A Primal-Dual Interior Method for Nonconvex Nonlinear Programming*

David M. Gay†
Michael L. Overton‡
Margaret H. Wright§

Dedicated to Michael J. D. Powell
on the occasion of his sixtieth birthday

"Is there any point to which you would wish to draw my attention?"
"To the curious incident of the dog in the night-time."
"The dog did nothing in the night-time."
"That was the curious incident," remarked Sherlock Holmes.[1]

Sir Arthur Conan Doyle, *Silver Blaze*

Abstract

Primal-dual interior methods for nonconvex nonlinear programming have recently been the subject of significant attention from the optimization community. Several different primal-dual methods have been suggested, along with some theoretical and numerical results. Although the underlying motivation for all of these methods is relatively uniform, there are nonetheless substantive variations in crucial details, including the formulation of the nonlinear equations, the form of the associated linear system, the choice of linear algebraic procedures for solving this system, strategies for adjusting the barrier parameter and the Lagrange multiplier estimates, the merit function, and the treatment of indefiniteness. Not surprisingly, each of these choices can affect the theoretical properties and practical performance of the method. This paper discusses the approaches to these issues that we took in implementing a specific primal-dual method.

1. Introduction

Primal-dual interior methods for nonconvex nonlinearly constrained optimization problems with both equality and inequality constraints have recently been discussed in [1, 2, 3, 6, 8, 11]. This paper provides some background on primal-dual methods and discusses how various algorithmic issues have been addressed in the authors' implementation of a specific method; no attempt is made to be comprehensive. Our results indicate that primal-dual methods are very promising, although several theoretical and practical issues remain to be resolved.

*This paper was presented at the International Conference on Nonlinear Programming, Beijing, China, September 1996, and will appear in the published proceedings of that conference.
†Bell Laboratories, Murray Hill, New Jersey 07974 (dmg@research.bell-labs.com).
‡New York University, New York, New York 10012 (overton@cs.nyu.edu).
§Bell Laboratories, Murray Hill, New Jersey 07974 (mhw@research.bell-labs.com).
[1] See Section 5.1.

We are interested in nonlinear nonconvex optimization problems with both equality and inequality constraints:

$$\underset{x\in\mathcal{R}^n}{\text{minimize}}\ f(x) \quad \text{subject to} \quad c(x)=0 \quad \text{and} \quad d(x)\geq 0, \tag{1.1}$$

where $f\colon\mathcal{R}^n \to \mathcal{R}$, $c\colon\mathcal{R}^n \to \mathcal{R}^m$, and $d\colon\mathcal{R}^n \to \mathcal{R}^p$ are smooth, and the relations $c(x)=0$ and $d(x)\geq 0$ in (1.1) are understood componentwise. The gradient of $f(x)$ is denoted by $g(x)$, the Jacobian of $c(x)$ by $A(x)$, and the Jacobian of $d(x)$ by $B(x)$. Exact second derivatives of f, c, and d are assumed to be available.

The *Lagrangian function* for problem (1.1) is defined as

$$\mathcal{L}(x,y,z) = f(x) - y^T c(x) - z^T d(x), \tag{1.2}$$

where y is an m-vector and z is a p-vector. The Hessian of the Lagrangian (with respect to x) is denoted by

$$W(x,y,z) \equiv \nabla^2 f(x) - \sum_{j=1}^{m} y_j \nabla^2 c_j(x) - \sum_{j=1}^{p} z_j \nabla^2 d_j(x). \tag{1.3}$$

We shall be concerned primarily with well-behaved local solutions of (1.1). Let x^* denote a point where the following conditions hold:

1. **Feasibility.** $c(x^*)=0$ and $d(x^*)\geq 0$.
2. **Constraint qualification.** The gradients of the constraints equal to zero at x^*—including all equalities and the active inequalities—are linearly independent.
3. **First-order Karush-Kuhn-Tucker (KKT) condition; complementarity.** There are vectors y^* and z^*, the Lagrange multipliers for the equality and inequality constraints, such that
$$g(x^*) = A^T(x^*)y^* + B^T(x^*)z^*, \tag{1.4}$$
where z^* satisfies
$$z^*\geq 0 \quad \text{and} \quad z_j^* d_j(x^*) = 0, \quad j=1,\ldots,p. \tag{1.5}$$
Using (1.2), condition (1.4) may be written as $\nabla_x\mathcal{L}(x^*,y^*,z^*)=0$, which means that x^* is a stationary point (with respect to x) of the Lagrangian function when $y=y^*$ and $z=z^*$.
4. **Strict complementarity.** $z_j^*>0$ if $d_j(x^*)=0$.
5. **Second-order KKT condition.** The matrix $N_J^{*T} W^* N_J^*$, the *reduced Hessian of the Lagrangian function*, is positive definite, where N_J^* is a basis for the null space of the Jacobian of all the constraints that are equal to zero at x^* (see (1.6) and (1.8)) and $W^*=W(x^*,y^*,z^*)$ is the Hessian of the Lagrangian (see (1.3)).

Under Conditions 1–5, it is well known (see, for example, [9] and [10]) that x^* is an isolated local constrained minimizer of problem (1.1) and that y^* and z^* are unique. We assume that at least one such isolated local constrained minimizer x^* of (1.1) exists.

Let $\hat{p}$ denote the number of inequality constraints equal to zero at x^*, and $\mathcal{A}$ the set of indices of those constraints, i.e., $d_i(x^*)=0$ for $i\in\mathcal{A}$ and $d_i(x^*)>0$ for $i\notin\mathcal{A}$. For any p-dimensional vector $v(x)$, $\hat{v}(x)$ denotes the $\hat{p}$-subvector of the components of $v(x)$ corresponding to indices in $\mathcal{A}$; for any matrix S with p rows, $\widehat{S}$ denotes the matrix whose $\hat{p}$ rows are the subset of rows of S corresponding to indices in $\mathcal{A}$. Thus $\hat{d}(x)$ contains the $\hat{p}$ components $d_i(x)$, $i\in\mathcal{A}$, and $\widehat{B}(x)$ denotes the $\hat{p}\times n$ Jacobian of $\hat{d}(x)$. Similarly, $\bar{v}(x)$ denotes the $(p-\hat{p})$-subvector of components of $v(x)$ corresponding to indices not in $\mathcal{A}$, with an analogous meaning for $\bar{S}$, so that $\bar{d}(x)$ contains the $p-\hat{p}$ components $d_i(x)$, $i\notin\mathcal{A}$, and $\bar{B}(x)$ denotes the Jacobian of $\bar{d}(x)$.

Some of our discussion involves subspaces associated with sets of constraint gradients. We define $J(x)$ as the $(m+\hat{p})\times n$ matrix

$$J(x) \equiv \begin{pmatrix} A(x) \\ \widehat{B}(x) \end{pmatrix}, \tag{1.6}$$

so that the rows are $J(x)$ are the gradients of all the constraints that are equal to zero at the solution. The constraint qualification and smoothness together imply that, when x is sufficiently close to x^*, the rows of $J(x)$ (and hence those of $A(x)$ and $\widehat{B}(x)$) are linearly independent. We assume henceforth that $J(x)$ has full rank, i.e., that $\operatorname{rank}(J) = m + \hat{p}$.

Let $N_A(x)$ ("null space of A") denote a matrix whose columns are an orthonormal basis for the null space of $A(x)$, and $N_{\hat{B}}(x)$ a matrix whose columns form an orthonormal basis for the null space of $\widehat{B}(x)$, so that

$$A(x)N_A(x) = 0 \quad \text{and} \quad \widehat{B}(x)N_{\hat{B}}(x) = 0. \tag{1.7}$$

Similarly, $N_J(x)$ denotes a matrix whose columns form an orthonormal basis for the null space of $J(x)$. Using (1.6), it follows that

$$J(x)N_J(x) = 0, \quad A(x)N_J(x) = 0, \quad \text{and} \quad \widehat{B}(x)N_J(x) = 0. \tag{1.8}$$

To avoid clutter, we may omit the argument (usually x) of vector or matrix functions, such as $g(x)$ or $A(x)$, when it is unnecessary or clear from context.

2. Primal-Dual Interior Methods

Our approach for locating x^* is to apply a primal-dual method to the inequality constraints of (1.1) and a sequential quadratic programming (SQP) approach to the equalities. Recent discussions of primal-dual methods for nonlinear problems are given in [1, 2, 3, 6, 8, 11].

