bright side, Theorem 3.10 appears to be, and is, very strong. It implies, in particular, that a barrier function method can converge to the solution of constrained problems for which the usual sufficient conditions (see section 2.4) do not hold. Barrier methods can converge, for example, when the constrained minimizer is not locally unique or does not satisfy a constraint qualification.

In the spirit of full disclosure, however, we need to note what has *not* been proved. Even within the compact set S whose existence is confirmed by the theorem, it is not necessarily true that every sequence of minimizers $\{y_k\}$ of the barrier function converges to a solution x^* of the original problem, even though y_k produces the smallest value of $B(x, \mu_k)$ in strict($\mathcal{F}$) $\cap S$; we know only that there is a convergent subsequence. Furthermore, it would be impossible in practice to guarantee that y_k

has been found rather than some other local minimizer of $B(x, \mu_k)$, and we must be content with sequences of *local* unconstrained minimizers of the barrier function. Unfortunately, these sequences may converge to points that are not local constrained minimizers.

This undesirable situation can happen when there are topologically inconsistent points. Consider the simple problem containing the badly behaved constraints (3.9) discussed earlier,

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

whose unique solution is obviously the point $x^* = -\gamma$. The barrier function $B(x, \mu)$ has two feasible minimizers:

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

For $\mu \to 0$, the sequence corresponding to the negative square root converges to $-\gamma$, the unique solution of the constrained problem. However, the minimizing sequence corresponding to the positive square root converges to the origin, which is a topologically inconsistent point but *not* a constrained minimizer.

Anomalous convergence in the barrier minimizers is not limited to topological inconsistencies. Consider, for instance, a problem given in [32, p. 399]:

$$\text{minimize} \quad x_1 + x_2 \quad \text{subject to} \quad x_1 x_2 \geq 0. \tag{3.16}$$

The feasible region consists of the upper right and lower left quadrants, and the objective function is unbounded below in the feasible region. Yet there is a minimizing sequence of the barrier function that converges to the origin. The failure does not result from topological inconsistency (see Definition 3.4), since the origin is not an interior point of $\mathcal{F}$. Rather, the failure can be attributed to the first peculiarity of the logarithmic interior function noted in section 3.2. The logarithmic interior function for the constraint $x_1 x_2 \geq 0$ is the same as for the two constraints $x_1 \geq 0$ and $x_2 \geq 0$, and the barrier function has no way to make the distinction. If the problem contained these two separate constraints, the origin would indeed be the (unique) solution.

3.4. The Barrier Trajectory. The major purpose of this section is to prove Theorem 3.12, which summarizes conditions under which a sequence $x(\mu)$ of barrier minimizers not only converges to x^* but also defines a differentiable path to x^*. Depending on the context, this path is called either the *central path* or the *barrier trajectory*; we shall use these terms more or less interchangeably. The term "central path" entered the broad optimization lexicon via the interior-point revolution, following from analysis of points that satisfy perturbed complementarity conditions for linear programs. It turns out that, modulo the presence of equality constraints in standard-form linear programs, these are precisely analogous to the relations that hold at barrier function minimizers as $\mu \to 0$—hence the older term "barrier trajectory," which tends to be used primarily when discussing nonlinear optimization. The central path can be defined without mentioning barrier functions, but in our view it is helpful to remember the connections.

The results in Theorem 3.12 were proved in [33, 110] for the case when the LICQ and strict complementarity hold at x^*. They were recently proved in [122] when only the MFCQ holds.

In sharp contrast to Theorem 3.10, which proved general results under a minimal set of assumptions, our focus now is on what happens under favorable circumstances,

and accordingly we impose conditions strong enough to ensure that x^* is an isolated local constrained minimizer. One of these, that the MFCQ holds at x^*, guarantees that x^* is in the closure of the strictly feasible region, as we now show.

LEMMA 3.11 (implication 4 of MFCQ). *If the MFCQ holds at a feasible point $\hat{x}$ that does not lie in* strict$(\mathcal{F})$*, then $\hat{x}$ lies in the closure of* strict$(\mathcal{F})$.

Proof. To show that $\hat{x}$ lies in the closure of strict$(\mathcal{F})$, we verify that every neighborhood of $\hat{x}$ contains strictly feasible points. It follows from continuity of $c(x)$ that the constraints inactive at $\hat{x}$ remain inactive in a sufficiently small neighborhood. Because the MFCQ holds at $\hat{x}$, there is a vector p such that $J_{\mathcal{A}}(\hat{x})p > 0$. By continuity of $c(x)$,

$$c_i(\hat{x} + \alpha p) = c_i(\hat{x}) + \alpha \nabla c_i(\hat{x})^T p + o(\alpha \|p\|).$$

Hence $c_i(\hat{x} + \alpha p) > 0$ for all sufficiently small α and all $i \in \mathcal{A}(\hat{x})$, and thus there are strictly feasible points in every neighborhood of $\hat{x}$. ∎

THEOREM 3.12 (properties of the central path/barrier trajectory). *Consider the problem of minimizing $f(x)$ subject to $c(x) \geq 0$. Let $\mathcal{F}$ denote the feasible region, and assume that the set* strict$(\mathcal{F})$ *of strictly feasible points is nonempty. Let x^* be a local constrained minimizer, with g^* denoting $g(x^*)$, J^* denoting $J(x^*)$, and so on, and let $\mathcal{A}$ denote $\mathcal{A}(x^*)$. Assume that the following sufficient optimality conditions hold at x^*:*

(a) *x^* is a KKT point, i.e., there exists a nonempty set $\mathcal{M}_\lambda$ of Lagrange multipliers λ satisfying*

$$\mathcal{M}_\lambda = \{\lambda : g^* = J^{*T}\lambda, \ \ \lambda \geq 0, \ \ \text{and} \ \ c(x^*) \cdot \lambda = 0\};$$

(b) *the MFCQ (Definition 2.10) holds at x^*, i.e., there exists p such that $J^*_{\mathcal{A}} p > 0$, where $J^*_{\mathcal{A}}$ denotes the Jacobian of the active constraints at x^*; and*

(c) *there exists $\omega > 0$ such that $p^T H(x^*, \lambda)p \geq \omega \|p\|^2$ for all $\lambda \in \mathcal{M}_\lambda$ and all nonzero p satisfying $g^{*T} p = 0$ and $J^*_{\mathcal{A}} p \geq 0$, where $H(x^*, \lambda)$ is the Hessian of the Lagrangian (2.11).*

Assume that a logarithmic barrier method is applied in which μ_k converges 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) *let $\{x^k\}$ denote such a convergent subsequence, with the obvious notation that c_i^k denotes $c_i(x^k)$, and so on. Then the sequence of barrier multipliers $\{\lambda^k\}$, whose ith component is μ_k / c_i^k, is bounded;*

(iii) $\lim_{k \to \infty} \lambda^k = \bar{\lambda} \in \mathcal{M}_\lambda$.

If, in addition, strict complementarity holds at x^*, i.e., there is a vector $\lambda \in \mathcal{M}_\lambda$ such that $\lambda_i > 0$ for all $i \in \mathcal{A}$, then

(iv) $\bar{\lambda}_{\mathcal{A}} > 0$;

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

(vi) *a unique, continuously differentiable vector function $x(\mu)$ of unconstrained minimizers of $B(x, \mu)$ exists for positive μ in a neighborhood of $\mu = 0$; and*

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

Proof. We prove this theorem in stages, starting with results (i)–(iii), which do not require strict complementarity.

The properties assumed about x^* ensure that it is an isolated constrained minimizer (Theorem 2.23). Because the MFCQ holds at x^*, we know from Lemma 3.11

that x^* is in the closure of strict($\mathcal{F}$). Theorem 3.10 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 $\lim_{k\to\infty} x^k = x^*$. As convenient, we denote quantities associated with x^k by a subscript or superscript k; the subscript i denotes the ith element. Thus, c_i^k denotes $c_i(x^k)$, ∇c_i^k denotes the ith row of the matrix J^k, and so on.

Each x^k is an unconstrained minimizer of $B(x, \mu_k)$, which means that the gradient of the barrier function vanishes at x^k:

$$g^k = \sum_{i=1}^{m} \nabla c_i^k \lambda_i^k, \quad \text{where} \quad \lambda_i^k = \frac{\mu_k}{c_i^k}. \tag{3.17}$$

Because $c^k > 0$ (from Theorem 3.10, result (ii)), λ_i^k is strictly positive for any $\mu_k > 0$.

Suppose that constraint i is *inactive* at x^*. Then, because x^k converges to x^*,

$$\lim_{k\to\infty} c_i^k = c_i(x^*) > 0, \quad \text{and hence} \quad \lim_{k\to\infty} \lambda_i^k = 0. \tag{3.18}$$

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

Otherwise, at least one constraint is active. Let the positive numbers ℓ^k and v_i^k be defined as

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

Note that $v_i^k > 0$ and $\sum_{i=1}^m v_i^k = 1$, so that $v_i^k \leq 1$. Since $\ell^k > 0$ and (3.17) holds at x^k, we have

$$\frac{1}{\ell^k} g^k - \sum_{i=1}^{m} \nabla c_i^k v_i^k = 0. \tag{3.19}$$

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

Now we show that $\limsup_{k\to\infty} \ell^k$, denoted by $\hat{\ell}$, must be finite. Suppose not, i.e., that $\{\ell^k\}$ is unbounded above, and consider (3.19) as $k \to \infty$. Because x^k converges to x^*, ∇c_i^k converges to ∇c_i^*. Under the assumption that ℓ^k is unbounded, the following relation must hold for any set $\{\widehat{v}_i\}$ of limit points of $\{v_i^k\}$:

$$\sum_{i=1}^{m} \nabla c_i^* \widehat{v}_i = 0, \quad \text{where} \quad \widehat{v}_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{m} \widehat{v}_i = 1. \tag{3.20}$$

Because $\widehat{v}_i = 0$ if constraint i is inactive at x^*, the first relation in (3.20) states that a nonnegative linear combination of the *active* constraint gradients at x^* is zero, with at least one positive coefficient, since $\widehat{v}$ satisfies the last condition in (3.20). However, since the MFCQ holds, we know from Lemma 2.12 that a nonnegative linear combination of the active constraint gradients can vanish only if *every* coefficient is zero—a contradiction, showing that $\hat{\ell}$ is finite.

Finiteness of $\hat{\ell}$ implies that each component λ_i^k is bounded for all k, proving (ii). As a result, the sequence $\{\lambda_i^k\}$ has at least one limit point, say $\bar{\lambda}_i$, with $\bar{\lambda}_i \geq 0$. It follows from convergence of x^k to x^*, (3.17), and (3.18) that $g^* = J^{*T}\bar{\lambda}$. Since $\bar{\lambda} \geq 0$ and $c^* \cdot \bar{\lambda} = 0$, it follows that $\bar{\lambda} \in \mathcal{M}_\lambda$, proving (iii).

We postpone the proofs of the remaining parts until after a short discussion of strict complementarity. ∎

3.5. Strict Complementarity. The stronger assumption of *strict complementarity* at x^* (Definition 2.6) means that there is a multiplier λ in $\mathcal{M}_\lambda$ with the property that $\lambda_i > 0$ for all $i \in \mathcal{A}$. When strict complementarity holds, Lemma 2.8 shows that condition (c) of Theorem 3.12 can be stated in a simpler form, where the definiteness of the Lagrangian Hessian is restricted only to the null space of $J_{\mathcal{A}}^*$.

THEOREM 3.12, condition (c′). *For all $\lambda \in \mathcal{M}_\lambda$, there exists $\omega > 0$ such that*

$$p^T H(x^*, \lambda) p \geq \omega \|p\|^2 \tag{3.21}$$

for all $\lambda \in \mathcal{M}_\lambda$ and all nonzero p satisfying $g^{*T} p = 0$ and $J_{\mathcal{A}}^* p = 0$.

This condition can be stated conveniently in terms of $N_{\mathcal{A}}$, a matrix whose columns form a basis for the null space of $J_{\mathcal{A}}^*$. The property (3.21) is equivalent to stating that

$$N_{\mathcal{A}}^T H(x^*, \mu) N_{\mathcal{A}} \text{ is positive definite,} \tag{3.22}$$

or, in words, that the reduced Hessian of the Lagrangian is positive definite.

To prove (iv) of Theorem 3.12, that the limiting multiplier estimate of the barrier method is strictly positive, we first prove an intermediate result: strict complementarity implies that x^k must converge to x^* at least as fast as μ_k converges to zero, namely,

$$\frac{\mu_k}{\|x^k - x^*\|} \geq M > 0 \quad \text{or, equivalently,} \quad \|x^k - x^*\| = O(\mu_k).$$

LEMMA 3.13 (implication 2 of strict complementarity). *Assume that the conditions of Theorem* 3.12 *are satisfied and, in addition, that strict complementarity holds at x^*. Then $\|x^k - x^*\| = O(\mu_k)$.*

Proof. This result will be proved by contradiction, by showing that the relationship $\mu_k / \|x^k - x^*\| \to 0$ cannot hold.

Let us define δ_k and s_k from

$$x^k = x^* + \delta_k = x^* + \|x^k - x^*\| s_k, \quad \text{so that } \delta_k = x^k - x^*, \; s_k = \frac{\delta_k}{\|\delta_k\|}, \tag{3.23}$$

and $\|s_k\| = 1$. Boundedness of $\|s_k\|$ means that there must be a convergent subsequence. Relabeling if necessary, we have $\lim_{k \to \infty} s_k = s$.

Because f and $\{c_k\}$ are twice-continuously differentiable, we have the usual first-order Taylor expansions around x^k and x^*:

$$f^* = f^k - \delta_k^T g_k + O(\|\delta_k\|^2) \quad \text{and} \quad f^k = f^* + \delta_k^T g^* + O(\|\delta_k\|^2); \tag{3.24}$$

$$c_i^* = c_i^k - \delta_k^T \nabla c_i^k + O(\|\delta_k\|^2) \quad \text{and} \quad c_i^k = c_i^* + \delta_k^T \nabla c_i^* + O(\|\delta_k\|^2). \tag{3.25}$$

Let $\mathcal{A}$ denote $\mathcal{A}(x^*)$. Since $c_i^k > 0$, it must be true that $(\nabla c_i^k)^T s_k \geq 0$ for all $i \in \mathcal{A}$, so that $J_{\mathcal{A}}^* s \geq 0$.

The desired result follows from a combination of mean-value theorems for scalar-valued functions, properties of the barrier minimizers, and strict complementarity.

To begin, we obtain a relationship between the difference in objective values $f^k - f^*$ and the sizes of μ_k and $\|x^k - x^*\|$ by writing

$$f^k - f^* = f^k - f^* - \sum_{i=1}^m \lambda_i^k (c_i^k - c_i^*) + \sum_{i=1}^m \lambda_i^k c_i^k - \sum_{i=1}^m \lambda_i^k c_i^*. \tag{3.26}$$

Using (3.24) and (3.25) gives

$$f^k - f^* = \delta_k^T \left(g_k - \sum_{i=1}^m \lambda_i^k \nabla c_i^k \right) + \sum_{i=1}^m \lambda_i^k c_i^k - \sum_{i=1}^m \lambda_i^k c_i^* + O\big(\|\delta_k\|^2\big). \tag{3.27}$$

By definition of x_k, $g_k - \sum_{i=1}^m \lambda_i^k \nabla c_i^k = 0$ and $\sum_{i=1}^m \lambda_i^k c_i^k = m\mu_k$. Further, since $c_i^* = 0$ for $i \in \mathcal{A}$ and $\lambda_i^k = O\big(\mu_k\big)$ for $i \notin \mathcal{A}$, relation (3.27) may be written as

$$f^k - f^* = O\big(\mu_k\big) + O\big(\|x^k - x^*\|^2\big).$$

Dividing by $\|x^k - x^*\|$, we have

$$\frac{f^k - f^*}{\|x^k - x^*\|} = \frac{O\big(\mu_k\big)}{\|x^k - x^*\|} + O\big(\|x^k - x^*\|\big). \tag{3.28}$$

Now we derive a second expression for $f^k - f^*$ using a mean-value theorem:

$$f^k - f^* = g(x^* + \theta\delta_k)^T \delta_k = \|x^k - x^*\| \; g(x^* + \theta\delta_k)^T s_k$$

for some θ satisfying $0 \le \theta \le 1$. Combining this with (3.28), we obtain

$$\frac{f^k - f^*}{\|x^k - x^*\|} = g(x^* + \theta\delta_k)^T s_k = \frac{O\big(\mu_k\big)}{\|x^k - x^*\|} + O\big(\|x^k - x^*\|\big). \tag{3.29}$$

Note that (3.28) and (3.29) are valid without strict complementarity.

Now we make the assumption (to be proved not to hold) that $\mu_k / \|x^k - x^*\| \to 0$ as $k \to \infty$. In this case, taking the limit in (3.29) and invoking continuity of g, it follows that

$$\lim_{k\to\infty} \frac{f^k - f^*}{\|x^k - x^*\|} = 0 \quad \text{and} \quad g^{*T} s = 0. \tag{3.30}$$

This relation is interesting because we have assumed strict complementarity, and we know that $J_{\mathcal{A}}^* s \ge 0$. Hence we can apply Lemma 2.8, which says that s satisfies $g^{*T} s = 0$ and $J_{\mathcal{A}}^* s \ge 0$ only if $J_{\mathcal{A}}^* s = 0$. Hence, if $\mu_k = o\big(\|x^k - x^*\|\big)$, it must be true that $J_{\mathcal{A}}^* s = 0$.

We now examine the consequences of the property that $J_{\mathcal{A}}^* s = 0$. Another version of the mean-value theorem, this time applied to the Lagrangian function expanded around x^* for a fixed value of its second argument, tells us that

$$\nabla L(x^k, \bar\lambda)^T \delta_k = \nabla L(x^*, \bar\lambda)^T \delta_k + \tfrac{1}{2} \delta_k^T \nabla^2 L(x^* + \theta\delta_k, \bar\lambda) \delta_k$$

for some $\theta \in [0, 1]$, where $\bar\lambda$ is the limit point of $\{\lambda^k\}$ whose existence was proved earlier. Since $\bar\lambda \in \mathcal{M}_\lambda$, we know that $\nabla L(x^*, \bar\lambda) = 0$, and thus

$$\nabla L(x^k, \bar\lambda)^T \delta_k = \tfrac{1}{2} \delta_k^T \nabla^2 L(x^* + \theta\delta_k, \bar\lambda) \delta_k.$$

Recall that $\delta_k = s_k \|\delta_k\|$, so that this expression may be written as

$$\|\delta_k\| \; \nabla L(x^k, \bar\lambda)^T s_k = \tfrac{1}{2} \|\delta_k\|^2 \; s_k^T \nabla^2 L(x^* + \theta\delta_k, \bar\lambda) s_k.$$

As $k \to \infty$, $s_k \to s$, and we have shown that $J_{\mathcal{A}}^* s = 0$ if $\mu_k = o\big(\|x^k - x^*\|\big)$. Because of this property, positive-definiteness of the Hessian of the Lagrangian along directions

orthogonal to $J_{\mathcal{A}}^*$ (see (3.21)), continuity of $\nabla^2 L$, and the fact that $\|s_k\| = 1$, we conclude that there is a positive number ω such that

$$\nabla L(x^k, \bar{\lambda})^T s_k \geq \omega \|\delta_k\|.$$

Dividing by $\|\delta_k\|$, we obtain

$$\frac{\nabla L(x^k, \bar{\lambda})^T s_k}{\|\delta_k\|} \geq \omega. \tag{3.31}$$

We now derive a different form for $\nabla L(x^k, \bar{\lambda})^T s_k$ by noting that $g^k = \sum \lambda_i^k \nabla c_i^k$, which means that

$$\nabla L(x^k, \bar{\lambda}) = g^k - \sum_{i=1}^{m} \bar{\lambda}_i \nabla c_i^k = \sum_{i=1}^{m} (\lambda_i^k - \bar{\lambda}_i) \nabla c_i^k. \tag{3.32}$$

For $i \notin \mathcal{A}$, $\lambda_i^k = O\big(\mu_k\big)$ and $\bar{\lambda}_i = 0$. Consequently, as $k \to \infty$,

$$\sum_{i \notin \mathcal{A}} (\lambda_i^k - \bar{\lambda}_i) \nabla c_i^k = O\big(\mu_k\big).$$

For $i \in \mathcal{A}$, we use the expansion $\nabla c_i^k = \nabla c_i^* + O\big(\delta_k\big)$. Since λ_i^k is bounded and converging to $\bar{\lambda}$, we have

$$\sum_{i \in \mathcal{A}} (\lambda_i^k - \bar{\lambda}_i) \nabla c_i^k = \sum_{i \in \mathcal{A}} o\big(1\big) \nabla c_i^* + o\big(\delta_k\big).$$

Combining these results for active and inactive constraints in (3.32) thus gives, for sufficiently large k,

$$\nabla L(x^k, \bar{\lambda}) = O\big(\mu_k\big) + \sum_{i \in \mathcal{A}} o\big(1\big) \nabla c_i^* + o\big(\delta_k\big).$$

Recalling the (temporary) conclusion that $J_{\mathcal{A}}^* s = 0$, where s is the limit of the unit vectors $\delta_k / \|\delta_k\|$, the inner product of $\nabla L(x^k, \bar{\lambda})$ and s satisfies

$$\nabla L(x^k, \bar{\lambda})^T s = O\big(\mu_k\big) + o\big(\delta_k\big), \quad \text{so that } \frac{\nabla L(x^k, \bar{\lambda})^T s}{\|\delta_k\|} = \frac{O\big(\mu_k\big)}{\|\delta_k\|} + o\big(1\big).$$

Because we have assumed that $\mu_k / \|\delta_k\| \to 0$, it follows that

$$\frac{\nabla L(x^k, \bar{\lambda})^T s}{\|\delta_k\|} \to 0.$$

But this contradicts (3.31), which asserts that the same ratio is strictly positive in the limit. We conclude that the relation $\mu_k = o\big(\|x^k - x^*\|\big)$ does not hold, thereby proving that $\|x^k - x^*\| = O\big(\mu_k\big)$. □

Having shown that $\mu_k = O\big(\|x^k - x^*\|\big)$, we proceed to show that the barrier multipliers converge to a strictly complementary multiplier.

THEOREM 3.12, result (iv). *If strict complementarity holds, then* $\bar{\lambda}_{\mathcal{A}} > 0$.

Proof. Let us define t_k as

$$t_k = \frac{x^k - x^*}{\mu_k}, \text{ so that } x^k = x^* + \mu_k t_k, \text{ with } \|t_k\| = \frac{\|x^k - x^*\|}{\mu_k}. \tag{3.33}$$

Lemma 3.13 has just shown that $\mu_k/\|x^k - x^*\| \geq M > 0$, which implies that $\|t_k\|$ is bounded, i.e., that $\|t_k\| \leq M_1$ for some constant M_1. Boundedness of $\|t_k\|$ means that there is a convergent subsequence, whose limit will be denoted by t, with $\lim_{k\to\infty} t_k = t$, relabeling the indices if necessary.

We show next that $\|t_k\|$ is bounded away from zero by expanding the ith constraint around x^* for $i \in \mathcal{A}(x^*)$. Since $c_i^* = 0$, $c(\cdot)$ is smooth, and $c_i^k \geq 0$, we have

$$c_i^k = \mu_k(\nabla c_i^k)^T t_k + O\big(\mu_k^2\|t_k\|^2\big) \quad \text{for } i \in \mathcal{A}(x^*), \tag{3.34}$$

where we know that $(\nabla c_i^k)^T t_k \geq 0$. Using the definition $\lambda_i^k = \mu_k/c_i^k$, it follows that

$$\lambda_i^k = \frac{1}{(\nabla c_i^k)^T t_k + \mu_k O\big(\|t^k\|^2\big)} \quad \text{for } i \in \mathcal{A}(x^*). \tag{3.35}$$

Because λ_i^k is bounded for all i, as shown previously, the denominator of (3.35) must be bounded away from zero, showing that $\|t_k\|$ cannot converge to zero, i.e.,

$$\|t_k\| \geq M_2 > 0. \tag{3.36}$$

Since $\{t_k\}$ is bounded and $\mu_k \to 0$, the term $\mu_k O\big(\|t^k\|^2\big)$ converges to zero. Boundedness of λ_i^k then implies that $(\nabla c_i^k)^T t_k$ cannot converge to zero for $i \in \mathcal{A}(x^*)$. Thus $J_{\mathcal{A}}^* t > 0$.

Putting together the result that $0 < M_2 \leq \|t_k\| \leq M_1$ and the fact that $\mu_k \to 0$, we see that the limiting barrier multiplier $\bar{\lambda}_i$ must satisfy

$$\bar{\lambda}_i = \frac{1}{(\nabla c_i^*)^T t} \quad \text{for } i \in \mathcal{A}(x^*). \tag{3.37}$$

Because $J_{\mathcal{A}}^* t > 0$, it follows that $\bar{\lambda}_i > 0$ if constraint i is active. This proves (iv) of Theorem 3.12. ∎

There is an immediate and useful corollary.

COROLLARY 3.14. *Under assumptions* (a)–(c) *of Theorem* 3.12 *and the added assumption of strict complementarity at* x^*, $\|x^k - x^*\| = \Theta(\mu_k)$.[4]

Proof. Lemma 3.13 has just verified that $\|x^k - x^*\| = O\big(\mu_k\big)$ when strict complementarity holds. The lower bound (3.36) on $\|t_k\|$ shows that $\mu_k = O\big(\|x^k - x^*\|\big)$, and the desired result is immediate. ∎

In [122], it is shown that $\bar{\lambda}_{\mathcal{A}}$ converges to the analytic center of the set of strictly complementary multipliers, i.e.,

$$\bar{\lambda}_{\mathcal{A}} = \arg\max \left\{ \sum_{i\in\mathcal{A}} \ln \lambda_i : \; g^* = J_{\mathcal{A}}^{*T}\lambda_{\mathcal{A}} \quad \text{and} \quad \lambda_{\mathcal{A}} > 0 \right\}. \tag{3.38}$$

We note for future reference that, following directly from (3.37),

$$(\nabla c_i^*)^T t = \frac{1}{\bar{\lambda}_i} \quad \text{for } i \in \mathcal{A}(x^*), \quad \text{so that} \quad J_{\mathcal{A}}^* t = 1 \mathbin{./} \bar{\lambda}_{\mathcal{A}}. \tag{3.39}$$

This relation is the foundation of the well-known nontangential approach of the barrier trajectory to x^*; see Figure 4.

Given that $\lambda_i^k > 0$ for all $i \in \mathcal{A}(x^*)$, we proceed to verify the remaining results (v)–(vii) of Theorem 3.12.

[4] See Definition A.9 for an explanation of the notation $\Theta(\cdot)$.

THEOREM 3.12, results (v)–(vii). *If strict complementarity holds, then*

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

(vi) *a unique, continuously differentiable vector function $x(\mu)$ of unconstrained minimizers of $B(x, \mu)$ exists for positive μ in a neighborhood of $\mu = 0$;*

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

Proof. The Hessian of the barrier function is

$$\nabla^2 B(x, \mu) = \nabla^2 f(x) - \sum_{i=1}^{m} \frac{\mu}{c_i(x)} \nabla^2 c_i(x) + J(x)^T C(x)^{-2} J(x). \tag{3.40}$$

Asymptotically, as $\mu_k \to 0$, $\nabla^2 B(x^k, \mu_k)$ approaches the sum of the Hessian of the Lagrangian function (with multiplier $\bar{\lambda}$) and a "large" matrix in the range space of the active constraint gradients. The crucial element in verifying positive-definiteness is positivity of the barrier multipliers corresponding to active constraints, which implies that, in effect, the range space of $J_{\mathcal{A}}^*$ is fully represented in the barrier Hessian.

We know that λ_i^k converges to a positive constant for all $i \in \mathcal{A}$, where $\mathcal{A}$ denotes $\mathcal{A}(x^*)$. Letting d_i^k denote the ratio $\mu_k/(c_i^k)^2 = \lambda_i^k / c_i^k$, it follows that, for $i \in \mathcal{A}$, d_i^k is unbounded:

$$\liminf_{k \to \infty} \; d_i^k = \liminf_{k \to \infty} \; \frac{\mu_k}{(c_i^k)^2} = \infty \quad \text{for } i \in \mathcal{A}. \tag{3.41}$$

For $i \notin \mathcal{A}$, $d_i^k \to 0$.

Denoting $\nabla^2 B(x^k, \mu_k)$ for compactness by H_B^k, we rewrite (3.40) as

$$H_B^k = \nabla^2 B(x^k, \mu_k) = H_0^k - \sum_{i=1}^{m} \frac{\mu_k}{c_i^k} \; H_i^k + \mu_k J_k^T C_k^{-2} J_k,$$

where H_0 denotes $\nabla^2 f$ and H_i denotes $\nabla^2 c_i$. To determine the properties of H_B^k as $k \to \infty$, we write it as

$$\begin{aligned} &H_B^k = H^* + M^* + M_1^k + M_2^k + M_3^k, \;\; \text{with} \\ &H^* = H_0^* - \sum_{i=1}^{m} \bar{\lambda}_i H_i^* \quad \text{and} \quad M^* = \gamma \sum_{i \in \mathcal{A}} \nabla c_i^* \nabla c_i^{*T} = \gamma J_{\mathcal{A}}^{*T} J_{\mathcal{A}}^*, \end{aligned} \tag{3.42}$$

where the positive constant γ will be defined later. The remaining three matrices in (3.42) are expressed as differences involving x_k, x^*, and γ:

$$\begin{aligned} &M_1^k = H_0^k - H_0^* - \sum_{i=1}^{m} \lambda_i^k \; H_i^k + \sum_{i=1}^{m} \bar{\lambda}_i H_i^*, \quad M_2^k = \gamma (J_{\mathcal{A}}^{k\,T} J_{\mathcal{A}}^k - J_{\mathcal{A}}^{*T} J_{\mathcal{A}}^*), \;\; \text{and} \\ &M_3^k = \sum_{i \in \mathcal{A}} (d_i^k - \gamma) \nabla c_i^k (\nabla c_i^k)^T + \sum_{i \notin \mathcal{A}} d_i^k \nabla c_i^k (\nabla c_i^k)^T. \end{aligned}$$

The matrix H^* is the Hessian of the Lagrangian function at $(x^*, \bar{\lambda})$. For sufficiently large k, the matrices M_1^k and M_2^k become arbitrarily small in norm; this follows from smoothness of f and $\{c_i\}$, convergence of x_k to x^*, convergence of λ_k to $\bar{\lambda}$, and (as we show later) boundedness of γ. Because d_i^k is unbounded above for $i \in \mathcal{A}$ (see (3.41)), the quantity $(d_i^k - \gamma)$ is positive for sufficiently large k; hence M_3^k is the sum of two positive semidefinite matrices and must itself be positive semidefinite.

Positive-definiteness of H_B^k follows if $H^* + \gamma J_{\mathcal{A}}^{*T} J_{\mathcal{A}}^*$ is guaranteed to be positive definite for some sufficiently large constant γ. Because strict complementarity holds, we invoke assumption (c′) (see (3.21)), which is that $p^T H^* p > 0$ for all p satisfying $J_{\mathcal{A}}^* p = 0$. Positive-definiteness of the barrier Hessian then follows directly from Debreu's lemma (Lemma A.14), which states that $p^T H^* p > 0$ for all p satisfying $J_{\mathcal{A}}^* p = 0$ if and only if there is a finite $\bar{\rho} > 0$ such that $H^* + \rho J_{\mathcal{A}}^T J_{\mathcal{A}}$ is positive definite for all $\rho \geq \bar{\rho}$. Thus result (v) is proved, showing that $\nabla^2 B(x^k, \mu_k)$ is positive definite and that x^k is an *isolated* unconstrained minimizer of $B(x, \mu_k)$ (Lemma A.7).

To verify the existence of a unique, differentiable function $x(\mu)$ for positive μ in a neighborhood of $x(\mu_k)$, we apply the implicit function theorem (see, for example, [81, p. 128] and [80, pp. 585–586]) to the $n+1$ variables (x, μ). At (x^k, μ_k), we know that the following system of nonlinear equations has a solution:

$$\Phi(x, \mu) = g(x) - \mu\, J_{\mathcal{A}}^T(x) C(x)^{-1} e.$$

The Jacobian of Φ with respect to x is the barrier Hessian H_B, 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 $x(\mu_k) \triangleq x^k$ such that $\Phi(x(\mu), \mu) = 0$ for all positive μ in a neighborhood 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 (vi).

Result (vii) is immediate from the local uniqueness of $x(\mu)$ and result (i), that x_k is a local unconstrained minimizer of the barrier function. □

Local uniqueness of the sequence $\{x^k\}$ has just been proved; let t denote the limit of the sequence $\{t_k\}$ defined in (3.33). For nonlinearly constrained problems, the nature of t, the tangent to the barrier trajectory at x^*, has been considered by various authors, including [109] (assuming the LICQ) and [122] (assuming the MFCQ). We mention two of its most interesting properties.

1. For any active constraint i, we know from (3.39) that $\nabla c_i^{*T} t = 1/\bar{\lambda}_i$. Letting θ_i denote the angle between t and ∇c_i^*, it follows that

$$\cos\theta_i \sim \frac{1}{\|\nabla c_i^*\| \bar{\lambda}_i}. \tag{3.43}$$

If all active constraint gradients are approximately equal in norm, relation (3.43) shows that the approach of the barrier trajectory to x^* is "closer to tangential" for active constraints with larger multipliers. This phenomenon will be illustrated in Figure 4.

2. Treating $x(\mu)$ as a differentiable function of μ for positive μ in the neighborhood of $\mu = 0$, one can derive the relation

$$H(x^*, \lambda^*) t = J_{\mathcal{A}}^{*T} w + J_{\text{in}}^{*T} (C_{\text{in}}^*)^{-1} e, \tag{3.44}$$

where J_{in}^* denotes the Jacobian of the inactive constraints at x^*, C_{in}^* denotes the diagonal matrix of inactive constraint values at x^*, and w represents the rate of change along t with respect to μ of the barrier multipliers corresponding to active constraints. Letting $N_{\mathcal{A}}$ denote a basis for the null space of $J_{\mathcal{A}}^*$ and multiplying (3.44) by $N_{\mathcal{A}}^T$, we have

$$N_{\mathcal{A}}^T H(x^*, \lambda^*) t = N_{\mathcal{A}}^T J_{\text{in}}^{*T} (C_{\text{in}}^*)^{-1} e. \tag{3.45}$$

Let us write $t = J_{\mathcal{A}}^{*T} t_{\mathcal{A}} + N_{\mathcal{A}} t_N$, splitting t into components lying in the orthogonal subspaces range$(J_{\mathcal{A}}^{*T})$ and null$(J_{\mathcal{A}}^*)$. Relation (3.39) states that $J_{\mathcal{A}}^* t = 1 \mathbin{./} \lambda_{\mathcal{A}}^*$, which

uniquely specifies $J_{\mathcal{A}}^{*T} t_{\mathcal{A}}$ (but not the vector $t_{\mathcal{A}}$, since $J_{\mathcal{A}}^*$ does not necessarily have full row rank). Substituting for t in (3.45), we obtain

$$N_{\mathcal{A}}^T H(x^*, \lambda^*) N_{\mathcal{A}} t_N = N_{\mathcal{A}}^T J_{\text{in}}^{*T} (C_{\text{in}}^*)^{-1} e - N_{\mathcal{A}}^T H(x^*, \lambda^*) J_{\mathcal{A}}^{*T} t_{\mathcal{A}}.$$

Our assumptions of second-order sufficiency and its special form due to strict complementarity (see (3.21) and (3.22)) imply that $N_{\mathcal{A}}^T H(x^*, \lambda^*) N_{\mathcal{A}}$ is positive definite, so that t_N is uniquely determined. Further details about t are given in [122].

3.6. The Effects of Null Multipliers. When a null multiplier is present, i.e., if there is an $i \in \mathcal{A}(x^*)$ such that $\lambda_i = 0$ for all $\lambda \in \mathcal{M}_\lambda$, the properties of the barrier trajectory change dramatically.

The situation in which the LICQ holds at x^* and there is a null multiplier was considered in [61]; the case of null multipliers when x^* satisfies the MFCQ is analyzed in detail in [122]. We summarize only the main points of interest, which are proved in [122].

At a critical juncture in the proof of Lemma 3.13, strict complementarity was needed to prove that $J_{\mathcal{A}}^* t > 0$. This eventually allowed us to show that $\|x^k - x^*\| = \Theta(\mu_k)$, the powerful result ensuring that (for example) the constraint values and barrier multipliers at x_k differ from their optimal values by approximately a constant times μ. Sadly, this property is destroyed if there is even a single null multiplier. Consider the very simple example of minimizing $\frac{1}{2}x^2$ subject to $x \geq 0$. The unique optimal solution is $x^* = 0$, the constraint is active, the LICQ holds (trivially), and $\lambda^* = 0$. The barrier function is

$$B(x, \mu) = \tfrac{1}{2}x^2 - \mu \ln x, \quad \text{with} \quad x(\mu) - x^* = \sqrt{\mu}.$$

This example suggests (correctly) the general property that

$$\mu_k = \Theta(\|x^k - x^*\|^2);$$

see [122] for details. In the null-multiplier case when the MFCQ holds, the barrier trajectory becomes tangent to the strongly active constraints (those with nonnull multipliers); see Figure 4. Further, an example in [69] illustrates that in this case the barrier multipliers do not necessarily converge to the analytic center of the set of strongly complementary multipliers, as they do (3.38) with strict complementarity.

3.7. More on the Central Path/Barrier Trajectory. The most convincing way to understand barrier trajectories is to see them. We now give examples to illustrate the differences in the central path's approach to the solution when strict complementarity does, and does not, hold.

An instance of the first is

$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \quad \tfrac{1}{2}(x_1^2 + x_2^2) \quad \text{subject to} \quad \tfrac{5}{4}x_1 + x_2 \geq 4 \quad \text{and} \quad x_2 \geq 2, \tag{3.46}$$

where both constraints are linear. The unique solution is $x^* = (1.6, 2)$. Both constraints are active at x^*, the LICQ holds, and the optimal multiplier is $\lambda^* = (1.28, 0.72)^T$. The left side of Figure 4 shows the contours of the objective function, the two constraints, and the cerulean trajectory of barrier minimizers converging to x^*. It is evident that the barrier trajectory is approaching the solution nontangentially to both active constraints. In contrast, the problem

$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \quad \tfrac{1}{2}(x_1^2 + x_2^2) \quad \text{subject to} \quad x_1 + x_2 \geq 4 \quad \text{and} \quad x_2 \geq 2, \tag{3.47}$$

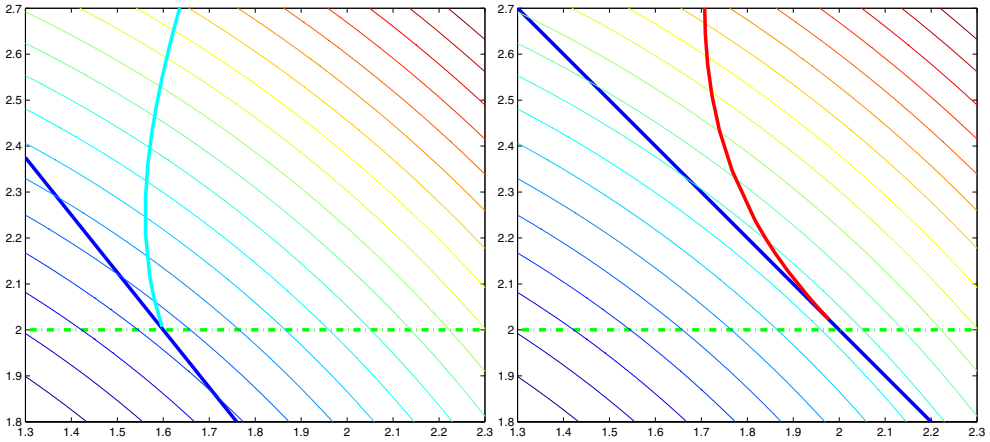

Fig. 4 *On the left, contours of $f(x)$ and the boundary of the feasible region are shown for problem* (3.46)*, along with the nontangential approach of the barrier trajectory. On the right, the tangential approach of the path of barrier minimizers is shown for the null-multiplier problem* (3.47).

stated earlier as (2.6), is the same as (3.46) except that one of the constraints has been perturbed so that it has a null multiplier. (Note that the LICQ holds at the solution for both problems.) For (3.47), the red trajectory of barrier minimizers approaches the strongly active constraint (the one with the nonnull multiplier) tangentially.

Appealing as the results of Theorem 3.12 are, we would be remiss to suggest that the barrier trajectory always behaves as well as in the examples depicted so far. As one would expect, analysis of the central path becomes more complicated as we increase the generality of the functions involved. For the case of linear programming, the central path is ultimately smooth and its approach to the optimal face is now completely understood (see, for example, [55]). For convex problems satisfying the LICQ, the barrier trajectory converges to the analytic center of the set of optimal solutions [70], but the very recent examples of [48] show that the central path can behave in strange and unexpected ways even for infinitely smooth convex functions.

Despite these caveats, the "centered" nature of the barrier trajectory, which provides at its best a differentiable, noncombinatorial approach to the solution, is arguably the great strength of barrier and barrier-related methods such as primal-dual methods (see section 5). Accordingly we present a different perspective on the central path as a preview of things to come.

At x_μ, the barrier gradient vanishes, which means that

$$\nabla B(x_\mu, \mu) = g(x_\mu) - \mu J(x_\mu)^T C(x_\mu)^{-1} e = 0. \tag{3.48}$$

This relationship explicitly involves only x, the so-called *primal variables*. (Note the linear programming terminology.) An alternative interpretation can be derived by defining m new independent variables λ and writing (3.48) as $n + m$ nonlinear equations in x and λ:

$$\begin{pmatrix} g(x) - J(x)^T \lambda \\ C(x)\lambda - \mu e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{3.49}$$

Recalling that the barrier multipliers λ_μ are defined as $\mu \cdot / c(x_\mu)$, it is easy to see that (x_μ, λ_μ) satisfy the nonlinear equations (3.49). Conversely, given any solution (x, λ) of (3.49), the associated vector x is a stationary point of the barrier function with parameter μ.

We mentioned earlier (see (3.8)) that the condition $c(x_\mu) \cdot \lambda_\mu = \mu$ represented in the second equation of (3.49) can be interpreted as "perturbed complementarity" because, as $\mu \to 0$, this relation corresponds to an increasingly accurate approximation of the complementarity condition $c(x^*) \cdot \lambda^* = 0$ that holds at x^*. An interesting historical note is that Fiacco and McCormick [33, p. 40] present the perturbed optimality conditions (3.49) as one possible motivation for the logarithmic barrier function.

We caution the reader that the *primal-dual* interpretation reflected in equations (3.49), where x are the primal variables and π the dual variables, captures only some of the properties of x_μ and λ_μ. In particular, the nonnegativity of c and π is not represented, nor are any second-order conditions associated with *minimizing* the barrier function (rather than simply finding a stationary point). We shall return later to these issues.

4. Newton's Method and Barrier Functions. In developing interior methods for nonlinearly constrained optimization, several aspects of Newton's method specialized to barrier functions will be important.

4.1. Background on Newton's Method. We briefly review the use of Newton's method for solving both nonlinear equations and unconstrained optimization problems.