Primal-dual interior methods can be interpreted in several ways; see, for example, [8] and the recent book [28] by S. Wright on primal-dual methods for linear programming. A particularly convenient interpretation for our purposes is based on the classical logarithmic barrier function associated with the objective function f and inequality constraints $d(x) \ge 0$:

$$\beta(x,\mu) = f(x) - \mu \sum_{j=1}^{p} \ln d_j(x), \tag{2.1}$$

where μ is a positive scalar. Interior methods for constrained optimization, many based on this barrier function, have been the subject of intense research since their revival in 1984; see [17, 23, 28]. For future reference we note that the barrier gradient and Hessian are given by

$$\nabla\beta(x,\mu) = g - \mu B^T D^{-1} e \quad \text{and} \quad \nabla^2\beta(x,\mu) = \nabla^2 f - \sum_{j=1}^{p} \frac{\mu}{d_j} \nabla^2 d_j + \mu B^T D^{-2} B, \tag{2.2}$$

where $e = (1, 1, \ldots, 1)^T$, all vector and matrix functions are evaluated at x, and $D(x)$ denotes the diagonal matrix $\operatorname{diag}(d_j(x))$.

The original problem (1.1) can be solved via a sequence of subproblems in which the barrier function (2.1), parameterized by μ, is minimized subject to the equality constraints $c(x) = 0$:

$$\operatorname{minimize} \beta(x,\mu) \quad \text{subject to} \quad c(x) = 0. \tag{2.3}$$

Note that the equality constraints of the original problem are carried over directly (without transformation) to the subproblem. Under Conditions 1–5 of Section 1, for a sequence of decreasing and sufficiently small barrier parameters μ, there is an associated sequence of isolated local minimizers x_μ of the equality-constrained subproblem (2.3) such that $A(x_\mu)$ has full rank and

$$c(x_\mu) = 0; \tag{2.4}$$

$$\nabla\beta(x_\mu,\mu) = A^T(x_\mu)y_\mu \quad \text{for some } m\text{-vector } y_\mu; \tag{2.5}$$

$$N_A^T\Big(\nabla^2\beta(x_\mu,\mu) - \sum_{j=1}^{m} y_\mu \nabla^2 c_j(x_\mu)\Big)N_A \quad \text{is positive definite;} \tag{2.6}$$

$$\lim_{\mu\to 0} x_\mu = x^*, \quad \lim_{\mu\to 0} y_\mu \to y^*, \quad \text{and} \quad \lim_{\mu\to 0} \frac{\mu}{d_j(x_\mu)} = z_j^*, \tag{2.7}$$

where y^* and z^* are the Lagrange multipliers appearing in (1.4). The points $\{x_\mu\}$ are said to define the barrier trajectory, or a (local) central path, for the subproblem (2.3). For proofs and additional details, see [9] and [23].

The third relation in (2.7) indicates that $\mu/d_j(x_\mu)$ serves as a Lagrange multiplier estimate for the jth inequality constraint. In this spirit, we define the positive p-vector z_μ as

$$z_\mu = \mu D(x_\mu)^{-1}e, \quad \text{so that its } j\text{th component satisfies} \quad (z_\mu)_j = \frac{\mu}{d_j(x_\mu)}. \tag{2.8}$$

Using (2.2), the first-order optimality condition (2.5) may be written in terms of z_μ as

$$g(x_\mu) - B^T(x_\mu)z_\mu = A^T(x_\mu)y_\mu. \tag{2.9}$$

Equation (2.8) may equivalently be written as

$$D(x_\mu)z_\mu = \mu e, \quad \text{or} \quad (z_\mu)_j\, d_j(x_\mu) = \mu. \tag{2.10}$$

Relation (2.10) is sometimes described as "centering" or "approximate complementarity"; the latter term arises because, as $\mu \to 0$, (2.10) imposes an increasingly accurate approximation of the complementarity condition $z_j^* d_j(x^*) = 0$ (1.5) that holds at optimality. Approximate complementarity can be represented in various mathematically equivalent forms. Of these, the form $Dz - \mu e = 0$ (2.10) is the "least nonlinear" in the sense that the associated primal-dual "Jacobian" matrix P (see (2.12)) is independent of μ and asymptotically reflects the condition of the original problem as $\mu \to 0$; for further discussion see [19, 23, 11].