For the nonlinear equations $F(x) = 0$, where F is a continuously differentiable function from $\mathbb{R}^n$ to $\mathbb{R}^n$, the *Newton step* p_k from an iterate x_k where $F'(x_k)$ is nonsingular is defined as the step to the zero of the local linear Taylor-series model of F:

$$F'(x_k)p_k = -F(x_k). \tag{4.1}$$

Similarly, Newton's method for minimizing a twice-continuously differentiable function $f(x)$ is based on the local quadratic Taylor-series model

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

where g_k denotes $\nabla f(x_k)$. If $\nabla^2 f(x_k)$ is positive definite, this quadratic function has a unique minimizer at $x_k + p_k$, where p_k satisfies

$$\nabla^2 f(x_k)p_k = -g_k. \tag{4.2}$$

In both (4.1) and (4.2), the iterates in a *pure* Newton method are $x_{k+1} = x_k + p_k$, and under well-known conditions these iterates converge quadratically to a zero of $F(x)$ or a minimizer of $f(x)$ (see, e.g., [81]).

Although Newton's method is famous for its fast local convergence, safeguards need to be included in any practical Newton method to encourage convergence from a generic starting point. A common globalization strategy in Newton-based methods for both nonlinear equations and unconstrained minimization is to include a *line search*, so that the new iterate is defined by

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

with $\alpha_k > 0$, where α_k is chosen to ensure that some suitable *merit function* is reduced by the move to x_{k+1}.

A further issue in line-search methods for unconstrained minimization is what to do when the current Hessian is not positive definite. Several approaches are widely used, among the most popular of which are so-called modified Newton methods in which p_k satisfies the linear system

$$\big(\nabla^2 f(x_k) + E\big)p_k = -g_k, \tag{4.4}$$

where E is a positive semidefinite matrix, usually diagonal, chosen so that $\nabla^2 f(x_k)+E$ is positive definite. An alternative globalization strategy in unconstrained optimization, embodied in *trust-region methods*, is to choose the step p_k to minimize the local Taylor-series model subject to a limitation on the size of $\|p\|$, where various norms may appear in the trust-region constraint (which, as its name suggests, specifies the local region in which the Taylor-series model can be trusted). See, e.g., [28, 34, 52, 77, 80] for material on line-search methods, and [25, 28, 34, 80] for discussion of trust-region methods.

The presentation in section 3.3 focused on unconstrained minimizers of the log barrier function; hence we consider how those points might be found.

4.2. The Classical Newton Barrier Method. A direct translation of the theory from section 3 into practice leads to a method in which minimizers x_μ of the barrier function are computed for a sequence of positive barrier parameters μ converging to zero. Such a method is structured into inner and outer iterations, where the inner iterations apply some variant of Newton's method to compute an unconstrained minimizer of $B(x,\mu)$ for a fixed value of μ, and the outer iterations test for convergence and adjust μ.

In recent algorithms, the idea is to improve efficiency by performing only an inexact minimization of the barrier function for each particular μ. With such a strategy, Newton iterations are executed until a suitable measure of improvement has been achieved; the barrier parameter is then reduced and the process repeated. The hope is that only a very small number of inner iterations (perhaps even one) will be needed for each value of μ.

The $n \times n$ *classical Newton barrier equations* $\nabla^2 B p = -\nabla B$ (see (4.2)) are

$$\left(\nabla^2 f - \sum_{i=1}^{m} \frac{\mu}{c_i}\nabla^2 c_i + \mu J^T C^{-2} J\right) p = -g + \mu J^T C^{-1} e, \tag{4.5}$$

where all arguments are evaluated at the current point. We call this the "classical" method because it was the form used in the 1960s; it is sometimes called a "primal" method because we are iterating solely on the original problem variables x. This system can be simplified in form by introducing an auxiliary m-vector $\pi(x,\mu)$, which can be viewed as a Lagrange multiplier estimate defined at an arbitrary strictly feasible point.

DEFINITION 4.1 (primal multipliers). *At any strictly feasible point x, the vector $\pi(x,\mu) = \mu \cdot/ c(x)$ is known as the vector of* primal multipliers. *The dependence of π on μ may be omitted if μ is obvious.*

For any sequence $\{x_k\}$ converging to x_μ, it must hold that $\lim_{k\to\infty} \pi(x_k,\mu) = \lambda_\mu$. Expressed in terms of π, the barrier gradient and Hessian take the forms

$$\nabla B(x,\mu) = g(x) - J(x)^T \pi(x,\mu) \quad \text{and} \tag{4.6a}$$

$$\nabla^2 B(x,\mu) = H\big(x,\pi\big) + J(x)^T\, \Pi(x,\mu)\, C(x)^{-1} J(x), \tag{4.6b}$$

where $H(x, \pi)$ is the Hessian of the Lagrangian evaluated with π as the multipliers, and $\Pi = \text{diag}(\pi_1, \pi_2, \ldots, \pi_m)$. These relations indicate that derivatives of the barrier function are intimately related to those of the Lagrangian evaluated with the primal multipliers.

At least in theory, the subproblem of minimizing $B(x, \mu)$ can be solved through a line-search or trust-region Newton method based on a quadratic model in which the gradient and Hessian are given by (4.6). However, several practical issues arise in applying a general-purpose unconstrained technique to the barrier subproblem.

A significant point that might be overlooked because of our emphasis on unconstrained barrier subproblems is that, although x_μ is a *locally unconstrained* minimizer of the barrier function (see result (ii) of Theorem 3.10), the barrier function cannot be treated without constraints. Since $B(x, \mu)$ exists only at strictly feasible points, there is always an *implicit* constraint $c(x) > 0$ that is not stated formally, but that nonetheless greatly influences the nature and design of a successful interior method.

For linear and quadratic programming, in which all the constraints are linear, it is possible to calculate exactly the closest point along p where any constraint becomes active; permissible steps are then restricted to be strictly less than this value. (This is the source of the frequently mentioned "0.99" rule in interior methods, meaning that the initial step is taken as 99% of the distance to the boundary, thereby retaining strict feasibility for linear constraints.) With nonlinear constraints, the step to the boundary cannot be calculated explicitly, and in any case its determination may require additional evaluations of the constraint functions.

Starting at a strictly feasible point x, it is straightforward to reduce α in a line-search method until $x + \alpha p$ is strictly feasible. But since most general-purpose line-search methods rely on quadratic, inverse quadratic, or cubic interpolation [72], they may not be well suited to the extreme behavior of barrier functions, e.g., the sharp rise near the boundary. Furthermore, a highly nonlinear constraint can be violated at points lying between two strictly feasible points—a property that creates additional complications for a standard line search. Various special-purpose line searches have been proposed for use in this context (see, e.g., [37, 74]). With a trust-region method, the discovery that the current trust region includes infeasible points typically requires formulation and solution of a new trust-region subproblem.

4.3. Properties of the Classical Newton Barrier Method.

4.3.1. Condition of the Barrier Hessian. As sketched in section 1.2, the barrier Hessian $\nabla^2 B$ is well known, even notorious, for being increasingly ill-conditioned and asymptotically singular as $\mu \to 0$. This property was observed for points lying on the barrier trajectory in the late 1960s [66, 73] and seems to have been a major reason for the decline in popularity of barrier methods. Despite recent complete analyses of the ill-conditioning associated with interior methods, its effects remain widely misunderstood. We therefore discuss them in some detail in this section and the next.

Using Definition 4.1, the classical Newton barrier equations (4.5) are

$$\big(H(\mu) + J^T \Pi(\mu)\, C^{-1}\, J\big)p = -g + J^T \pi(\mu), \tag{4.7}$$

where all functions are evaluated at x and the argument μ is included for emphasis, so that

$$\pi(\mu) = \mu \,/\!\cdot\, c \quad \text{and} \quad H(\mu) = \nabla^2 f - \sum_{i=1}^{m} \pi(\mu) \nabla^2 c_i. \tag{4.8}$$

First, we examine the nature of the ill-conditioning by considering the barrier gradient and Hessian near a solution x^* that satisfies the strong second-order sufficient conditions of Theorem 2.24, limiting our attention to strictly feasible points x close to x^* in the sense that

$$c(x) > 0 \quad \text{and} \quad \|x - x^*\| \leq \delta \quad \text{for suitably small } \delta \ll 1. \tag{4.9}$$

For small enough δ, each inactive constraint $c_i(x) = \Theta(1)$ in this region. (See Definition A.9 for an explanation of $\Theta(\cdot)$ and $\Omega(\cdot)$.)

We further assume that x is close enough to the barrier trajectory so that the smallest active constraint value is not too small compared to μ; formally,

$$c_{\min} \stackrel{\triangle}{=} \min_{i \in \mathcal{A}} \; c_i(x) = \Omega(\mu), \quad \text{so that} \quad \max_{i \in \mathcal{A}} \; \pi_i(x) = O\big(1\big), \tag{4.10}$$

where $\mathcal{A}$ denotes $\mathcal{A}(x^*)$. Under the assumptions of Theorem 3.12, we have that $\|x_\mu - x^*\| = \Theta(\mu)$. Since our intent is to move from x toward x_μ, x_μ should be closer to x^* than x, so we assume that $\mu = O\big(\delta\big)$.

Given that x satisfies (4.9) and $\mu = O\big(\delta\big)$, full column rank of $J_{\mathcal{A}}(x^*)^T$, continuity of $g(x)$ and $J(x)$, and the optimality conditions at x^* imply that the objective gradient satisfies

$$g(x) = J_{\mathcal{A}}(x)^T \lambda_{\mathcal{A}}^* + O\big(\delta\big), \quad \text{so that} \quad g(x) \approx J_{\mathcal{A}}(x)^T \lambda_{\mathcal{A}}^*,$$

where $J_{\mathcal{A}}(x)$ denotes the Jacobian of the constraints active at x^*. (No constraints are active at x.) Throughout this discussion, we take for granted that $\|g(x^*)\| = \Theta(1)$.

When (4.10) applies at x, the primal multiplier π_i is $\Theta(\mu)$ for inactive constraints and $O\big(1\big)$ for active constraints. Hence we may write

$$\nabla B = g - J^T \pi \approx J_{\mathcal{A}}^T \lambda_{\mathcal{A}}^* - J_{\mathcal{A}}^T \pi_{\mathcal{A}} = J_{\mathcal{A}}^T \lambda_{\mathcal{A}}^* - \mu J_{\mathcal{A}}^T C_{\mathcal{A}}^{-1} e, \tag{4.11}$$

where unstarred functions are evaluated at x, and we see that the barrier gradient lies almost entirely in the range of $J_{\mathcal{A}}(x)^T$.

Let J_{in} denote the Jacobian of the constraints that are inactive at x^*, with a similar meaning for C_{in}. Based on the same reasoning used in deriving (4.11), $H(x, \pi) = O\big(1\big)$, where π denotes the primal multipliers (Definition 4.1). At points satisfying (4.9) and (4.10),

$$J^T \Pi C^{-1} J = J_{\mathcal{A}}^T \Pi_{\mathcal{A}} C_{\mathcal{A}}^{-1} J_{\mathcal{A}} + J_{\text{in}}^T \Pi_{\text{in}} C_{\text{in}}^{-1} J_{\text{in}},$$

and the matrix $J^T \Pi C^{-1} J$ will be dominated for sufficiently small μ by $J_{\mathcal{A}}^T \Pi_{\mathcal{A}} C_{\mathcal{A}}^{-1} J_{\mathcal{A}}$, which is $O(1/\mu)$. The barrier Hessian accordingly resembles a large matrix whose column space is the range of $J_{\mathcal{A}}^T$:

$$\nabla^2 B \approx J_{\mathcal{A}}^T \Pi_{\mathcal{A}} C_{\mathcal{A}}^{-1} J_{\mathcal{A}} = \mu J_{\mathcal{A}}^T C_{\mathcal{A}}^{-2} J_{\mathcal{A}}. \tag{4.12}$$

As $\mu \to 0$, $c_{\mathcal{A}} \to 0$, so that $\nabla^2 B$ becomes unbounded in norm, with its row and column spaces converging to the row and column spaces of $J_{\mathcal{A}}^T \Pi_{\mathcal{A}} C_{\mathcal{A}}^{-1} J_{\mathcal{A}}$, which is rank-deficient if $m_{\mathcal{A}}$, the number of active constraints, satisfies $0 < m_{\mathcal{A}} < n$. A point to remember is that the ill-conditioning is created not by the *size* of $\nabla^2 B$—a matrix with a huge norm can be well-conditioned—but by its *rank-deficiency*.

Arguing more formally, one can show that, when evaluated at suitable points near x^* for small values of μ, the barrier Hessian has two widely separated sets of

eigenvalues—$m_{\mathcal{A}}$ asymptotically unbounded eigenvalues and $n - m_{\mathcal{A}}$ eigenvalues that are $\Theta(1)$. Further, the invariant subspace corresponding to the large eigenvalues is close to the range of $J_{\mathcal{A}}^T$ and the complementary invariant subspace corresponding to the small eigenvalues is close to the null space of $J_{\mathcal{A}}$ [111]. The following theorem summarizes properties of the eigenvalues of $\nabla^2 B$ at "nice" points; further details about $\nabla^2 B$ in a general neighborhood of x^* are given in [111].

THEOREM 4.2 (eigenvalues of the barrier Hessian). *Consider a strictly feasible point x_δ satisfying*

$$\|x_\delta - x^*\| \le \delta \text{ with } \delta \ll 1, \quad \mu = O(\delta), \quad \text{and} \quad \min c_i = \Theta(\mu) \text{ for } i \in \mathcal{A}(x^*), \tag{4.13}$$

where δ and μ are sufficiently small. Let $m_{\mathcal{A}}$ denote the number of constraints active at x^*, and assume that $0 < m_{\mathcal{A}} < n$. Finally, let $N_{\mathcal{A}}(x)$ denote a matrix whose columns form an orthonormal basis for the null space of $J_{\mathcal{A}}(x)$, where $\mathcal{A}$ is the set of constraints active at x^*. Then

(i) *$\nabla^2 B(x_\delta, \mu)$ is positive definite;*

(ii) *$\nabla^2 B(x_\delta, \mu)$ has $m_{\mathcal{A}}$ eigenvalues that are $\Theta(1/\mu)$, and $n - m_{\mathcal{A}}$ eigenvalues that are $\Theta(1)$; thus $\operatorname{cond}\big(\nabla^2 B(x_\delta, \mu)\big) = \Theta(1/\mu)$;*

(iii) *let $\{\nu_k\}$ denote the $n - m_{\mathcal{A}}$ eigenvalues of $N_{\mathcal{A}}^T(x_\delta) H(x_\delta, \pi) N_{\mathcal{A}}(x_\delta)$, which is the Lagrangian Hessian $H(x_\delta, \pi)$ projected into the null space of $J_{\mathcal{A}}(x_\delta)$. The $n - m_{\mathcal{A}}$ $\Theta(1)$ eigenvalues of $\nabla^2 B(x_\delta, \mu)$, denoted by $\xi_1, \ldots, \xi_{n-m_{\mathcal{A}}}$, satisfy*

$$|\xi_k - \nu_k| = O(\mu), \quad k = 1, \ldots, n - m_{\mathcal{A}}. \tag{4.14}$$

□

Result (iii) is especially interesting because it shows that the bounded ($\Theta(1)$) eigenvalues of the barrier Hessian are close to those of the reduced Hessian of the Lagrangian, assuming that $H(x, \pi)$ is close to $H(x^*, \lambda^*)$. In effect, crucial information about the reduced Hessian of the Lagrangian function is buried in the barrier Hessian.

At "bad" points inordinately near the boundary (where (4.10) does not hold, so that some active constraint values are very small relative to μ), the barrier Hessian can become arbitrarily ill-conditioned [111]. This property provides support for the classical strategy, still valid today, of avoiding iterates that are too close to the boundary.

Ill-conditioning in the barrier Hessian translates into extreme behavior of the barrier function. For the problem

$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \quad \tfrac{3}{20} x_1 x_2 + x_2 \quad \text{subject to} \quad 1 - 10x_1^2 - x_2^2 \ge 0, \tag{4.15}$$

Figure 5 displays the contours of the barrier function very close to the solution, and very close to the boundary of the active constraint, for $\mu = 10^{-2}$. The smaller the value of μ, the more the level curves follow the curvature of the nonlinear constraint; note that the significant nonlinearity of the constraint is disguised by the elongated scales needed to show the crowded contour lines.

As reflected by the ill-conditioning of the Hessian, it is fair to say that barrier functions are generically badly behaved for small values of μ. Speaking informally, the associated widely divergent scaling in different subspaces underlies recent analyses showing that the sphere of convergence for Newton's method applied to the log barrier function is $\Theta(\mu)$ ([105]; also see [120]). This monotonic decrease in the size of the neighborhood where the local quadratic model is guaranteed to be accurate confirms our earlier remarks about the importance of keeping close to the trajectory.

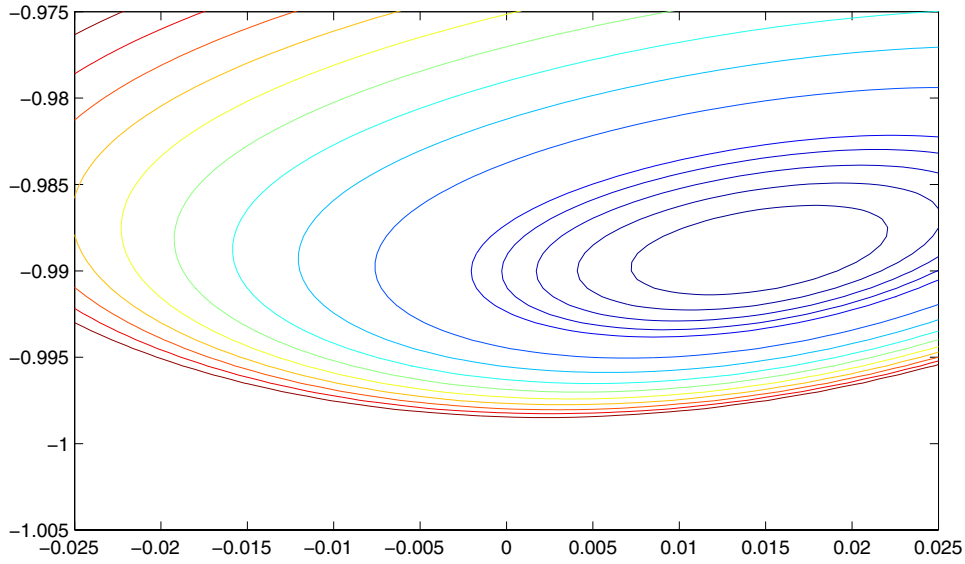

Fig. 5 *This figure shows the contours of the log barrier function for the problem* (4.15) *for* $\mu = 10^{-2}$. *Even for this relatively large value of* μ, *still a long way from* $\mu = 0$, *the contours of the barrier function hug the curvature of the nonlinear constraint.*

4.3.2. Numerical Solution of the Classical Newton Barrier Equations. Given the ill-conditioning just described, it would be entirely reasonable to expect serious numerical trouble when solving for the Newton step. Since the revival of barrier methods, however, it has been repeatedly observed that the computed solutions in interior methods are almost always much more accurate than they should be (see, for example, [90, 110]). To explain why, numerous papers, e.g., [41, 85, 113, 114, 116, 118, 119], have examined ill-conditioning in several contexts, including linear programming, linear complementarity, and primal-dual systems for nonconvex nonlinear optimization. One of the surprising results that has emerged is that, despite extreme ill-conditioning, it is usually possible to solve the Newton barrier equations with acceptable accuracy.

To understand the full effect of an ill-conditioned matrix, we need to consider its *computed* version. In forming the barrier Hessian in finite precision, errors in the computed active constraints play a central role, since their values are by definition converging to zero. For most practical problems, the computed constraint values will almost certainly be subject to *cancellation*, except for special cases like simple bound constraints $x \geq 0$, where the constraint and variable have the same value.

The properties of the computed barrier Hessian are exactly analogous to those analyzed in [113, 119] for the condensed primal-dual Hessian (see (5.4)). A key result is that, when the active constraints are computed with cancellation error, the computed barrier Hessian and gradient at "nice" points (in the sense of Theorem 4.2) are likely to experience relative perturbations *much larger than the unit roundoff*. But—and this is the saving grace—these lie almost entirely in the range space of $J_{\mathcal{A}}^T$ and hence are not blown up by the ill-conditioning.

For completeness, we quote a result, derived as in [113], about the accuracy of the computed Newton direction.

RESULT 4.1. *Consider applying a log barrier method to the problem of minimizing* $f(x)$ *subject to* $c(x) \geq 0$, *where the strong sufficient conditions of Theorem* 2.24 *are satisfied at* x^*. *Assume that, for sufficiently small* μ,

(i) *the barrier Hessian and gradient are evaluated at a strictly feasible point* x *satisfying* (4.13);
(ii) *the value of* $c_{\min}$ *in* (4.13) *is sufficiently large relative to the unit roundoff* $\mathbf{u}$ *so that the chosen factorization of the barrier Hessian runs successfully to completion;*
(iii) *the computed constraint functions are subject to cancellation error; and*
(iv) *the computed step* $\tilde{p}$ *is obtained by applying a backward-stable direct method to solve the Newton equations, and any growth factor associated with the factorization is bounded by a reasonable constant.*

Let p *denote the exact solution of the Newton equations. As long as* $\|p\| = \Omega(\delta)$, *then*

$$\|\tilde{p} - p\| \;\leq\; O(\mathbf{u}) + \|p\| O(\mathbf{u}/\delta). \qquad \square \tag{4.16}$$

The (happy) result is that the computed search direction typically differs from the exact solution by an $O(1)$ multiple of the unit roundoff. Note that even though $\tilde{p}$ and p are similar in norm, they need not display a small *relative* error when both are small.

An interesting corollary is that solving alternative, well-conditioned systems such as (4.21) that are satisfied by the classical Newton barrier step is unlikely to produce more accurate answers; close to x^*, the computed right-hand side in such a system will almost always contain absolute errors that are of the order of machine precision, and so will the computed solution. Hence, even though the alternative matrix is well-conditioned, the associated computed step is likely to contain absolute errors of order machine precision—i.e., errors not much smaller than the errors (4.16) resulting from solving the ill-conditioned Newton barrier system.

To summarize: because of a fortuitous combination of special structure in the matrix, the right-hand side, and the errors arising in computing them, the much-feared ill-conditioning of barrier methods is not harmful when the Newton system (4.5) is solved at reasonable points in a carefully implemented algorithm.

4.3.3. Flaws in the Exact Newton Direction. Knowing that ill-conditioning is not usually harmful, one might conclude that it makes sense to use a classical Newton barrier method. However, as we shall now see, serious inefficiencies occur in the classical barrier method because of fundamental defects in the *exact* Newton step.

Assume that the strong second-order sufficient conditions of Theorem 2.24 hold at x^*. Using the approximate expressions (4.11) and (4.12) for the barrier gradient and Hessian that apply when $\mu = O(\delta)$ and x satisfies (4.9) and (4.10), the Newton equations at x "look like" the following relation, which involves only vectors that lie in the range of $J_{\mathcal{A}}^T$:

$$\mu J_{\mathcal{A}}^T C_{\mathcal{A}}^{-2} J_{\mathcal{A}} p \approx -J_{\mathcal{A}}^T \lambda_{\mathcal{A}}^* + \mu J_{\mathcal{A}}^T C_{\mathcal{A}}^{-1} e, \tag{4.17}$$

where $\mathcal{A}$ denotes the set of constraints active at x^*. Since by assumption $J_{\mathcal{A}}^T$ has full column rank at x^*, it also has full column rank near x^*, and may be canceled from both sides of (4.17). The Newton equations at x are thus (approximately)

$$\mu C_{\mathcal{A}}^{-2} J_{\mathcal{A}} p \approx -\lambda_{\mathcal{A}}^* + \mu C_{\mathcal{A}}^{-1} e. \tag{4.18}$$

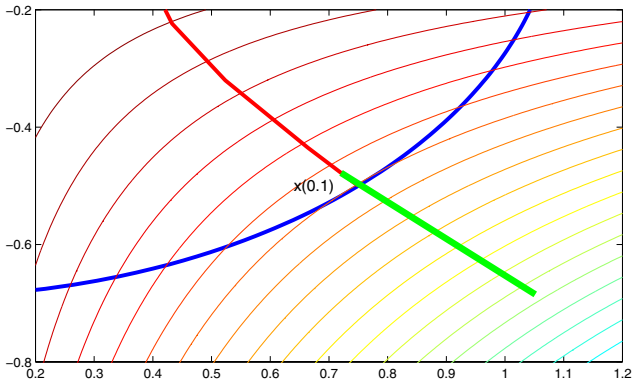

Fig. 6 *For problem* (3.4)*, the exact Newton step is shown for* $\mu = 0.0125$*, starting at the exact barrier minimizer* $x(0.1)$*, which lies on the red barrier trajectory. As predicted, the full Newton step moves to a point that violates, by a long way, the active ellipsoidal constraint.*

For $i \in \mathcal{A}$, the approximate Newton equation from (4.18) is

$$(\nabla c_i)^T p \approx c_i - \frac{c_i^2 \lambda_i^*}{\mu}.$$

Now suppose that the current iterate x is equal to the *exact barrier minimizer* x_μ, which means that $\mu/c_i \approx \lambda_i^*$. If we then reduce the barrier parameter to $\widehat{\mu}$, where $\widehat{\mu} < \mu$, and compute the Newton step with the barrier parameter taken as $\widehat{\mu}$, the Newton step satisfies

$$(\nabla c_i)^T p \approx -c_i\Big(\frac{\mu}{\widehat{\mu}} - 1\Big). \tag{4.19}$$

When μ exceeds $\widehat{\mu}$ by some reasonable factor, e.g., the ratio $\mu/\widehat{\mu}$ is greater than (say) 2, the relationship (4.19) strongly suggests that taking the full Newton step to $x + p$ will produce an *infeasible* point, since the linearized step to the boundary satisfies $(\nabla c_i)^T p \approx -c_i$. For example, if $\widehat{\mu}$ is smaller than μ by a factor of 10, as in a practical barrier algorithm, then $(\nabla c_i)^T p \approx -9c_i$ and the Newton step is likely to produce substantial infeasibility. (See [22, 112] for further details and comments on the exact Newton step.)

This phenomenon can be seen in Figure 6, which shows a closeup of the region near the solution of problem (3.4), along with the (red) trajectory of barrier minimizers. At the strictly interior barrier function minimizer $x(0.1)$, the green line shows the exact Newton step corresponding to $\mu = 0.0125$, so that the ratio between old and new μ is 8. As predicted by our analysis, the full Newton step moves to a significantly infeasible point. Similar results apply when x is very close to a minimizer of the barrier function.

Classical barrier methods are indeed inefficient—but, by a strange twist of fate, ill-conditioning, their longtime hobgoblin, is *not* the culprit.

4.4. The Augmented Newton Barrier Equations. As noted in section 4.3.1, the Hessian matrix in the classical Newton barrier equations (4.7) becomes increasingly ill-conditioned as the solution is approached. Instead of solving the classical

Newton system directly, it is possible to define an *augmented* or *stretched* system that is not inevitably ill-conditioned as $\mu \to 0$ (see, e.g., [57]). Recalling the definition

$$H\big(x, \pi(x,\mu)\big) = \nabla^2 f(x) - \sum_{i=1}^{m} \pi_i(x,\mu)\nabla^2 c_i(x), \quad \text{with} \quad \pi(x,\mu) = \mu \,\text{⁄}\, c(x), \tag{4.20}$$

and using block elimination, it is easy to see that the Newton barrier direction p is part of the solution of the following $(n+m)\times(n+m)$ augmented system:

$$\begin{pmatrix} H\big(x,\pi(x,\mu)\big) & -J(x)^T \\ \Pi(x,\mu)J(x) & C(x) \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = -\begin{pmatrix} g(x) - J(x)^T\pi(x,\mu) \\ 0 \end{pmatrix}, \tag{4.21}$$

where the matrix in (4.21) can be symmetrized using diagonal row and column scaling (see section 5.1).

Now consider what happens as $\mu \to 0$. For a constraint that is inactive at the solution, $\pi_i(\mu) \to 0$ because $\pi_i = \mu/c_i$ and $c_i \not\to 0$. For each such constraint, the associated row of ΠJ converges to zero. Consequently, the matrix of (4.21) converges to a matrix that can be symmetrically permuted into a 2×2 block-diagonal matrix, where one diagonal block contains the values of the inactive constraints and the other diagonal block is

$$M_{\mathcal{A}} \triangleq \begin{pmatrix} H\big(x,\pi(x,\mu)\big) & -J_{\mathcal{A}}(x)^T \\ \Pi_{\mathcal{A}}(x,\mu)J_{\mathcal{A}}(x) & 0 \end{pmatrix}. \tag{4.22}$$

Assume now that the strong sufficient optimality conditions of Theorem 2.24 are satisfied at x^*: full row rank of $J_{\mathcal{A}}(x^*)$, strict complementarity, and positive-definiteness of $N_{\mathcal{A}}^T H(x^*,\lambda^*)N_{\mathcal{A}}$, where the columns of $N_{\mathcal{A}}$ form a basis for the null space of $J_{\mathcal{A}}(x^*)$. Then, $\pi_{\mathcal{A}}\big(x(\mu),\mu\big) \to \lambda_{\mathcal{A}}^* > 0$. Hence, $\Pi_{\mathcal{A}}\big(x(\mu),\mu\big)$ remains nonsingular as $\mu \to 0$. Lemma A.15 then implies that $M_{\mathcal{A}}$ is nonsingular, with a bounded condition number, as $\mu \to 0$.

Expressing the primal Newton step p as part of the solution of (4.21) provides an alternative perspective on the difficulties discussed in section 4.3.3. The nonlinear equations associated with the perturbed optimality conditions (3.49) are

$$\begin{aligned} g\big(x(\mu)\big) - J\big(x(\mu)\big)^T\lambda(\mu) = 0 \quad &\text{and} \\ C\big(x(\mu)\big)\lambda(\mu) - \mu e = 0. & \end{aligned}$$

Differentiating $x(\mu)$ and $\lambda(\mu)$ with respect to μ leads to the following equations for the *tangent step* $\big(x'(\mu),\lambda'(\mu)\big)$:

$$\begin{pmatrix} H\big(x(\mu),\lambda(\mu)\big) & -J(x(\mu))^T \\ \Lambda(\mu)J(x(\mu)) & C(x(\mu)) \end{pmatrix} \begin{pmatrix} x'(\mu) \\ \lambda'(\mu) \end{pmatrix} = \begin{pmatrix} 0 \\ e \end{pmatrix}. \tag{4.23}$$

Applying the same arguments as in the analysis of (4.21), it follows that if the strong sufficient optimality conditions of Theorem 2.24 hold at x^*, the matrix of (4.23) has a bounded condition number as $\mu \to 0$.

Consider the situation where a solution $\big(x(\mu),\lambda(\mu)\big)$ of the perturbed optimality conditions (3.49) is known for a particular value μ. Our purpose is to study the primal Newton step as the barrier parameter is reduced from μ to $\widehat{\mu}$, where $\widehat{\mu} < \mu$

and $\widehat{\mu}$ is close to μ. Since the matrix in (4.23) is expected to be well-conditioned, the tangent step defined by (4.23) should be a good prediction of the step from the barrier minimizer $x(\mu)$ to the barrier minimizer $x(\widehat{\mu})$, i.e.,

$$x(\widehat{\mu}) \approx x(\mu) + (\widehat{\mu} - \mu)x'(\mu). \tag{4.24}$$

Since $\pi(x,\mu)$ is identical to $\lambda(\mu)$ when $x = x(\mu)$, it follows that $H\big(x,\pi(x,\mu)\big) = H\big(x(\mu),\lambda(\mu)\big)$, and the system (4.23) for the tangent step may be written as

$$\begin{pmatrix} H & -J^T \\ \Pi J & C \end{pmatrix} \begin{pmatrix} x'(\mu) \\ \lambda'(\mu) \end{pmatrix} = \begin{pmatrix} 0 \\ e \end{pmatrix}, \tag{4.25}$$

where $H = H\big(x(\mu),\pi\big)$ with $\pi = \pi\big(x(\mu),\mu\big)$, and J and C are defined at $x(\mu)$.

Now suppose that we proceed to minimize $B(x,\widehat{\mu})$ using Newton's method, starting at $x = x(\mu)$. In order to simplify the notation in the augmented barrier equations, we define $\widehat{\pi} = \widehat{\mu} \mathbin{./} c\big(x(\mu)\big)$ and $\widehat{H} = H\big(x(\mu),\widehat{\pi}\big)$. Since $x = x(\mu)$ we have $\widehat{\pi} = (\widehat{\mu}/\mu)\pi$. Consider the situation where the objective is *linear*, in which case $\widehat{H} = (\widehat{\mu}/\mu)H$. If we take into account that $g = \mu J^T C^{-1} e$, it follows from (4.21) that the primal Newton direction p at the point $x(\mu)$ satisfies

$$\begin{pmatrix} \dfrac{\widehat{\mu}}{\mu}H & -J^T \\ \dfrac{\widehat{\mu}}{\mu}\Pi J & C \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = -(\mu - \widehat{\mu}) \begin{pmatrix} J^T C^{-1} e \\ 0 \end{pmatrix},$$

or, equivalently,

$$\begin{pmatrix} H & -J^T \\ \Pi J & C \end{pmatrix} \begin{pmatrix} \widehat{p} \\ \widehat{q} \end{pmatrix} = -(\mu - \widehat{\mu}) \begin{pmatrix} 0 \\ e \end{pmatrix}, \tag{4.26}$$

where $\widehat{p} = (\widehat{\mu}/\mu)p$ and $\widehat{q} = q - (\pi - \widehat{\pi})$. Comparing (4.25) and (4.26) we see that the primal Newton direction satisfies

$$p = \frac{\mu}{\widehat{\mu}}(\widehat{\mu} - \mu)x'(\mu) \tag{4.27}$$

and is a factor of $\mu/\widehat{\mu}$ larger than the "ideal" step suggested by (4.24). Thus we see another explanation, from a different view than the one taken in section 4.3.3, of why the Newton barrier step is poorly scaled immediately after a reduction in the barrier parameter.

4.5. Extrapolation Methods. The poor step that occurs immediately following a reduction in the barrier parameter can also be explained intuitively in terms of the quality of the multiplier estimates provided by a barrier method. When the current point x is equal or very close to a barrier minimizer $x(\mu)$ for small μ, the value $\pi(x,\mu) = \mu \mathbin{./} c(x)$ is likely to be a good estimate of the optimal multipliers, but $\pi(x,\widehat{\mu}) = \widehat{\mu} \mathbin{./} c(x)$, which is smaller by a factor of $\widehat{\mu}/\mu$, will be a poor estimate. To finesse this difficulty, *extrapolation methods* can be devised that retain the multiplier estimate $\pi\big(x(\mu),\mu\big)$ for the *first* iteration following a reduction in the barrier parameter from μ to $\widehat{\mu}$ [7, 30, 58]. At this iteration, instead of the conventional strategy of solving (4.21) with $\mu = \widehat{\mu}$, a special step Δx is defined that satisfies

$$\begin{pmatrix} H\big(x,\pi(x,\mu)\big) & -J(x)^T \\ \Pi(x,\mu)J(x) & C(x) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = -\begin{pmatrix} g(x) - J(x)^T\pi(x,\widehat{\mu}) \\ 0 \end{pmatrix}, \tag{4.28}$$

i.e., each occurrence of $\widehat{\mu}$ in the augmented matrix is replaced by the old barrier parameter μ.

To illustrate the effect of this strategy, consider computing the extrapolated Δx at a minimizer $x = x(\mu)$ of the barrier function. As in the previous section we use a new barrier parameter $\widehat{\mu}$ such that $\widehat{\mu} < \mu$, and define $H = H\big(x(\mu), \pi\big)$ with $\pi = \mu \cdot/ c\big(x(\mu)\big)$ and $\widehat{\pi} = \widehat{\mu} \cdot/ c\big(x(\mu)\big)$. Simple rearrangement of the augmented equations (4.28) and substitution of the relation $\widehat{\pi} = (\widehat{\mu}/\mu)\pi$ gives the equivalent system

$$\begin{pmatrix} H & -J^T \\ \Pi J & C \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\widehat{\lambda} \end{pmatrix} = -(\mu - \widehat{\mu}) \begin{pmatrix} 0 \\ e \end{pmatrix}, \tag{4.29}$$

where $\Delta\widehat{\lambda} = \Delta\lambda - (\pi - \widehat{\pi})$. Comparing this equation with (4.23) shows that Δx satisfies $\Delta x = (\widehat{\mu} - \mu)x'(\mu)$, giving the desired step from (4.24).

By not updating the multiplier estimates until after the first Newton iteration, it is thus possible to correct an inherent deficiency of the classical Newton barrier method. Other forms of extrapolation have been proposed in [33, 78]. A related idea is to convert the general problem into an equivalent problem with a linear objective function. (This can be done by creating a new variable ζ, adding the inequality constraint $\zeta - f(x) \geq 0$, and minimizing with respect to x and ζ.) With this approach, the conventional barrier step is computed immediately after a reduction in the barrier parameter, but the direction is then rescaled by the factor $\widehat{\mu}/\mu$ (see (4.27)). For more details, see [121].

4.6. Modified Barrier Functions. The unconstrained minimizers of the classical log barrier function converge to a solution of the original constrained problem only if the barrier parameter μ goes to zero. By contrast, *modified barrier methods* [13, 23, 54, 76, 84] define a sequence of unconstrained problems in which the value of μ remains *bounded away from zero*, thereby avoiding the need to solve a problem whose Hessian becomes increasingly ill-conditioned as μ is decreased.

Modified barrier methods are based on the observation that for a fixed positive μ, the constraints $c_i(x) \geq 0$ and $\mu \ln\big(1 + c_i(x)/\mu\big) \geq 0$ are equivalent, i.e., their associated sets of feasible points are identical. Moreover, a KKT point for the original problem (2.1) is also a KKT point for the modified problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ f(x) \ \text{ subject to } \ \mu \ln\big(1 + c_i(x)/\mu\big) \geq 0, \quad i = 1, 2, \ldots, m. \tag{4.30}$$

This motivates the definition of the *modified barrier function*,

$$M(x, \lambda) = f(x) - \mu \sum_{i=1}^{m} \lambda_i \ln\big(1 + c_i(x)/\mu\big), \tag{4.31}$$

which can be interpreted as the conventional Lagrangian function for the modified problem (4.30).

A complete theory analogous to that of section 3 exists for the modified barrier function (see [84]). A crucial property of the modified barrier function is that if λ^* is a multiplier vector in $\mathcal{M}_\lambda(x^*)$ (see Definition 2.5, section 2.1), then there exists a fixed μ^* such that for all $\mu < \mu^*$, the corresponding x^* is a local minimizer of $M(x, \lambda^*)$, which implies that $\nabla M(x^*, \lambda^*) = 0$ and $\nabla^2 M(x^*, \lambda^*)$ is positive semidefinite (see Lemma A.7). It follows that if an optimal multiplier is known, x^* can be found from just *one* unconstrained minimization. (The motivation for modified barrier methods is similar to that for augmented Lagrangian methods; see, e.g., [34, 52, 77, 80].)

In practice, neither the optimal multiplier vector nor an upper bound on μ is known in advance. As a result, a sequence of problems must be solved in which each $M(x,\lambda)$ is defined with estimates of λ^* and μ^*. The multiplier estimate is updated following each subproblem, and the barrier parameter is reduced if $\nabla^2 M(x,\lambda)$ is not sufficiently positive definite. (For details, see, e.g., [13, 54, 76, 84].)

If the barrier parameter needs to be reduced to a new value $\widehat{\mu}$, the current x may not satisfy $1+c_i(x)/\widehat{\mu} > 0$, and the modified barrier function will not be well-defined. In this situation, either x must be reset to a known point in strict($\mathcal{F}$) or a special feasibility procedure must be used to find an acceptable new point. This difficulty can be alleviated using the techniques of section 6 for the treatment of inequalities with no known interior point. However, modified barrier methods may be inefficient for so-called *critical inequalities* that must be satisfied at all times.

The problem (4.31) implicitly imposes the *shifted* constraints $c(x)+\mu e \geq 0$. A related approach [23] is to define a set of explicit positive shifts $\{s_i\}$ and find an unconstrained minimizer of

$$M(x,\lambda,s) = f(x) - \sum_{i=1}^{m} \lambda_i s_i \ln\bigl(1 + c_i(x)/s_i\bigr). \tag{4.32}$$

In this case, the value of s_i can be interpreted as either a shift (see section 6.4) or as a barrier parameter for the ith constraint.

We have now covered the main features of interior methods based on minimizing the log barrier function through either the classical Newton equations (4.5) or the stretched system (4.21). By far the most popular interior-point methods today are *primal-dual methods*, which we discuss next.

5. Primal-Dual Interior Methods. Because of inherent flaws in the classical primal barrier method, primal-dual methods based on properties of $x(\mu)$ are increasingly popular for solving general nonlinear programming problems; see, for example, the recent papers [14, 24, 31, 39, 46, 93, 97, 103]. As in primal-dual methods for linear programming, the original (primal) variables x and the dual variables λ (representing the Lagrange multipliers) are treated as independent.

The usual motivation for primal-dual methods is to find (x,λ) satisfying the equations that hold at $x(\mu)$. In the spirit of the perturbed optimality conditions (3.49), we seek to compute a feasible solution $(x(\mu),\lambda(\mu))$ of the $n+m$ nonlinear equations $F^\mu(x,\lambda)=0$, where

$$F^\mu(x,\lambda) = \begin{pmatrix} g(x) - J(x)^T\lambda \\ C(x)\lambda - \mu e \end{pmatrix}. \tag{5.1}$$

Let v denote the $(n+m)$-vector of the combined unknowns (x,λ) at a point that is strictly feasible in both x and λ, i.e., $c(x)>0$ and $\lambda>0$. If $F^\mu(v)$ denotes the function $F^\mu(x,\lambda)$, then a Newton direction $\Delta v = (\Delta x, \Delta\lambda)$ is defined by the Newton equations $F^\mu(v)'\Delta v = -F^\mu(v)$. After collecting terms on the right-hand side, the Newton *primal-dual equations* may be expressed as

$$\begin{pmatrix} H(x,\lambda) & -J(x)^T \\ \Lambda J(x) & C(x) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} g(x) - J(x)^T\lambda \\ C(x)\bigl(\lambda - \pi(x,\mu)\bigr) \end{pmatrix}, \tag{5.2}$$

where $H(x,\lambda)$ is the Hessian of the Lagrangian evaluated at (x,λ) and $\pi = \mu \cdot/ c(x)$.

All primal-dual methods are based on more or less the idea just described, which is sometimes presented in terms of the logarithmic barrier function (hence leading to

properties of $x(\mu)$) or else in terms of perturbed complementarity (3.8) as a desired property in itself.

The success of primal-dual methods is due in part to their effectiveness at following the barrier trajectory. Consider a point $(x, \lambda) = \big(x(\mu), \lambda(\mu)\big)$ on the trajectory, where μ is now a specific value, and suppose that the barrier parameter is reduced from μ to $\widehat{\mu}$. The primal-dual direction computed at this point with the *new* barrier parameter $\widehat{\mu}$ (i.e., the solution of (5.2) with $\mu = \widehat{\mu}$) is *tangent* to the trajectory at (x, λ). This property is easily shown by noting that $g(x) - J(x)^T\lambda = 0$, $\lambda = \pi(x, \mu)$, $C(x)\lambda = \mu e$, and $\pi(x, \widehat{\mu}) = (\widehat{\mu}/\mu)\pi(x, \mu)$. Hence Δx and $\Delta\lambda$ satisfy

$$\begin{pmatrix} H(x,\lambda) & -J(x)^T \\ \Lambda J(x) & C(x) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} 0 \\ (\mu - \widehat{\mu})e \end{pmatrix}. \tag{5.3}$$

Comparing this equation with (4.23) and (4.24) shows that $\Delta x = (\widehat{\mu} - \mu)x'(\mu)$ and $\Delta\lambda = (\widehat{\mu} - \mu)\lambda'(\mu)$. Hence, Δx and $\Delta\lambda$ will usually give a good approximation of the step to $\big(x(\widehat{\mu}), \lambda(\widehat{\mu})\big)$, the next point on the trajectory. This property does not hold for the classical barrier method (see the discussion of sections 4.4–4.5).

Naturally, the primal-dual equations (5.2) do not begin to constitute a complete algorithm for nonlinear programming. Primal-dual methods are the object of active research today and involve a wide range of algorithmic issues, including

- formulation of the linear system that defines the Newton steps (section 5.1);
- choice of the scheme used to measure and ensure progress toward the solution (section 5.2);
- treatment of nonconvexity; and
- treatment of equality constraints (section 6).

5.1. Formulation of the Primal-Dual Equations. As in the classical Newton barrier method, primal-dual methods have a two-level structure of inner and outer iterations, with the inner iterations corresponding to the iterations of Newton's method for a given value of μ. Primal-dual methods exhibit excellent performance in the neighborhood of a trajectory of minimizers of the barrier function, which we shall subsequently call "a trajectory of minimizers." In particular, under the assumption of strict complementarity and a suitable constraint qualification, the inner iterations converge at a Q-quadratic rate; see, e.g., [31]. (See Definition A.10 of Q-quadratic and Q-superlinear convergence.) Moreover, the inner iterations can be terminated so that the combined sequence of inner iterates ultimately converges to x^* Q-superlinearly if μ is reduced at an appropriate rate; see, e.g., [16, 59, 115, 124, 125].

Beyond the work associated with evaluating the objective function, constraint functions, and their first and second derivatives, the cost of a primal-dual iteration is dominated by the cost of solving the linear system (5.2), and effective sparse linear system software is the key to efficiency for large problems. A common approach is to use block elimination to obtain smaller "condensed" systems. Since $c(x) > 0$, the $(2, 2)$ block of (5.2) may be eliminated to give the following $n \times n$ system for Δx:

$$H_C(x, \lambda)\Delta x = -\big(g(x) - J(x)^T\pi(x, \mu)\big), \tag{5.4}$$

where the *condensed primal-dual matrix* $H_C(x, \lambda)$ is defined as

$$H_C(x, \lambda) \triangleq H(x, \lambda) + J(x)^T D(x, \lambda)^{-1} J(x), \quad \text{with} \quad D(x, \lambda) = \Lambda^{-1}C(x). \tag{5.5}$$

The matrix D, which is introduced for later convenience, is diagonal and positive definite, with diagonal elements $d_i = c_i/\lambda_i$. The symmetric matrix H_C is equal to the

barrier Hessian at any minimizer of the barrier function and so is positive definite at points on the trajectory for sufficiently small μ. Observe that the right-hand side of (5.4) is the negative gradient of the barrier function, so that (5.4) strongly resembles the classical Newton barrier equations.

Like the barrier Hessian, H_C becomes increasingly ill-conditioned in a highly structured way as the iterates converge. However, this ill-conditioning is usually harmless [113, 119]. Furthermore, since the $n - m_{\mathcal{A}}$ eigenvalues of H_C that are $\Theta(1)$ converge to the eigenvalues of $N(x^*)^T H(x^*, \lambda^*) N(x^*)$ (see [113] for details), H_C contains precisely the information needed by the algorithm to determine whether the computed iterates are converging to a point satisfying second-order optimality conditions; see section 2.1. The condensed primal-dual system can be solved by either direct or iterative methods, using (for example) an off-the-shelf Cholesky factorization or preconditioned conjugate-gradient method.

A drawback with block elimination is that significant fill-in can occur in H_C (except when the constraints are simple bounds, in which case H_C is simply H plus a positive diagonal matrix). An alternative strategy is to factorize the full $(n+m)\times(n+m)$ system in (5.2) (see, e.g., [43, 49, 50]), typically after symmetrizing the system. A symmetric matrix can be created by multiplying the second block of equations in (5.2) by Λ^{-1} and changing the sign of the second block of columns, giving

$$\begin{pmatrix} H & J^T \\ J & -D \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta\lambda \end{pmatrix} = - \begin{pmatrix} g - J^T\lambda \\ D(\lambda - \pi) \end{pmatrix}, \tag{5.6}$$

where dependencies on x, λ, and μ have been suppressed for brevity. As $\mu \to 0$, the diagonals of D corresponding to the active constraints grow without bound, and so this particular form of symmetrization produces an increasingly ill-conditioned system. However, it can be shown that the ill-conditioning is benign as long as certain direct methods are used to factorize the matrix. (For more details, see [41, 85, 114, 119].)

The structure of (5.6) is useful in later analysis, but in practical computation it is better to symmetrize without increasing the condition of the system—for example, by premultiplying the second block of equations in (5.2) by $\Lambda^{-1/2}$ and simultaneously scaling the λ-variables by $\Lambda^{-1/2}$ [91]. This symmetrization gives

$$\begin{pmatrix} H & J^T\Lambda^{1/2} \\ \Lambda^{1/2}J & -C \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Lambda^{-1/2}\Delta\lambda \end{pmatrix} = - \begin{pmatrix} g - J^T\lambda \\ \Lambda^{-1/2}C(\lambda - \pi) \end{pmatrix}. \tag{5.7}$$

If strict complementarity holds at the solution, and the gradients of the active constraints are linearly independent, the matrix of (5.7) remains well-conditioned as the solution is approached. A benign ill-conditioned (diagonal) scaling remains when forming the right-hand side.

Let K denote the symmetric matrix in (5.6). Since $\lambda > 0$, the inertias of the matrices in (5.6) and (5.7) are identical, by Sylvester's law of inertia (Lemma A.12). Moreover, the inertia of the condensed matrix H_C can be deduced from either of the matrices of (5.6) or (5.7), since $\mathrm{In}(K) = \mathrm{In}(H + J^T D^{-1} J) + (0, m, 0)$ by positive-definiteness of D (Lemma A.16).

If (2.1) is a convex program (Definition A.8), the matrix K of (5.6) is *symmetric quasi-definite*, which means that for every row and column permutation P there exists a factorization $P^T K P = LBL^T$, where L is unit lower-triangular and B is *diagonal* (see [53, 102]). This allows P to be selected solely for the purposes of obtaining a sparse factor L. For nonconvex problems, the relevant factorization is $P^T K P = LBL^T$,

where B is now a block-diagonal matrix with diagonal blocks of order 1×1 or 2×2. Efficient sparse-matrix methods are also available for this case (see, e.g., [5, 29]), but additional overhead is likely since the choice of P depends on numerical considerations unrelated to the sparsity of the factors.

In the neighborhood of a trajectory of minimizers we can expect nonsingularity of the primal-dual system and Q-quadratic convergence of the inner iterates. However, when the problem is nonconvex and the current primal-dual iterate is far from the trajectory, there is no guarantee that a solution of the primal-dual system or condensed system exists. Moreover, when a merit function is used to guarantee convergence from arbitrary starting points, the solution of the primal-dual system—even if it exists—may not be a direction of improvement for the merit function (see section 5.2). In this case, the solution of (5.2) or (5.4) is not useful, and systems based on certain modified Hessians $\bar{H}$ must be formulated.

5.2. Forcing Global Convergence. An important component of a practical primal-dual method for nonconvex optimization is the technique used to ensure convergence from a general starting point. One of the most popular choices is to require, through a line search, a sufficient decrease in a *merit function* that encourages the early iterates to move toward the trajectory. Letting Δv denote the full Newton step $(\Delta x, \Delta\lambda)$, a typical strategy for a convex program (Definition A.8) is to choose a positive α such that $c(x + \alpha\Delta x) > 0$ and some norm of the residual associated with the nonlinear equations $F^\mu(x, \lambda) = 0$ (5.1) is sufficiently reduced, i.e., $\|F^\mu(v + \alpha\Delta v)\| < \|F^\mu(v)\|$ (see, e.g., [31, 115]).

The definition of a merit function for nonconvex problems is considerably more complicated. Since the majority of primal-dual methods require the provision of second derivatives, algorithms have all the information necessary to find second-order points—i.e., points that satisfy the second-order necessary conditions for optimality (Lemma 2.17). But the merit function $\|F^\mu(v)\|$ is not appropriate in this case since it ensures convergence only to points satisfying the first-order conditions.

A number of primal-dual methods use the classical barrier function $B(x, \mu)$ as a merit function (see, e.g., [21, 62]). Since $B(x, \mu)$ does not involve the multipliers, these methods use a separate measure to safeguard the dual variables after the primal step has been taken. For example, in [21], given an interior primal-dual estimate (x, λ), a search direction p is generated from the trust-region subproblem:

$$\min_{p \in \mathbb{R}^n} \quad p^T \nabla B + \tfrac{1}{2} p^T (H + J^T D^{-1} J) p \ \text{ subject to } \ \|p\|_W \leq \delta,$$

where $\|p\|_W = \left(p^T W p\right)^{1/2}$, with W a positive-definite symmetric approximation to the condensed Hessian $H + J^T D^{-1} J$. After completion of the primal step, the dual variables are updated by a separate procedure that ensures convergence of $\|\lambda - \pi(x)\|$ to zero.

A different approach is based on the merit function

$$M^\mu(x, \lambda) = f(x) - \mu \sum_{i=1}^{m} \ln c_i(x) - \mu \sum_{i=1}^{m} \left(\ln \left(\frac{c_i(x)\lambda_i}{\mu} \right) + 1 - \frac{c_i(x)\lambda_i}{\mu} \right), \tag{5.8}$$

which includes both primal and dual variables (see [39]). The function $M^\mu(x, \lambda)$ is the classical barrier function $B(x, \mu)$ augmented by a weighted proximity term that measures the distance of (x, λ) to the trajectory $\big(x(\mu), \lambda(\mu)\big)$. (Similar proximity measures have been used in convex programming; see, e.g., [104].) The defining property of $M^\mu(x, \lambda)$ is that it is minimized with respect to both x and λ at any

point $(x(\mu), \lambda(\mu))$ on the trajectory of minimizers, which implies that a decrease in $M^\mu(x,\lambda)$ can be used to encourage progress toward a minimizer of $B(x,\mu)$. The gradient and Hessian of $M^\mu(x,\lambda)$ are

$$\nabla M^\mu(x,\lambda) = \begin{pmatrix} g - J^T(2\pi - \lambda) \\ \Lambda^{-1}C(\lambda - \pi) \end{pmatrix}, \quad \nabla^2 M^\mu(x,\lambda) = \begin{pmatrix} H_M & J^T \\ J & \mu\Lambda^{-2} \end{pmatrix},$$

where π is the vector of primal multipliers and $H_M = H(x, 2\pi - \lambda) + 2J^T C^{-1}\Pi J$. Since $\big(x(\mu), \lambda(\mu)\big)$ minimizes $M^\mu(x,\lambda)$, it follows that $\nabla M^\mu(x,\lambda) = 0$ and $\nabla^2 M^\mu(x,\lambda)$ is positive semidefinite at every point on the trajectory. As a result, line-search or trust-region methods can be devised in which the local quadratic model is $\mathcal{Q}(s) = s^T \nabla M + \frac{1}{2} s^T G s$, where

$$G = \begin{pmatrix} H + 2J^T D^{-1} J & J^T \\ J & D \end{pmatrix}, \tag{5.9}$$

i.e., G is $\nabla^2 M^\mu(x,\lambda)$ with π replaced by λ and $\mu\Lambda^{-1}$ replaced by C. It can be shown that if $H + J^T D^{-1} J$ is positive definite, the solution of $Gs = -\nabla M$ is the unique minimizer of $\mathcal{Q}(s)$ and that $s = (\Delta x, \Delta\lambda)$ also solves the primal-dual system (5.2). To see this, note that a premultiplication of both sides of (5.2) by the nonsingular matrix

$$\begin{pmatrix} I & 2J^T D^{-1} \\ 0 & \Lambda^{-1} \end{pmatrix}$$

gives the equivalent equation

$$\begin{pmatrix} H + 2J^T D^{-1} J & J^T \\ J & D \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} g - J^T(2\pi - \lambda) \\ D(\lambda - \pi) \end{pmatrix}. \tag{5.10}$$

These properties suggest a line-search method that uses the solution of the primal-dual system (5.2) as a search direction (for more details, see [39]). If G of (5.9) is sufficiently positive definite, the search direction is the unique solution of $Gs = -\nabla M$ (which is equivalent to the primal-dual system (5.2)). Otherwise, the search direction is a linear combination of two vectors that are by-products of the relevant factorization of (5.6); the first vector is the solution of a related positive-definite system $\bar{G}s = -\nabla M$, and the second vector is a direction of negative curvature for the quadratic model $\mathcal{Q}(s)$ (see Definition A.13). If the condensed matrix is formed, it can be modified "on the fly" during the factorization so that its factors are those of a positive-definite $\bar{H} + J^T D^{-1} J$ for some implicitly defined $\bar{H}$ (see, e.g., [40]). Alternatively, the inertia-controlling LBL^T factorization discussed in [38, 39, 42] detects and modifies indefiniteness of the (implicitly defined) matrix $H + J^T D^{-1} J$ while factorizing the full system (5.6). A potential drawback is that the row and column interchanges needed by the inertia-controlling factorization may interfere with the row and column ordering used to maintain sparsity, producing factors that are generally less sparse than those obtained by off-the-shelf sparse-matrix software.

A related strategy is to apply a trust-region method based on finding an approximate solution of the subproblem

$$\underset{s \in \mathbb{R}^{n+m}}{\text{minimize}} \quad \mathcal{Q}(s) \quad \text{subject to} \quad \|s\|_T \leq \delta,$$

where $\|s\|_T = (s^T T s)^{1/2}$, δ is the trust-region radius, and T is the block-diagonal matrix $T = \text{diag}(I, D)$, where $D = \Lambda^{-1}C$ as before. In [47], it is shown that a trust-region step can be computed by factorizing a sequence of systems with diagonally modified primal-dual structure, where the inertia of these systems can be determined without recourse to a special factorization method.

A quite different procedure for forcing convergence is to use a *filter*, first proposed in [36] in the context of SQP methods. Constrained optimization is a "two-objective" problem in which we would like to minimize (i) the objective function and (ii) the norm of the constraint violations. Since these goals usually conflict, most merit functions define a compromise based on a weighted combination of (i) and (ii). In the SQP context, a filter is a collection of ordered pairs, each consisting of the objective value and violated constraint norm at a particular point, with the property that no entry in the filter is *dominated* by any other entry. (A point x is said to dominate a point $\hat{x}$ if the objective and constraint violation norm at x are both less than or equal to the corresponding values at $\hat{x}$.) A new iterate is accepted only if its objective function value or constraint violation yields a strict improvement compared to some point in the filter. If the new iterate qualifies, it is added to the filter, and all filter entries that it dominates are removed. In practice many details need to be specified in order to define a convergent filter algorithm; we refer the interested reader to [35, 36]. Recently, several algorithms were proposed that use a filter to force convergence of a primal-dual interior method (see [8, 100, 107]). In the primal-dual context, there are several ways to define the filter pair, including the norm of the gradient of the Lagrangian and the proximity norm $\|C(x)\lambda - \mu e\|$.

6. Treatment of Equality Constraints. We now turn to the general problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ f(x) \quad \text{subject to} \quad c_i(x) = 0,\ i \in \mathcal{E} \quad \text{and} \quad c_i(x) \geq 0,\ i \in \mathcal{I}, \tag{6.1}$$

which contains both inequality and equality constraints. Although interior-point methods are, strictly speaking, relevant only to inequality constraints, equality constraints are an essential part of the picture for several reasons.

1. Equality constraints may arise naturally as part of an optimization problem that also contains inequalities—e.g., one wants to find an optimal point with nonnegative coordinates *on*, not inside, a sphere, or the variables are nonnegative probabilities that must sum to one.

2. Although our discussion so far has consistently assumed that an initial strictly feasible point is known, this may not always be true. Inequality constraints arising in practice can (informally) be divided into two classes: *critical* inequalities that must be satisfied at all times and *noncritical* inequalities that can be violated at intermediate iterates. Critical inequalities arise when the constraint functions not only characterize the desired properties of the solution but also define a region in which the problem statement is meaningful. Fortunately, it is typical in these circumstances that a strictly feasible point can easily be determined—for example, in structural design, the initial specification of a bridge may be overengineered to guarantee that an inordinately heavy load can be carried. However, in other situations a strictly feasible point may not be available, even if such a point exists.

3. Finally, it is sometimes more efficient to convert an inequality constraint into an alternative form involving one or more equality constraints. The two most common reformulations of the single inequality constraint $c_i(x) \geq 0$ are

$$c_i(x) - s_i = 0, \quad s_i \geq 0, \quad \text{where } s_i \text{ is called a } \textit{slack variable}; \text{ or}$$
$$c_i(x) + s_i \geq 0, \quad s_i = 0, \quad \text{where } s_i \text{ is called a } \textit{shift variable}.$$

Both of these forms are "equivalent" to the original inequality in the sense that the same set of points in x-space satisfies them. We return to them in section 6.4.

6.1. Optimality Conditions for Mixed Constraints. Before turning to algorithms, we state selected definitions and results that are relevant to optimality conditions for the mixed-constraint problem.

6.1.1. KKT Points and Constraint Qualifications. Let $c_{\mathcal{I}}(x)$ denote the subvector $c_i(x)$, $i \in \mathcal{I}$, and let $c_{\mathcal{E}}(x)$ denote the subvector $c_i(x)$, $i \in \mathcal{E}$. The following is directly analogous to the definition of a KKT point for an inequality-constrained problem (Definition 2.3).

DEFINITION 6.1 (first-order KKT point). *The* first-order KKT conditions *for problem* (6.1) *hold at the point* x^**, or, equivalently,* x^* *is a (first-order)* KKT point*, if there exists a Lagrange multiplier vector* λ^* *such that*

$$c_{\mathcal{I}}(x^*) \geq 0 \ \textit{and} \ c_{\mathcal{E}}(x^*) = 0 \quad \text{(feasibility)}, \tag{6.2a}$$

$$g(x^*) = J(x^*)^T\lambda^* \quad \text{(stationarity)}, \tag{6.2b}$$

$$\lambda^*_{\mathcal{I}} \geq 0 \quad \text{(nonnegativity of the inequality multipliers)}, \ \textit{and} \tag{6.2c}$$

$$c_{\mathcal{I}}(x^*) \cdot \lambda^*_{\mathcal{I}} = 0 \quad \text{(complementarity for the inequality constraints)}. \tag{6.2d}$$

In contrast to Definition 2.3 for the inequality-constrained case, the Lagrange multipliers $\lambda^*_{\mathcal{E}}$ associated with the equality constraints are not restricted in sign. Complementarity holds automatically for equality constraints, which vanish at the solution.

The first-order KKT conditions may be written more compactly as $F(x, \lambda) = 0$, $c_{\mathcal{I}}(x^*) \geq 0$, $\lambda_{\mathcal{I}} \geq 0$, with

$$F(x, \lambda) = \begin{pmatrix} g(x) - J(x)^T\lambda \\ c_{\mathcal{I}}(x) \cdot \lambda_{\mathcal{I}} \\ c_{\mathcal{E}}(x) \end{pmatrix}. \tag{6.3}$$

The KKT conditions are based on the properties of constraint linearizations, and hence they are necessary conditions for optimality only when the local constraint linearizations are sufficiently good, i.e., when a *constraint qualification* holds. (See the analogous discussion in section 2.2.)

DEFINITION 6.2 (active, inactive, and violated constraints). *For the inequality constraints* $c_{\mathcal{I}}(x) \geq 0$*, the* i*th constraint is said to be* active *at* $\bar{x}$ *if* $c_i(\bar{x}) = 0$*,* inactive *if* $c_i(\bar{x}) > 0$*, and* violated *if* $c_i(\bar{x}) < 0$*. For the equality constraints* $c_{\mathcal{E}}(x) = 0$*, the* i*th constraint is* active *at* $\bar{x}$ *if* $c_i(\bar{x}) = 0$ *and* violated *at* $\bar{x}$ *if* $c_i(\bar{x}) \neq 0$*. The* active set $\mathcal{A}(\bar{x})$ *is the set of indices of the constraints active at* $\bar{x}$*, i.e.,* $\mathcal{A}(\bar{x}) = \{i : c_i(\bar{x}) = 0\}$*. The set of active inequality constraints at* $\bar{x}$ *is denoted by* $\mathcal{A}_{\mathcal{I}}(\bar{x})$*, i.e.,* $\mathcal{A}_{\mathcal{I}}(\bar{x}) = \{i \in \mathcal{I} : c_i(\bar{x}) = 0\}$*. The arguments of* $\mathcal{A}$ *and* $\mathcal{A}_{\mathcal{I}}$ *are omitted when they are obvious.*

Note that all equality constraints are active at any feasible point, i.e., if $\bar{x}$ is feasible for (6.1), then $\mathcal{A}(\bar{x}) = \mathcal{A}_{\mathcal{I}}(\bar{x}) \cup \mathcal{E}$. We now define two standard constraint qualifications for mixed constraints.

DEFINITION 6.3 (LICQ for mixed constraints). *Consider a constrained problem with constraints* $c_i(x) \geq 0$*,* $i \in \mathcal{I}$*, and* $c_i(x) = 0$*,* $i \in \mathcal{E}$*. The* linear independence constraint qualification *holds at the feasible point* $\bar{x}$ *if* $J_{\mathcal{A}}(\bar{x})$*, the Jacobian of the active constraints at* $\bar{x}$*, has full row rank, i.e., if the combined gradients of the active inequality constraints and the equality constraints are linearly independent.*

DEFINITION 6.4 (MFCQ for mixed constraints). *Consider a problem with constraints* $c_{\mathcal{I}}(x) \geq 0$ *and* $c_{\mathcal{E}}(x) = 0$. *The* Mangasarian–Fromovitz constraint qualification *holds at the feasible point* $\bar{x}$ *if the gradients of the equality constraints at* $\bar{x}$, $\nabla c_i(\bar{x})$, $i \in \mathcal{E}$, *are linearly independent and if there exists a nonzero vector* p *such that* $\nabla c_i(\bar{x})^T p > 0$ *for all* $i \in \mathcal{A}_{\mathcal{I}}(\bar{x})$ *and* $\nabla c_i(\bar{x})^T p = 0$ *for all* $i \in \mathcal{E}$.

It should be emphasized that full row rank of $J_{\mathcal{E}}(\bar{x})$ is needed for the MFCQ to hold at $\bar{x}$. Further, satisfaction of the LICQ implies that the MFCQ holds as well.

6.1.2. Necessary Optimality Conditions for Mixed-Constraint Problems. As in the inequality-constrained case, we can state necessary conditions for optimality only if a constraint qualification holds.

LEMMA 6.5 (first-order necessary conditions for a local constrained minimizer). *If* x^* *is a local minimizer of problem* (6.1) *and the MFCQ holds at* x^*, *then* x^* *must be a KKT point.* ∎

By analogy with Definition 2.5 for the inequality case, we define the set of multipliers that satisfy the KKT conditions of Definition 6.1.

DEFINITION 6.6 (acceptable Lagrange multipliers). *Given a KKT point* x^* *for problem* (6.1), *the set of acceptable multipliers is defined as*

$$\mathcal{M}_\lambda(x^*) \triangleq \{\lambda \in \mathbb{R}^m : g(x^*) = J(x^*)^T\lambda, \ \ \lambda_{\mathcal{I}} \geq 0, \ \ \textit{and} \ \ \lambda_{\mathcal{I}} \cdot c_{\mathcal{I}}(x^*) = 0\}. \tag{6.4}$$

Using $\mathcal{M}_\lambda(x^*)$, second-order necessary conditions for optimality can be stated when the LICQ holds.

LEMMA 6.7 (second-order necessary conditions). *Consider problem* (6.1), *and suppose that* x^* *is a local minimizer where the LICQ holds. Then there is a vector* λ^* *which satisfies* $\lambda^*_{\mathcal{I}} \geq 0$, $c^*_{\mathcal{I}} \cdot \lambda^*_{\mathcal{I}} = 0$, *and* $g^* = J^{*T}\lambda^*$, *and* $p^T H(x^*, \lambda^*)p \geq 0$ *for all* p *satisfying* $J^*_{\mathcal{A}} p = 0$. ∎

6.1.3. Sufficient Optimality Conditions for Mixed-Constraint Problems. Second-order sufficient optimality conditions for problem (6.1) are similar to the sufficient conditions for the all-inequality problem; see section 2.4.

DEFINITION 6.8 (second second-order sufficient condition (SSC2)). *Consider problem* (6.1). *Let* x^* *denote a KKT point, so that* $c^*_{\mathcal{I}} \geq 0$ *and* $c^*_{\mathcal{E}} = 0$. *We say that SSC2 holds at* x^* *if for* every *Lagrange multiplier* λ *satisfying* $\lambda_{\mathcal{I}} \geq 0$, $c^*_{\mathcal{I}} \cdot \lambda_{\mathcal{I}} = 0$, *and* $g^* = J^{*T}\lambda$, *there exists* $\omega > 0$ *such that* $p^T H(x^*, \lambda)p \geq \omega\|p\|^2$ *for all nonzero* p *such that* $g^{*T}p = 0$, $J^*_{\mathcal{E}}p = 0$, *and* $J^*_{\mathcal{A}_{\mathcal{I}}}p \geq 0$.

Proving that x^* is an *isolated* local constrained minimizer requires a constraint qualification. The next theorem is analogous to Theorem 2.23; see [88, Theorem 2.4] for the proof.

THEOREM 6.9 (sufficient conditions for an isolated constrained minimizer). *The point* x^* *is an isolated local constrained minimizer of problem* (6.1) *if*

(i) x^* *is a KKT point, i.e.,* $c^*_{\mathcal{I}} \geq 0$, $c^*_{\mathcal{E}} = 0$, *and there exists a nonempty set* $\mathcal{M}_\lambda$ *of multipliers* λ *satisfying* $\lambda_{\mathcal{I}} \geq 0$, $c^*_{\mathcal{I}} \cdot \lambda_{\mathcal{I}} = 0$, *and* $g^* = J^{*T}\lambda$;

(ii) *the MFCQ holds at* x^*;

(iii) *for all* $\lambda \in \mathcal{M}_\lambda(x^*)$ *of* (6.4) *and all nonzero* p *satisfying* $g^{*T}p = 0$, $J^*_{\mathcal{E}}p = 0$, *and* $J^*_{\mathcal{A}_{\mathcal{I}}}p \geq 0$, *there exists* $\omega > 0$ *such that* $p^T H(x^*, \lambda)p \geq \omega\|p\|^2$. ∎

Finally, we state the following theorem separately for future reference.

THEOREM 6.10 (strong sufficient conditions for an isolated constrained minimizer). *The point* x^* *is an isolated local constrained minimizer of problem* (6.1) *if*

(i) x^* *is feasible and the LICQ holds at* x^*, *i.e.,* $J_{\mathcal{A}}(x^*)$ *has full row rank;*

(ii) x^* *is a KKT point and strict complementarity holds, i.e., the (necessarily unique) multiplier* λ^* *has the property that* $\lambda^*_i > 0$ *for all* $i \in \mathcal{A}_{\mathcal{I}}(x^*)$;

(iii) *for all nonzero vectors* p *satisfying* $J_{\mathcal{A}}(x^*)p = 0$, *there exists* $\omega > 0$ *such that* $p^T H(x^*, \lambda^*)p \geq \omega \|p\|^2$. □

Condition (iii) of Theorem 6.10 is equivalent to stating that the reduced Hessian of the Lagrangian, $N_{\mathcal{A}}(x^*)^T H(x^*, \lambda^*) N_{\mathcal{A}}(x^*)$, is positive definite, where $N_{\mathcal{A}}(x^*)$ is a matrix whose columns form a basis for the null space of $J_{\mathcal{A}}(x^*)$.

6.2. A Barrier-SQP Approach.

6.2.1. A Very Short Word on SQP Methods. To focus our attention for the moment on the treatment of equality constraints, we consider the problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\ \phi(x) \quad \text{subject to} \quad c_{\mathcal{E}}(x) = 0, \tag{6.5}$$

where the objective is called ϕ so that it is not confused with f. SQP is a powerful technique for solving mixed-constraint problems; here we consider how SQP might be applied to the all-equality problem (6.5). We emphasize that we address only a minute fraction of what can be said about SQP (see, e.g., [6, 10, 34, 52, 77, 80]) and that we do *not* exploit the full generality of SQP methods.

SQP methods are based directly on optimality conditions, and accordingly they maintain estimates of both x and λ. The step p in the x-variables calculated by a Newton SQP method (the only kind considered here) can be viewed in two ways. First, p is the solution of the following quadratic program defined at the point x:

$$\underset{p \in \mathbb{R}^n}{\text{minimize}}\ \tfrac{1}{2} p^T H_\phi(x, \lambda) p + \nabla \phi(x)^T p \quad \text{subject to} \quad J_{\mathcal{E}}(x)p = -c_{\mathcal{E}}(x), \tag{6.6}$$

where λ is the current Lagrange multiplier estimate, $H_\phi(x, \lambda)$ is the Hessian of the Lagrangian for ϕ and $c_{\mathcal{E}}$,

$$H_\phi(x, \lambda) \triangleq \nabla^2 \phi(x) - \sum_{i \in \mathcal{E}} \lambda_i \nabla^2 c_i(x), \tag{6.7}$$

and $J_{\mathcal{E}}$ is the Jacobian of $c_{\mathcal{E}}$. (SQP methods take their name from this interpretation.) Second, and equivalently, p is part of the Newton step associated with solving the nonlinear equations that specify the KKT conditions for the equality-constrained problem (6.5):

$$\nabla \phi(x) = J_{\mathcal{E}}(x)^T \lambda_{\mathcal{E}} \quad \text{and} \quad c_{\mathcal{E}}(x) = 0, \tag{6.8}$$

where $\lambda_{\mathcal{E}}$ is a Lagrange multiplier vector associated with the constraints $c_{\mathcal{E}}(x) = 0$. (See Definition 6.1.)

6.2.2. Combining Barrier and SQP Methods. The original motivation for a barrier method was to eliminate inequality constraints by blending them into a composite objective and then to solve a sequence of unconstrained subproblems. An analogue for mixed constraints is to treat the inequalities through a barrier transformation but to retain the equalities, leading to a sequence of *equality-constrained* subproblems of the form

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) - \mu \sum_{i \in \mathcal{I}} \ln c_i(x) \\ \text{subject to} \quad & c_i(x) = 0, \quad i \in \mathcal{E}. \end{aligned} \tag{6.9}$$

(The constraints $c_i(x) > 0$, $i \in \mathcal{I}$, are not shown because they are treated implicitly by the barrier term.) To solve the overall problem (6.1), we need to solve the subproblem

(6.9) for a sequence of μ values converging to zero. (This was the approach taken in [51] for standard-form linear programming, which has the mixed constraints $Ax = b$ and $x \geq 0$.)

The next question is, How can we solve (6.9)? Based on the motivation sketched in section 6.2.1, an SQP method can be applied to the equality-constrained problem (6.9) whose objective function ϕ is the log barrier function. Substituting in (6.8), the associated KKT conditions have the form

$$g(x) - J_{\mathcal{I}}(x)^T\big(\mu \,\diagup\!\!\cdot\, c_{\mathcal{I}}(x)\big) = J_{\mathcal{E}}(x)^T\lambda_{\mathcal{E}} \quad \text{and} \quad c_{\mathcal{E}}(x) = 0. \tag{6.10}$$

To define a primal-dual method as in section 5, we now create an *independent variable* $\lambda_{\mathcal{I}}$ of multipliers for the inequality constraints from the additional equation

$$c_{\mathcal{I}}(x) \cdot \lambda_{\mathcal{I}} = \mu e,$$

which holds at KKT points of (6.9). With this definition, the KKT conditions (6.10) can be rewritten as a system of nonlinear equations

$$F^\mu(x,\lambda) = \begin{pmatrix} g(x) - J(x)^T\lambda \\ c_{\mathcal{I}}(x) \cdot \lambda_{\mathcal{I}} - \mu e \\ c_{\mathcal{E}}(x) \end{pmatrix} = 0. \tag{6.11}$$

A method based on applying Newton's method to (6.11) will be called a *barrier-SQP* approach. As before in (3.49), the second equation can be viewed as perturbing the complementarity condition for the inequality constraints in the KKT conditions (6.3).

If the constraints satisfy the LICQ (see section 6.1.1) and strong second-order conditions hold at x^* (see Theorem 6.10), a differentiable trajectory of solutions (x_μ, λ_μ) exists when μ is small enough, (x_μ, λ_μ) converges to (x^*, λ^*) as $\mu \to 0$, and x_μ is a local constrained minimizer of (6.9). Barrier-SQP interior methods attempt to follow this trajectory of minimizers by finding an approximate solution of $F^\mu(x,\lambda) = 0$ for a suitable sequence $\mu \to 0$. For a given μ, Newton's method can be applied to (6.11), imposing a further requirement that x and $\lambda_{\mathcal{I}}$ remain strictly feasible with respect to the implicit inequality constraints $c_{\mathcal{I}}(x) \geq 0$ and $\lambda_{\mathcal{I}} \geq 0$.

Letting v denote the $n+m$ unknowns (x,λ), so that $F^\mu(v)$ denotes $F^\mu(x,\lambda)$, the Newton direction $\Delta v = (\Delta x, \Delta\lambda)$ is defined as the solution of $F^\mu(v)'\Delta v = -F^\mu(v)$. The explicit Newton equations are

$$\begin{pmatrix} H & -J_{\mathcal{I}}^T & -J_{\mathcal{E}}^T \\ \Lambda_{\mathcal{I}} J_{\mathcal{I}} & C_{\mathcal{I}} & 0 \\ J_{\mathcal{E}} & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda_{\mathcal{I}} \\ \Delta\lambda_{\mathcal{E}} \end{pmatrix} = - \begin{pmatrix} g - J^T\lambda \\ \lambda_{\mathcal{I}} \cdot c_{\mathcal{I}} - \mu e \\ c_{\mathcal{E}} \end{pmatrix}, \tag{6.12}$$

where we have suppressed the arguments x and λ.

Let $W(x)$ and $Z(\lambda)$ denote diagonal matrices whose entries are

$$w_{\mathcal{I}}(x) = c_{\mathcal{I}}(x), \;\; w_{\mathcal{E}}(x) = 0, \;\; z_{\mathcal{I}}(\lambda) = \lambda_{\mathcal{I}}, \;\; \text{and} \;\; z_{\mathcal{E}}(\lambda) = 1, \tag{6.13}$$

and define the vector $r^\mu(x,\lambda)$ as

$$r^\mu_{\mathcal{I}}(x,\lambda) \triangleq c_{\mathcal{I}}(x) \cdot \lambda_{\mathcal{I}} - \mu e \quad \text{and} \quad r^\mu_{\mathcal{E}}(x,\lambda) \triangleq c_{\mathcal{E}}(x). \tag{6.14}$$

The equations (6.12) may then be written in a compact form similar to (5.2):

$$\begin{pmatrix} H & -J^T \\ ZJ & W \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} g - J^T\lambda \\ r^\mu \end{pmatrix}. \tag{6.15}$$

Since Z is diagonal and positive definite and W is diagonal and positive semidefinite, (6.15) can be symmetrized as (5.2) was in the inequality-constrained case, and the discussion of section 5.1 applies with $D = Z^{-1}W$. Because W has zeros on its diagonal, the $(2,2)$ block in (6.15) cannot be eliminated to form a condensed system analogous to (5.4), but inertial properties can be deduced using Lemma A.16.

For future reference, we define $\pi_{\mathcal{I}}(x,\mu)$ as

$$\pi_{\mathcal{I}}(x,\mu) \triangleq \mu \cdot/ c_{\mathcal{I}}(x), \tag{6.16}$$

so that $\pi_{\mathcal{I}}$ is an estimate of $\lambda_{\mathcal{I}}$ at the current iterate x for a specific value of μ; see Definition 4.1. (Definition of $\pi_{\mathcal{E}}$ will be deferred until it is needed.) Section 6.2.1 mentions that the SQP direction can be interpreted as the solution of a quadratic programming problem (6.6) that includes the Hessian of the Lagrangian and linearizations of the equality constraints. The barrier-SQP Newton step Δx then solves the subproblem

$$\begin{array}{ll} \underset{\Delta x \in \mathbb{R}^n}{\text{minimize}} & \frac{1}{2}\Delta x^T (H + J_{\mathcal{I}}^T C_{\mathcal{I}}^{-1} \Lambda_{\mathcal{I}} J_{\mathcal{I}}) \Delta x + (g - J_{\mathcal{I}}^T \pi_{\mathcal{I}})^T \Delta x \\ \text{subject to} & J_{\mathcal{E}} \Delta x = -c_{\mathcal{E}}. \end{array} \tag{6.17}$$

The solution of (6.17) produces the same result as if the SQP formulation (6.6) were applied directly to (6.9). Interior-point algorithms based on the quadratic programming subproblem (6.17) typically utilize SQP-based merit functions to enforce convergence (see, e.g., [15, 27, 46, 83, 123]). If indefiniteness is addressed with a suitable line-search or trust-region technique, the limit points of the sequence of iterates satisfy the second-order necessary optimality conditions for (6.9) with a fixed value of μ.

Rather than factorize the matrices in (6.12) and (6.15), one might use an iterative method. Since the $(3,3)$ block of the matrix in (6.12) is zero and W in (6.15) is singular, it is not clear how to obtain a matrix guaranteed to be positive definite. One strategy is to consider the KKT conditions for the quadratic programming subproblem (6.17), which can be written as

$$\begin{pmatrix} H + J_{\mathcal{I}}^T C_{\mathcal{I}}^{-1} \Lambda_{\mathcal{I}} J_{\mathcal{I}} & J_{\mathcal{E}}^T \\ J_{\mathcal{E}} & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ -(\lambda_{\mathcal{E}} + \Delta\lambda_{\mathcal{E}}) \end{pmatrix} = - \begin{pmatrix} g - J_{\mathcal{I}}^T \pi_{\mathcal{I}} \\ c_{\mathcal{E}} \end{pmatrix}, \tag{6.18}$$

where the "new" multiplier $\lambda_{\mathcal{E}} + \Delta\lambda_{\mathcal{E}}$ constitutes the multipliers for the quadratic programming subproblem. Note that (6.18) is identical to the result of eliminating the $(2,2)$ block of the matrix in (6.12).

The matrix of (6.18) is symmetric but necessarily indefinite (Lemma A.15), so that a standard conjugate-gradient method cannot be applied directly. However, conjugate-gradient methods can be extended to this situation by introducing a symmetric indefinite *preconditioning matrix*, which may take the form

$$\begin{pmatrix} M & J_{\mathcal{E}}^T \\ J_{\mathcal{E}} & 0 \end{pmatrix}, \tag{6.19}$$

where M is a positive-definite "approximation" to $H + J_{\mathcal{I}}^T C_{\mathcal{I}}^{-1} \Lambda_{\mathcal{I}} J_{\mathcal{I}}$. In choosing M, there is usually a tradeoff between ease of factorization—for example, taking M as diagonal—and speed of convergence of the conjugate-gradient iterations. (See, e.g., [21, section 5].)

6.3. A Penalty-Barrier Approach. In contrast to the SQP approach just described, the classical treatment of equality constraints is to eliminate them through

unconstrained minimization of a composite function that includes a *penalty* for violating $c_\varepsilon(x) = 0$—most commonly, the quadratic penalty function $\|c_\varepsilon\|_2^2/\mu$. Reverting in part to the constraint-removal spirit of the 1960s, we now consider dealing with equality constraints by unconstrained minimization of a *penalty-barrier* function, exactly as proposed in [33]:

$$\Phi_{\text{PB}}(x,\mu) \triangleq f(x) - \mu \sum_{i\in\mathcal{I}} \ln c_i(x) + \frac{1}{2\mu}\,\|c_\varepsilon\|_2^2. \tag{6.20}$$

(The appearance of $\frac{1}{2}$ in the penalty term avoids irritating factors of two in the derivatives.) The implicit constraints $c_\mathcal{I}(x) > 0$ are handled by the barrier term and hence are not included in (6.20). Let x_μ denote an unconstrained minimizer of $\Phi_{\text{PB}}(x,\mu)$. A detailed analysis, analogous to the results in sections 3.3 and 3.4, is given in [33] of the conditions under which, for sufficiently small μ, the sequence $\{x_\mu\}$ defines a differentiable *penalty-barrier trajectory* converging to x^*.

To find x_μ, we exploit its stationarity (Lemma A.7). Writing out $\nabla\Phi_{\text{PB}}(x)$ and rearranging, we obtain a system of nonlinear equations equivalent to the condition that $\nabla\Phi_{\text{PB}}(x) = 0$:

$$F^\mu(x,\lambda) = \begin{pmatrix} g(x) - J(x)^T\lambda \\ c_\mathcal{I}(x) \cdot \lambda_\mathcal{I} - \mu e \\ c_\varepsilon(x) + \mu\lambda_\varepsilon \end{pmatrix} = 0, \tag{6.21}$$

where $\lambda_\mathcal{I}$ and λ_ε represent multiplier estimates that converge to $\lambda_\mathcal{I}^*$ and λ_ε^* as $\mu \to 0$ and, at x_μ, satisfy the relations

$$c_\mathcal{I}(x_\mu) \cdot \lambda_\mathcal{I} = \mu e \quad \text{and} \quad \mu\lambda_\varepsilon = -c_\varepsilon(x_\mu). \tag{6.22}$$

A useful interpretation of (6.21) is that we have perturbed the complementarity portions of the KKT conditions (6.3) corresponding to *both* inequality and equality constraints.

To complete the definition of $\pi(x,\mu)$ for the equality constraints (see (6.16)), we define

$$\pi_\mathcal{I}(x,\mu) \triangleq \mu \,/\cdot\, c_\mathcal{I}(x) \quad \text{and} \quad \pi_\varepsilon(x,\mu) \triangleq -c_\varepsilon(x)/\mu. \tag{6.23}$$

Application of Newton's method for equations to (6.21) gives

$$\begin{pmatrix} H & -J^T \\ ZJ & W \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = -\begin{pmatrix} g - J^T\lambda \\ W(\lambda - \pi) \end{pmatrix}, \tag{6.24}$$

where W and Z are defined in (6.13). The matrix in (6.24) can be symmetrized into the form (5.6), where $D = Z^{-1}W$; thus the discussion of section 5.1 applies and the linear algebraic work extends straightforwardly from the inequality-constrained case.

Treating equalities via a quadratic penalty function tends to regularize the problem in the sense that, as long as μ is nonzero, the matrix $(ZJ \;\; W)$ may have full row rank even if the Jacobian J_ε is rank-deficient. Consequently, one needs to modify only H to make the matrix in (6.24) nonsingular [39]. (A related regularization approach for general SQP methods is described in [117].)

We began this section with a discussion of unconstrained minimization of the classical penalty-barrier function Φ_{PB} (6.20), but the techniques presented have all been

based on solving the nonlinear equations (6.21) through Newton's method. As noted previously, either line-search or trust-region techniques must be applied to ensure convergence.

In the context of a line-search method, Φ_{PB} itself can be used as a merit function. Alternatively, the merit function $M^\mu(x,\lambda)$ of (5.8) can be generalized by adding

$$\frac{1}{2\mu}\|c_\varepsilon(x)\|_2^2 + \frac{1}{2\mu}\|c_\varepsilon(x) + \mu\lambda_\varepsilon\|_2^2, \tag{6.25}$$

which represents a combination of the original quadratic penalty term from (6.20) and a term that reflects proximity to the condition $\mu\lambda_\varepsilon = -c_\varepsilon(x_\mu)$ that holds along the penalty-barrier trajectory. If x_μ is a point on the penalty-barrier trajectory and λ_μ is the associated multiplier defined by (6.21), then (x_μ, λ_μ) is an unconstrained minimizer of $M^\mu(x,\lambda)$ for sufficiently small μ. See [39, 47] for further details.

6.4. The Slack and Shift Reformulations. We observed at the beginning of section 6 that an inequality constraint $c_i(x) \geq 0$ can be reformulated as two "equivalent" constraints—one inequality and one equality—by adding a new variable. The *slack* and *shift* reformulations of $c_i(x) \geq 0$ are

$$c_i(x) - s_i = 0, \quad s_i \geq 0 \quad \text{(slack)}, \tag{6.26a}$$

$$c_i(x) + s_i \geq 0, \quad s_i = 0 \quad \text{(shift)}. \tag{6.26b}$$

Both of these ideas have a long history in optimization. Slack variables pervade linear programming, and the idea of shifted constraints underlies the motivation in [86] of augmented Lagrangian methods for nonlinearly constrained optimization. We consider only shifts of the form (6.26b); for more general discussions of shifted constraints in interior methods, see, e.g., [23, 46, 60, 84, 98] and section 4.6.

The appeal of altering the constraints via either slacks or shifts may not be evident at first, since the dimension of the problem appears to have increased. However, two major benefits accrue:

1. Within an interior method, it is extremely simple to retain strict feasibility with respect to the bound constraint $s_i \geq 0$ imposed on a slack variable, since the step to the boundary of such a constraint along any direction can be calculated exactly. In contrast, the step to the boundary of a general nonlinear constraint must be calculated by iteration, so that maintaining strict feasibility may lead to many wasted evaluations of the constraint function at infeasible points.

2. The slack and shift reformulations allow interior-point methods to be applied to inequality constraints even if no initial strictly feasible point is known, as long as suitable techniques are available for dealing with equality constraints. Suppose that $c_i(x_0) < 0$ at the initial point x_0. For the shift reformulation, the constraint $c_i(x) + s_i \geq 0$ is strictly satisfied at x_0 for any s_i greater than $-c_i(x_0)$, and any positive value of the slack variable s_i is strictly feasible for the inequality $s_i \geq 0$. But the equality constraint $s_i = 0$ is not satisfied in the case of a shift, nor is $c_i(x_0) - s_i = 0$ in the slack case. Thus we can interpret both reformulations as placing the burden of ultimate feasibility on whatever method is used to satisfy the *equality* constraints.

For general constraint functions, these reformulations may dramatically change the behavior of an interior method—as we have repeatedly stressed, equality and inequality constraints are treated differently. With a shift reformulation, for example, if a shift variable ever becomes zero during the iterations of a barrier-SQP method (section 6.2.2), it will remain zero thereafter, which means that the original inequality constraint $c_i(x) \geq 0$ will be strictly satisfied at subsequent iterates. This is because

Δs_i, the component of the barrier-SQP search direction for the shift, satisfies the linearization $s_i + \Delta s_i = 0$; if $s_i = 0$, then $\Delta s_i = 0$ and the shift variable will never move away from zero. With a slack reformulation, on the other hand, even if $c_i(x_k) - s_i = 0$ at the kth iteration, this equality constraint is unlikely to remain satisfied at later iterations when $c_i(x)$ is nonlinear. It is not difficult to construct examples in which, starting from a strictly feasible point for the constraint $c_i(x) \geq 0$, the Newton step resulting from a slack reformulation subsequently violates the constraint even though the corresponding slack variable remains positive. (See [18] for a discussion of related issues.)

Other relevant aspects of these reformulations involve the nature of the feasible region. If the constraint function $c_i(x)$ is concave (Definition A.5), both reformulations preserve convexity of the feasible region for that constraint. With a slack reformulation, the feasible region defined by the inequality constraints $\{s_i \geq 0\}$ on the slack variables is convex, and any nonconvexity of the feasible set for the original problem is reflected in the equality constraints $c_i(x) - s_i = 0$. Certain difficulties associated with these reformulations will be considered in section 7.

As already noted, a strictly feasible initial point with respect to the (reformulated) inequality constraints can always be found. Thus, assuming that equality constraints can be coped with, there is no loss of generality in assuming that the initial point of the mixed-constraint problem (6.1) strictly satisfies any inequality constraints. (Of course, no such assurance can be given concerning feasibility for the equality constraints.)

It might seem that these reformulations lead to an undesirable increase in work because of the larger numbers of variables and constraints. Luckily, the linear algebraic calculations can be arranged so that there is no change in the dimension of the linear systems to be solved. Consider, for example, using a barrier-SQP method (section 6.2.2) to solve a problem that originally contained m inequality constraints, all of which have been reformulated using slack variables, and suppose that we need to solve the linear system (6.15). Following a straightforward elimination of the unknowns Δs (the changes in the m slack variable), the system to be solved may be written as

$$\begin{pmatrix} H & -J^T \\ \Lambda J & S \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \end{pmatrix} = - \begin{pmatrix} g - J^T \lambda \\ C\lambda - \mu e \end{pmatrix}, \tag{6.27}$$

where S is the positive diagonal matrix of slack values. After solving (6.27), Δs can be recovered from the relation

$$\Delta s = -s + \mu \Lambda^{-1} e - \Lambda^{-1} S \Delta \lambda.$$

Note that the system (6.27) has the same form as the primal-dual system (5.2) where no slack variables are present, except that the diagonal matrix C of constraint values in the $(2, 2)$ block of (5.2) is replaced by S in (6.27). Hence, the only extra expenses come from storing and updating the slack variables. Similar arguments can be made for the shift reformulation (see [96]).

7. Complications in the Nonconvex Case. For convex programming problems, interior methods display particularly nice properties; see, e.g., [3, 4, 79]. Even in this case, however, the trajectory may show undesirable properties if only smoothness and convexity are assumed [48].

The theoretical results for general nonconvex problems are weaker and mostly asymptotic, as exemplified by Theorems 3.10 and 3.12. In spite of this potential

concern, interior methods successfully and efficiently solve large nonconvex nonlinear programming problems every day, but the possibility of strange or even pathological behavior should not be ignored.

We have alluded previously to the fact that many interior-point methods are based on satisfying the first-order KKT conditions, but verification that a solution is a minimizer requires satisfaction of second-order conditions. An important issue in the nonconvex case is thus how to treat indefiniteness, a topic that would consume too many pages to allow its detailed treatment in this article.

A more fundamental difficulty arises at the theoretical level. In [106], a class of one-dimensional examples was recently defined for which certain line-search barrier-SQP methods (section 6.2.2) fail to find a feasible point and, much more seriously, do not even converge to a meaningful point.

The examples in [106] have the general form

$$\begin{array}{ll} \underset{x\in\mathbb{R}}{\text{minimize}} & f(x) \\ \text{subject to} & x^2+\alpha \geq 0, \quad x-\beta \geq 0, \end{array} \tag{7.1}$$

where $\beta \geq 0$ and f is unspecified for the moment.

If $\alpha \leq 0$, $x^2+\alpha$ can be written as $(x+\sqrt{-\alpha})(x-\sqrt{-\alpha})$, which means that the constraint $x^2+\alpha \geq 0$ is satisfied if and only if $x \geq \sqrt{-\alpha}$ or $x \leq -\sqrt{-\alpha}$. In effect, two disjunctive relations have been lumped together together into one constraint; see section 3.2 and the constraint of (3.16).

Suppose that a slack reformulation (section 6.4) is applied. Then we have

$$\begin{array}{ll} \underset{x\in\mathbb{R},s\in\mathbb{R}^2}{\text{minimize}} & f(x) \\ \text{subject to} & x^2+\alpha-s_1=0, \\ & x-\beta-s_2=0, \\ & s_1,s_2 \geq 0. \end{array} \tag{7.2}$$

A necessary ingredient of a barrier-SQP method is linearization of equality constraints, which for (7.2) means that Δx and Δs must satisfy

$$2x\Delta x - \Delta s_1 = -x^2+s_1-\alpha, \tag{7.3a}$$

$$\Delta x - \Delta s_2 = -x+s_2+\beta. \tag{7.3b}$$

If $\beta \geq 0$ and the initial x and s satisfy

$$x<0, \quad x<\beta+s_2, \quad \text{and} \tag{7.4a}$$

$$\alpha + \frac{\beta(x^2-s_1+\alpha)}{(x-s_2-\beta)} \leq \min\{\,0,-\tfrac{1}{2}\alpha\,\}, \tag{7.4b}$$

then any step $(\Delta x,\Delta s)$ satisfying (7.3) has the property that $s+\Delta s \not\geq 0$. It is further shown in [107] that a feasible point will never be reached by a line-search barrier-SQP method that maintains feasibility by reducing the step length, i.e., by choosing iterate $k+1$ as

$$x_{k+1}=x_k+\alpha_k\Delta x_k, \quad s_{k+1}=s_k+\alpha_k\Delta s_k,$$

where $\alpha_k>0$ and $s_k+\alpha_k\Delta s_k>0$. Instead, the full step $(\Delta x,\Delta s)$ persistently moves outside the region of strict feasibility for the slacks, the values of α_k converge to zero, and the iterates converge to a point without any apparent desirable properties (for instance, it does not minimize the constraint violations).

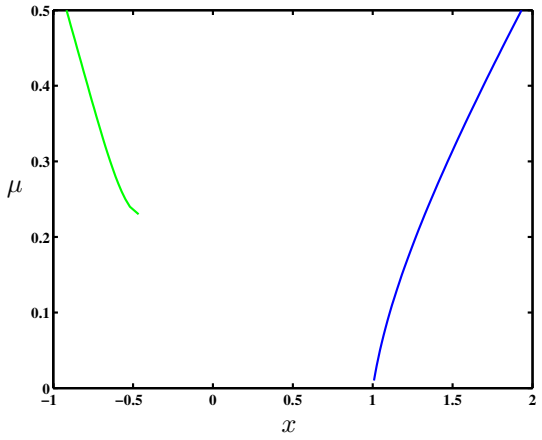

Fig. 7 *The figure illustrates two numerically computed penalty-barrier trajectories $x(\mu)$ of minimizers of (7.5). For each larger value of μ, there are two distinct minimizers, but there is only one minimizer below a threshold value of $\mu \approx 0.23$. The trajectory to the right leads to the optimal solution $x^* = 1$, whereas the trajectory to the left ceases to exist.*

Although these particular difficulties can be avoided, it is interesting to see what can go wrong in such a small and seemingly innocuous problem. It can also be shown that this form of "false convergence" implies that the multiplier search directions diverge [96]. For further discussion, see [9, 17, 68].

We now consider a particular instance of (7.1) with $f(x) = x$, $\alpha = -1$, and $\beta = \frac{1}{2}$, so that the optimal solution is $x^* = 1$ and the first constraint is active. The log barrier function for the original inequality-constrained problem is

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

and there is a well-defined trajectory of barrier function minimizers (see section 3.4) converging to x^* as $\mu \to 0$. In contrast, a penalty-barrier method applied to the slack formulation (7.2) seeks unconstrained minimizers of the penalty-barrier function (6.20):

$$\Phi_{\text{PB}}(x, s, \mu) = f(x) - \mu \sum_{i=1}^{2} \ln s_i + \frac{1}{2\mu}(x^2 - 1 - s_1)^2 + \frac{1}{2\mu}(x - \tfrac{1}{2} - s_2)^2. \tag{7.5}$$

This particular Φ_{PB} has *two* distinct local minimizers for larger values of μ, and so there are two penalty-barrier trajectories. As μ is reduced to approximately 0.23, one of these trajectories suddenly disappears at the point where the Hessian of the augmented penalty-barrier function of (6.25) becomes indefinite. This phenomenon is illustrated in Figure 7, which shows the x coordinates of the two penalty-barrier trajectories.

With a slack reformulation, the feasible region with respect to the inequality constraints, $s \geq 0$, is always convex, but a shift reformulation (see section 6.4) may lead to inequalities that specify a nonconvex, possibly even disconnected, feasible region. In such a situation, even if a step of unity along the search direction produces a point that satisfies the inequalities, there may exist points corresponding to smaller values of the step that do *not* satisfy the inequality constraints. The phenomenon of crossing the infeasible region with respect to the inequality constraints is called

tunneling, a phenomenon seen when analyzing the shift reformulation of (7.1) [96]. It is evident that even deciding on the best problem formulation can be extremely difficult for nonconvex problems.

8. Of Things Not Covered. Because of limitations on space and time, this article does not include everything that should be said about interior methods for nonlinearly constrained optimization. Our omission of any discussion of software is the most obvious lack. Several codes are available today (late 2002) that have been successfully applied to solve very large, nonlinear, nonconvex optimization problems; we urge interested readers to visit the Web site of "Optimization online" at www.optimization-online.org.

Other important aspects of interior methods that have not been discussed include complexity analyses for convex programs, the initial choice of barrier and/or penalty parameter, strategies for reducing the barrier parameter, criteria for terminating inner iterations, more detailed global convergence results, and methods that do not require second derivatives.

9. Summary. The interior-point revolution has had many highly positive results, including

- a deeper, more unified understanding of constrained optimization;
- continuing improvements in theory and methods;
- more algorithmic options for familiar problems;
- the ability to solve new problems; and
- closer ties between discrete and continuous optimization.

On the (possibly) negative side, life has become much more complicated for at least two classes of people: those who use optimization, and those who teach it.

Anyone wishing to solve a nonlinearly constrained problem, or even a linear program, is faced with a greatly expanded set of algorithmic options, especially since researchers continue to show that apparently different problem forms can be transformed into one another. Our focus throughout has been the very general area of nonlinearly constrained optimization, but newly emerging problem classes such as "mathematical programs with equilibrium constraints" (MPECs) are even more general. Beyond MPECs, there are nonlinear mixed integer programming problems, the extremely challenging union of continuous and discrete optimization. We have every expectation that interior-point methods will be helpful for both of these categories.

For teachers of optimization, the added pedagogical responsibilities are substantial; knowledge of Newton's method is needed today just to teach linear programming. But for those of us who enjoy nonlinearity and its difficulties, this is a privilege rather than a burden.

On balance, the interior-point revolution has unquestionably energized the field of constrained optimization, not to mention its creation of new connections with other areas of optimization and, more broadly, with numerical analysis, computer science, and scientific and engineering applications.

Appendix. Useful Definitions and Results. This appendix contains a set of definitions and results that may be useful for the reader.

DEFINITION A.1 (neighborhood of a point). *Given a point $\bar{x} \in \mathbb{R}^n$, a (closed)* neighborhood *of $\bar{x}$ is the set of points x satisfying $\|x - \bar{x}\| \leq \delta$ for some $\delta > 0$, with an analogous definition of an open neighborhood.*

DEFINITION A.2 (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 neighborhood of x that is entirely contained in S. The* interior *of S, denoted by* $\mathrm{int}(S)$*, is the collection of all interior points of S.*

DEFINITION A.3 (boundary of a set). *Given a set S, a point x is a* boundary point *of S if every neighborhood of x contains at least one point in S and at least one point not in S. The* boundary *of S, denoted by* $\mathrm{bnd}(S)$*, is the collection of all boundary points of S. A closed set contains all its boundary points.*

DEFINITION A.4 (convex set). *The set $S \subseteq \mathbb{R}^n$ is* convex *if, for every x_1 and x_2 in S and for all $\theta \in [0,1]$, the point $(1-\theta)x_1 + \theta x_2$ is also in S.*

DEFINITION A.5 (convex and concave functions). *The function $\varphi(x)$, defined for x in a nonempty open convex set S, is a* convex function *if, for every two points x_1 and x_2 in S and all $\theta \in [0,1]$,*

$$\varphi\big((1-\theta)x_1 + \theta x_2\big) \le (1-\theta)\varphi(x_1) + \theta\varphi(x_2).$$

The function $\varphi(x)$ is concave *if $-\varphi(x)$ is convex.*

LEMMA A.6 (special properties of smooth convex functions). *If the function $\varphi(x)$ is convex and twice-continuously differentiable on $\mathbb{R}^n$, then $\nabla^2\varphi(x)$ is positive semidefinite. Further, every local unconstrained minimizer $\bar{x}$ of φ is a* global *unconstrained minimizer in the sense that $\varphi(\bar{x}) = \{\min\ \varphi(x) : x \in \mathbb{R}^n\}$.*

LEMMA A.7 (optimality conditions for unconstrained optimization). *Consider unconstrained minimization of the twice-continuously differentiable function $f(x)$.*

(i) *If x^* is a local unconstrained minimizer of $f(x)$, then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semidefinite.*

(ii) *If $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite, x^* is an isolated local unconstrained minimizer of $f(x)$.*

Proofs of Lemma A.7 and related results may be found in, for example, [25, 28, 34, 52, 77, 80].

DEFINITION A.8 (convex program). *The optimization problem*

$$\text{(A.1)} \qquad \underset{x\in\mathbb{R}^n}{\text{minimize}}\ f(x) \quad \text{subject to} \quad c_i(x) = 0,\ i \in \mathcal{E}, \quad \textit{and} \quad c_i(x) \ge 0,\ i \in \mathcal{I},$$

is said to be a convex program*, or a* convex programming problem*, if*

(i) *f is convex,*

(ii) *all of the equality constraint functions $c_i(x)$, $i \in \mathcal{E}$, are* affine*, and*

(iii) *all of the inequality constraint functions $c_i(x)$, $i \in \mathcal{I}$, are* concave*.*

DEFINITION A.9 (order notation). *Let ϕ be a scalar, vector, or matrix function of a positive variable h, let p be fixed, and let κ_u and κ_l denote constants.*

(i) *If there exists $\kappa_u > 0$ such that $\|\phi\| \le \kappa_u h^p$ for all sufficiently small h, we write $\phi = O(h^p)$ and say, "ϕ is big oh of h^p."*

(ii) *If, for any $\epsilon > 0$, $\|\phi\|/h^p < \epsilon$ for all sufficiently small h, we write $\phi = o(h^p)$ and say, "ϕ is little oh of h^p."*

(iii) *If there exists $\kappa_l > 0$ such that $\|\phi\| \ge \kappa_l h^p$ for all sufficiently small h, we write $\phi = \Omega(h^p)$ and say, "ϕ is omega of h^p."*

(iv) *If there exist $\kappa_l > 0$ and $\kappa_u > 0$ such that $\kappa_l h^p \le \|\phi\| \le \kappa_u h^p$ for all sufficiently small h, we write $\phi = \Theta(h^p)$ and say, "ϕ is theta of h^p."*

DEFINITION A.10 (rates of convergence). *Let $\{x_k\}$ be a sequence that converges to x^*, with $x_k \ne x^*$, and assume that*

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} = \gamma_2, \quad \textit{where} \quad 0 < \gamma_2 < +\infty.$$

Then $\{x_k\}$ converges Q-quadratically *to x^*.*

A sequence $\{x_k\}$ converges to x^* Q-superlinearly *if there exists a sequence $\{\beta_k\}$ converging to zero such that for all sufficiently large k,*

$$\|x_{k+1} - x^*\| \le \beta_k \|x_k - x^*\|.$$

A Q-superlinearly convergent sequence $\{x_k\}$ thus satisfies $\|x_{k+1} - x^*\| = o(\|x_k - x^*\|)$.

DEFINITION A.11 (inertia of a matrix). *Given a symmetric matrix H, its* inertia, *denoted by* $\mathrm{In}(H)$, *is the integer triple* (i_p, i_n, i_z), *where* $i_p(H)$, $i_n(H)$, *and* $i_z(H)$ *are the numbers of positive, negative, and zero eigenvalues of* H.

LEMMA A.12 (Sylvester's law of inertia). *Given a symmetric matrix H and a nonsingular matrix U of the same dimension, then* $\mathrm{In}(H) = \mathrm{In}(UHU^T)$.

DEFINITION A.13 (directions of curvature). *Given a symmetric matrix H, we say that the vector p is a* direction of positive curvature *(with respect to H) if $p^THp > 0$, a* direction of zero curvature *if $p^THp = 0$, and a* direction of negative curvature *if $p^THp < 0$. When H is the Hessian of a nonlinear function f, a similar terminology applies with respect to f. For example, if $H = \nabla^2 f$, a direction p satisfying $p^THp > 0$ is said to be a* direction of positive curvature for f, *and so on. The scalar p^THp is sometimes called the* curvature of f along p.

LEMMA A.14 (Debreu's lemma [26]). *Given an $m \times n$ matrix J and an $n \times n$ symmetric matrix H, then $x^THx > 0$ for all nonzero x satisfying $Jx = 0$ if and only if there is a finite $\bar{\rho} \geq 0$ such that $H + \rho J^TJ$ is positive definite for all $\rho \geq \bar{\rho}$.*

LEMMA A.15 (inertia of the augmented system [56, Lemma 3.4]). *Given an $n \times n$ symmetric matrix H and an $m \times n$ matrix J, let r denote the rank of J and let N be a matrix whose columns form a basis for the null space of J. If we define K as*

$$K = \begin{pmatrix} H & J^T \\ J & 0 \end{pmatrix}, \quad \textit{then} \quad \mathrm{In}(K) = \mathrm{In}(N^THN) + (r, r, m - r).$$

LEMMA A.16 ([38, Proposition 2]). *Given an $n \times n$ symmetric matrix H, an $m \times n$ matrix J, and an $n \times n$ symmetric positive semidefinite matrix D, let r denote the rank of $\begin{pmatrix} J & -D \end{pmatrix}$. Further, let U_0 be a matrix whose columns form a basis for the null space of D and let N be a matrix whose columns form a basis for the null space of U_0^TJ. Finally, let $D^\dagger$ denote the pseudoinverse of D, m_0 the dimension of the null space of D, and define H_C as $H + J^TD^\dagger J$. Then* $\mathrm{rank}(U_0^TJ) = m_0 - m + r$ *and, if*

$$K \triangleq \begin{pmatrix} H & J^T \\ J & -D \end{pmatrix}, \quad \textit{then} \quad \mathrm{In}(K) = \mathrm{In}\left(N^TH_CN\right) + (m_0 - m + r, r, m - r).$$

Acknowledgments. We are grateful to Joshua Griffin, Dieter Kraft, and Göran Sporre for their reading of the manuscript and a number of helpful suggestions. We also appreciate the suggestions from an anonymous referee. Finally, we thank Nick Trefethen, our *SIAM Review* editor, for his insightful comments, encouragement, and enthusiasm.

REFERENCES

[1] E. L. ALLGOWER AND K. GEORG, *Continuation and path following*, in Acta Numerica, 1993, Cambridge University Press, Cambridge, UK, 1993, pp. 1–64.

[2] E. L. ALLGOWER AND K. GEORG, *Numerical path following*, in Handbook of Numerical Analysis, Vol. 5, P. G. Ciarlet and J. L. Lions, eds., North-Holland, Amsterdam, 1997, pp. 3–207.

[3] K. M. ANSTREICHER AND J.-P. VIAL, *On the convergence of an infeasible primal-dual interior-point method for convex programming*, Optim. Methods Softw., 3 (1994), pp. 273–283.

[4] P. ARMAND, J. C. GILBERT, AND S. JAN-JÉGOU, *A feasible BFGS interior point algorithm for solving convex minimization problems*, SIAM J. Optim., 11 (2000), pp. 199–222.

[5] C. ASHCRAFT, R. G. GRIMES, AND J. G. LEWIS, *Accurate symmetric indefinite linear equation solvers*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 513–561.

[6] M. S. BAZARAA, H. D. SHERALI, AND C. M. SHETTY, *Nonlinear Programming: Theory and Algorithms*, 2nd ed., John Wiley, New York, 1993.

[7] A. BENCHAKROUN, J.-P. DUSSAULT, AND A. MANSOURI, *A two parameter mixed interior-exterior penalty algorithm*, ZOR—Math. Methods Oper. Res., 41 (1995), pp. 25–55.

[8] H. Y. Benson, D. F. Shanno, and R. J. Vanderbei, *Interior-Point Methods for Nonconvex Nonlinear Programming: Filter Methods and Merit Functions*, Report ORFE-00-06, Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ, 2000.

[9] H. Y. Benson, D. F. Shanno, and R. J. Vanderbei, *Interior-Point Methods for Nonconvex Nonlinear Programming: Jamming and Comparative Numerical Testing*, Report ORFE-00-02, Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ, 2000.

[10] P. T. Boggs and J. W. Tolle, *Sequential quadratic programming*, in Acta Numerica, 1995, Cambridge University Press, Cambridge, UK, 1995, pp. 1–51.

[11] K.-H. Borgwardt, *The average number of pivot steps required by the simplex-method is polynomial*, Z. Oper. Res. Ser. A-B, 26 (1982), pp. A157–A177.

[12] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, Springer-Verlag, New York, 2000.

[13] M. G. Breitfeld and D. F. Shanno, *Computational experience with penalty-barrier methods for nonlinear programming*, Ann. Oper. Res., 62 (1996), pp. 439–463.

[14] R. H. Byrd, J. C. Gilbert, and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming*, Math. Program., 89 (2000), pp. 149–185.

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

[16] R. H. Byrd, G. Liu, and J. Nocedal, *On the local behavior of an interior point method for nonlinear programming*, in Numerical Analysis 1997 (Dundee, 1997), D. F. Griffiths, D. J. Higham, and G. A. Watson, eds., Addison Wesley Longman, Harlow, UK, 1998, pp. 37–56.

[17] R. H. Byrd, M. Marazzi, and J. Nocedal, *On the Convergence of Newton Iterations to Non-Stationary Points*, Report OTC 2001/01, Optimization Technology Center, Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, 2001.

[18] R. H. Byrd, J. Nocedal, and R. A. Waltz, *Feasible Interior Methods Using Slacks for Nonlinear Optimization*, Report OTC 2000/11, Optimization Technology Center, Northwestern University, Evanston, IL, 2000.

[19] C. W. Carroll, *The created response surface technique for optimizing nonlinear restrained systems*, Oper. Res., 9 (1961), pp. 169–184.

[20] V. Chvátal, *Linear Programming*, W. H. Freeman, New York, 1983.

[21] A. R. Conn, N. I. M. Gould, D. Orban, and P. L. Toint, *A primal-dual trust-region algorithm for non-convex nonlinear programming*, Math. Program., 87 (2000), pp. 215–249.

[22] A. R. Conn, N. I. M. Gould, and P. L. Toint, *A note on using alternative second-order models for the subproblems arising in barrier function methods for minimization*, Numer. Math., 68 (1994), pp. 17–33.

[23] A. R. Conn, N. I. M. Gould, and P. L. Toint, *A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds*, Math. Comp., 66 (1997), pp. 261–288, S1–S11.

[24] A. R. Conn, N. I. M. Gould, and P. L. Toint, *A primal-dual algorithm for minimizing a non-convex function subject to bound and linear equality constraints*, in Nonlinear Optimization and Related Topics (Erice, 1998), Kluwer Academic, Dordrecht, The Netherlands, 2000, pp. 15–49.

[25] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust-Region Methods*, SIAM, Philadelphia, PA, 2000.

[26] G. Debreu, *Definite and semidefinite quadratic forms*, Econometrica, 20 (1952), pp. 295–300.

[27] J. E. Dennis, Jr., M. Heinkenschloss, and L. N. Vicente, *Trust-region interior-point SQP algorithms for a class of nonlinear programming problems*, SIAM J. Control Optim., 36 (1998), pp. 1750–1794.

[28] J. E. Dennis, Jr., and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice–Hall, Englewood Cliffs, NJ, 1983.

[29] I. S. Duff, N. I. M. Gould, J. K. Reid, J. A. Scott, and K. Turner, *The factorization of sparse symmetric indefinite matrices*, IMA J. Numer. Anal., 11 (1976), pp. 181–204.

[30] J.-P. Dussault, *Numerical stability and efficiency of penalty algorithms*, SIAM J. Numer. Anal., 32 (1995), pp. 296–317.

[31] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang, *On the formulation and theory of the Newton interior-point method for nonlinear programming*, J. Optim. Theory Appl., 89 (1996), pp. 507–541.

[32] A. V. Fiacco, *Barrier methods for nonlinear programming*, in Operations Research Support Methodology, A. Holzman, ed., Marcel Dekker, New York, 1979, pp. 377–440.
[33] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming*, Classics Appl. Math. 4, SIAM, Philadelphia, PA, 1990. Reprint of the 1968 original.
[34] R. Fletcher, *Practical Methods of Optimization*, 2nd ed., John Wiley, Chichester, 1987.
[35] R. Fletcher, N. I. M. Gould, S. Leyffer, and P. L. Toint, *Global Convergence of Trust-Region SQP-Filter Algorithms for General Nonlinear Programming*, Tech. Report 99/03, Département de Mathématique, Facultés Universitaires Notre-Dame de la Paix (FUNDP), Namur, Belgium, 1999.
[36] R. Fletcher and S. Leyffer, *Nonlinear programming without a penalty function*, Math. Program., 91 (2002), pp. 239–269.
[37] R. Fletcher and A. P. McCann, *Acceleration techniques for nonlinear programming*, in Optimization, R. Fletcher, ed., Academic Press, London, New York, 1969, pp. 203–213.
[38] A. Forsgren, *Inertia-controlling factorizations for optimization algorithms*, Appl. Numer. Math., 43 (2002), pp. 91–107.
[39] A. Forsgren and P. E. Gill, *Primal-dual interior methods for nonconvex nonlinear programming*, SIAM J. Optim., 8 (1998), pp. 1132–1152.
[40] A. Forsgren, P. E. Gill, and W. Murray, *Computing modified Newton directions using a partial Cholesky factorization*, SIAM J. Sci. Comput., 16 (1995), pp. 139–150.
[41] A. Forsgren, P. E. Gill, and J. R. Shinnerl, *Stability of symmetric ill-conditioned systems arising in interior methods for constrained optimization*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 187–211.
[42] A. Forsgren and W. Murray, *Newton methods for large-scale linear equality-constrained minimization*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 560–587.
[43] R. Fourer and S. Mehrotra, *Solving symmetric indefinite systems in an interior-point method for linear programming*, Math. Program., 62 (1993), pp. 15–39.
[44] K. R. Frisch, *The Logarithmic Potential Method of Convex Programming*, memorandum, University Institute of Economics, Oslo, Norway, 1955.
[45] J. Gauvin, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Math. Program., 12 (1977), pp. 136–138.
[46] D. M. Gay, M. L. Overton, and M. H. Wright, *A primal-dual interior method for nonconvex nonlinear programming*, in Advances in Nonlinear Programming (Beijing, 1996), Y. Yuan, ed., Kluwer Academic, Dordrecht, The Netherlands, 1998, pp. 31–56.
[47] E. M. Gertz and P. E. Gill, *A Primal-Dual Trust Region Algorithm for Nonlinear Programming*, Numerical Analysis Report NA 02-1, University of California, San Diego, 2002.
[48] J. C. Gilbert, C. C. Gonzaga, and E. Karas, *Examples of ill-behaved central paths in convex optimization*, Math. Program., to appear.
[49] P. E. Gill, W. Murray, D. B. Ponceleón, and M. A. Saunders, *Solving Reduced KKT Systems in Barrier Methods for Linear and Quadratic Programming*, Report SOL 91-7, Department of Operations Research, Stanford University, Stanford, CA, 1991.
[50] P. E. Gill, W. Murray, D. B. Ponceleón, and M. A. Saunders, *Solving reduced KKT systems in barrier methods for linear programming*, in Numerical Analysis 1993 (Dundee, 1993), D. F. Griffiths and G. A. Watson, eds., Longman Scientific and Technical, Harlow, UK, 1994, pp. 89–104.
[51] P. E. Gill, W. Murray, M. A. Saunders, J. Tomlin, and M. H. Wright, *On projected Newton methods for linear programming and an equivalence to Karmarkar's projective method*, Math. Program., 36 (1986), pp. 183–209.
[52] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*, Academic Press, London, New York, 1981.
[53] P. E. Gill, M. A. Saunders, and J. R. Shinnerl, *On the stability of the Cholesky factorization for symmetric quasidefinite systems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 35–46.
[54] D. Goldfarb, R. Polyak, K. Scheinberg, and I. Yuzefovich, *A modified barrier-augmented Lagrangian method for constrained minimization*, Comput. Optim. Appl., 14 (1999), pp. 55–74.
[55] C. C. Gonzaga, *Path-following methods for linear programming*, SIAM Rev., 34 (1992), pp. 167–224.
[56] N. I. M. Gould, *On practical conditions for the existence and uniqueness of solutions to the general equality quadratic programming problem*, Math. Program., 32 (1985), pp. 90–99.
[57] N. I. M. Gould, *On the accurate determination of search directions for simple differentiable penalty functions*, IMA J. Numer. Anal., 6 (1986), pp. 357–372.
[58] N. I. M. Gould, *On the convergence of a sequential penalty function method for constrained minimization*, SIAM J. Numer. Anal., 26 (1989), pp. 107–128.

[59] N. I. M. Gould, D. Orban, A. Sartenaer, and P. L. Toint, *Superlinear convergence of primal-dual interior point algorithms for nonlinear programming*, SIAM J. Optim., 11 (2001), pp. 974–1002.
[60] F. Jarre and M. A. Saunders, *A practical interior-point method for convex programming*, SIAM J. Optim., 5 (1995), pp. 149–171.
[61] K. Jittorntrum, *Sequential Algorithms in Nonlinear Programming*, Ph.D. thesis, Department of Computer Science, Australian National University, Canberra, 1978.
[62] C. A. Johnson and A. Sofer, *A primal-dual method for large-scale image reconstruction in emission tomography*, SIAM J. Optim., 11 (2000), pp. 691–715.
[63] N. Karmarkar, *A new polynomial-time algorithm for linear programming*, Combinatorica, 4 (1984), pp. 373–395.
[64] V. Klee and G. J. Minty, *How good is the simplex algorithm?*, in Inequalities III, O. Shisha, ed., Academic Press, London, New York, 1972, pp. 159–175.
[65] A. S. Lewis and M. L. Overton, *Eigenvalue optimization*, in Acta Numerica, 1996, Vol. 5, Cambridge University Press, Cambridge, UK, 1996, pp. 149–190.
[66] F. A. Lootsma, *Hessian matrices of penalty functions for solving constrained-optimization problems*, Philips Res. Rep., 24 (1969), pp. 322–330.
[67] O. L. Mangasarian and S. Fromovitz, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints*, J. Math. Anal. Appl., 17 (1967), pp. 37–47.
[68] M. Marazzi and J. Nocedal, *Feasibility control in nonlinear optimization*, in Foundations of Computational Mathematics (Oxford, 1999), Cambridge University Press, Cambridge, UK, 2001, pp. 125–154.
[69] G. P. McCormick and C. Witzgall, *Logarithmic SUMT limits in convex programming*, Math. Program., 90 (2001), pp. 113–145.
[70] L. McLinden, *An analogue of Moreau's proximation theorem, with application to the nonlinear complementarity problem*, Pacific J. Math., 88 (1980), pp. 101–161.
[71] N. Megiddo, *Pathways to the optimal set in linear programming*, in Progress in Mathematical Programming: Interior Point and Related Methods, N. Megiddo, ed., Springer-Verlag, New York, 1989, pp. 131–158.
[72] J. J. Moré and D. J. Thuente, *Line search algorithms with guaranteed sufficient decrease*, ACM Trans. Math. Software, 20 (1994), pp. 286–307.
[73] W. Murray, *Analytical expressions for the eigenvalues and eigenvectors of the Hessian matrices of barrier and penalty functions*, J. Optim. Theory Appl., 7 (1971), pp. 189–196.
[74] W. Murray and M. H. Wright, *Line search procedures for the logarithmic barrier function*, SIAM J. Optim., 4 (1994), pp. 229–246.
[75] K. G. Murty and S. N. Kabadi, *Some NP-complete problems in quadratic and nonlinear programming*, Math. Program., 39 (1987), pp. 117–129.
[76] S. G. Nash, R. Polyak, and A. Sofer, *Numerical comparison of barrier and modified-barrier methods for large-scale bound-constrained optimization*, in Large-Scale Optimization: State of the Art, D. W. Hearn and P. M. Pardalos, eds., Kluwer Academic, Dordrecht, The Netherlands, 1994, pp. 319–338.
[77] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*, McGraw–Hill, New York, 1996.
[78] S. G. Nash and A. Sofer, *Why Extrapolation Helps Barrier Methods*, Tech. Report, George Mason University, Fairfax, VA, 1998.
[79] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM Stud. Appl. Math. 13, SIAM, Philadelphia, PA, 1994.
[80] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.
[81] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Classics Appl. Math. 30, SIAM, Philadelphia, PA, 2000. Reprint of the 1970 original.
[82] P. M. Pardalos and G. Schnitger, *Checking local optimality in constrained quadratic programming is NP-hard*, Oper. Res. Lett., 7 (1988), pp. 33–35.
[83] T. Plantenga, *A trust region method for nonlinear programming based on primal interior-point techniques*, SIAM J. Sci. Comput., 20 (1998), pp. 282–305.
[84] R. Polyak, *Modified barrier functions (theory and methods)*, Math. Program., 54 (1992), pp. 177–222.
[85] D. B. Ponceleón, *Barrier Methods for Large-Scale Quadratic Programming*, Ph.D. thesis, Department of Computer Science, Stanford University, Stanford, CA, 1990.
[86] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletcher, ed., London, Academic Press, New York, 1969, pp. 283–298.

[87] M. J. D. Powell, *Problems related to unconstrained optimization*, in Numerical Methods for Unconstrained Optimization, W. Murray, ed., Academic Press, London, New York, 1972, pp. 29–55.
[88] S. M. Robinson, *Generalized equations and their solutions.* II. *Applications to nonlinear programming*, Math. Programming Stud., 19 (1982), pp. 200–221.
[89] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[90] M. A. Saunders, *Major Cholesky would feel proud*, ORSA J. Comput., 6 (1994), pp. 23–27.
[91] M. A. Saunders, *Private communication*, 1998.
[92] A. Schrijver, *Theory of Linear and Integer Programming*, John Wiley, New York, 1986.
[93] D. F. Shanno and R. J. Vanderbei, *Interior-point methods for nonconvex nonlinear programming: Orderings and higher-order methods*, Math. Program., 87 (2000), pp. 303–316.
[94] S. Smale, *On the average number of steps of the simplex method of linear programming*, Math. Program., 27 (1983), pp. 241–262.
[95] D. A. Spielman and S.-H. Teng, *Smoothed analysis: Why the simplex method usually takes polynomial time*, in Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing, Crete, Greece, 2001, pp. 296–305. The full paper is available at www-math.mit.edu/~spielman/simplex.
[96] G. Sporre and A. Forsgren, *Relations between Divergence of Multipliers and Convergence to Infeasible Points in Primal-Dual Interior Methods for Nonconvex Nonlinear Programming*, Report TRITA-MAT-2002-OS7, Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden, 2002.
[97] A. L. Tits, A. Wächter, S. Bakhtiari, T. J. Urban, and C. T. Lawrence, *A Primal-Dual Method for Nonlinear Programming with Strong Global and Local Convergence Properties*, Tech. Report 2002–29, Institute for Systems Research, University of Maryland, College Park, MD, 2002.
[98] M. J. Todd, *Scaling, shifting and weighting in interior-point methods*, Comput. Optim. Appl., 3 (1994), pp. 305–315.
[99] M. J. Todd, *Semidefinite optimization*, in Acta Numerica, 2001, Cambridge University Press, Cambridge, UK, 2001, pp. 515–560.
[100] M. Ulbrich, S. Ulbrich, and L. Vicente, *A Globally Convergent Primal-Dual Interior-Point Filter Method for Nonlinear Programming*, manuscript, 2000.
[101] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Rev., 38 (1996), pp. 49–95.
[102] R. J. Vanderbei, *Symmetric quasidefinite matrices*, SIAM J. Optim., 5 (1995), pp. 100–113.
[103] R. J. Vanderbei and D. F. Shanno, *An interior-point algorithm for nonconvex nonlinear programming*, Comput. Optim. Appl., 13 (1999), pp. 231–252.
[104] J.-P. Vial, *Computational experience with a primal-dual interior-point method for smooth convex programming*, Optim. Methods Softw., 3 (1994), pp. 285–310.
[105] M. C. Villalobos, R. A. Tapia, and Y. Zhang, *The Sphere of Convergence of Newton's Method on Two Equivalent Systems from Nonlinear Programming*, Tech. Report CRPC-TR9915, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1999.
[106] A. Wächter and L. T. Biegler, *Failure of global convergence for a class of interior point methods for nonlinear programming*, Math. Program., 88 (2000), pp. 565–587.
[107] A. Wächter and L. T. Biegler, *Global and Local Convergence for a Class of Interior Point Methods for Nonlinear Programming*, Tech. Report B-01-09, CAPD, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA, 2001.
[108] D. P. Williamson, *Lecture Notes on Approximation Algorithms*, Research Report RC 21409, IBM T. J. Watson Research Center, Yorktown Heights, NY, 1998. Available online from www.almaden.ibm.com/cs/people/dpw.
[109] M. H. Wright, *Numerical Methods for Nonlinearly Constrained Optimization*, Ph.D. thesis, Department of Computer Science, Stanford University, Stanford, CA, 1976.
[110] M. H. Wright, *Interior methods for constrained optimization*, in Acta Numerica, 1992, Cambridge University Press, New York, 1992, pp. 341–407.
[111] M. H. Wright, *Some properties of the Hessian of the logarithmic barrier function*, Math. Program., 67 (1994), pp. 265–295.
[112] M. H. Wright, *Why a pure primal Newton barrier step may be infeasible*, SIAM J. Optim., 5 (1995), pp. 1–12.
[113] M. H. Wright, *Ill-conditioning and computational error in interior methods for nonlinear programming*, SIAM J. Optim., 9 (1998), pp. 84–111.
[114] S. J. Wright, *Stability of linear equations solvers in interior-point methods*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1287–1307.
[115] S. J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, PA, 1997.

[116] S. J. Wright, *Stability of augmented system factorizations in interior-point methods*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 191–222.
[117] S. J. Wright, *Superlinear convergence of a stabilized SQP method to a degenerate solution*, Comput. Optim. Appl., 11 (1998), pp. 253–275.
[118] S. J. Wright, *Modified Cholesky factorizations in interior-point algorithms for linear programming*, SIAM J. Optim., 9 (1999), pp. 1159–1191.
[119] S. J. Wright, *Effects of finite-precision arithmetic on interior-point methods for nonlinear programming*, SIAM J. Optim., 12 (2001), pp. 36–78.
[120] S. J. Wright, *On the convergence of the Newton/log barrier method*, Math. Program., 90 (2001), pp. 71–100.
[121] S. J. Wright and F. Jarre, *The role of linear objective functions in barrier methods*, Math. Program., 84 (1999), pp. 357–373.
[122] S. J. Wright and D. Orban, *Properties of the Log-Barrier Function on Degenerate Nonlinear Programs*, Preprint ANL/MCS-P772-0799, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, 1999.
[123] H. Yamashita and H. Yabe, *A primal-dual interior point method for nonlinear optimization: Global convergence, convergence rate and numerical performance for large scale problems*, in Parametric Optimization and Related Topics, V (Tokyo, 1997), Lang, Frankfurt am Main, Germany, 2000, pp. 213–250.
[124] Y. Zhang, R. Tapia, and F. Potra, *On the superlinear convergence of interior-point algorithms for a general class of problems*, SIAM J. Optim., 3 (1993), pp. 413–422.
[125] Y. Zhang and R. A. Tapia, *A superlinearly convergent polynomial primal-dual interior-point algorithm for linear programming*, SIAM J. Optim., 3 (1993), pp. 118–133.