Together, (2.9), (2.4), and (2.10) specify $n+m+p$ nonlinear equations that hold at (x_μ, y_μ, z_μ):

$$F_\mu(x_\mu, y_\mu, z_\mu) = 0, \quad \text{with} \quad F_\mu(x,y,z) = \begin{pmatrix} A^Ty + B^Tz - g \\ c \\ Dz - \mu e \end{pmatrix}, \tag{2.11}$$

where A, B, g, c, and d are evaluated at x. Note that the equations (2.11) do not capture the inequalities associated with (x_μ, y_μ, z_μ)—the sign restrictions on d and z, and the positive definiteness of (2.6).

Applying Newton's method to the nonlinear system (2.11), we obtain $n+m+p$ linear equations for the Newton steps $(\Delta x, \Delta y, \Delta z)$:

$$P\begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = \begin{pmatrix} g - A^Ty - B^Tz \\ -c \\ \mu e - Dz \end{pmatrix}, \quad \text{where} \quad P \equiv \begin{pmatrix} -W & A^T & B^T \\ A & 0 & 0 \\ ZB & 0 & D \end{pmatrix} \tag{2.12}$$

and $Z = \operatorname{diag}(z_j)$. The matrix P of (2.12) will be called the primal-dual matrix. In the context of interior methods for inequality-constrained problems, the designation "primal-dual" usually means that both the primal variables x and the dual variables z are treated as independent, although there are several interpretations of the term.

To the best of our knowledge, the idea of applying Newton's method to the full set of equations (2.11), even with $\mu = 0$, did not appear in the literature until the late 1980s. The book [28] gives some details of the history of primal-dual methods, mostly in the context of linear programming. In the 1970s, Robinson [20] proposed the idea of treating z as independent and applying Newton's method to the complementarity condition expressed as a single scalar equation $z^T d(x) = 0$ rather than the p separate equations $z_j d_j(x) = 0$ of (2.10). The suggestion of applying Newton's method to the square system (2.11) in the nonlinear programming context was apparently first made in a 1991 report by McCormick [19]; see also [23].

Expressing the subproblem (2.3) in primal-dual form is beneficial in several ways compared to a primal formulation involving only x and y. When the optimality conditions for (2.3) are expressed as nonlinear equations in terms of x and y, the associated Newton search direction is likely to be poorly scaled except when x is very close to x_μ for the current value of μ; for further details, see [25]. A primal-dual formulation can be motivated by other philosophical issues; see [8].

We sometimes use the shorthand notation

$$v = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad v^* = \begin{pmatrix} x^* \\ y^* \\ z^* \end{pmatrix}, \quad \text{and we define} \quad \delta = \|v - v^*\|, \tag{2.13}$$

so that δ measures the (unknown) distance from the current primal-dual iterate to the optimal solution.

3. Solving the Primal-Dual Equations

To calculate the primal-dual steps $(\Delta x, \Delta y, \Delta z)$, the unsymmetric linear system (2.12) can be solved at every iteration. An advantage of this approach is that, as already mentioned, the matrix P has a bounded condition number as $(x, y, z) \to (x^*, y^*, z^*)$ and $\mu \to 0$. Because the dimension of P is $n + m + p$, the associated linear algebraic work might be viewed as excessive, although efficient sparse matrix techniques should be able to exploit both the sparsity and symmetric structure (since Z is diagonal) of the matrix P. Comparatively little attention has been devoted to this possibility.

A more substantive difficulty arises when solving nonconvex problems. If the original problem (1.1) is convex (i.e., $f(x)$ is convex, $\{c_i(x)\}$ are linear, and $\{d_i(x)\}$ are concave), then the Lagrangian Hessian $W(x, y, z)$ is automatically positive semidefinite and every constrained stationary point is a minimizer. Otherwise, however, our experience indicates strongly that a robust second-derivative method for (1.1) needs to be able not only to verify the second-order optimality conditions of Condition 5 (see Section 1), but also to move away from non-minimizing stationary points. All methods known for checking positive-definiteness of the reduced Hessian rely on symmetry of an associated matrix (see [12]); no way has been devised to obtain this information from direct factorization of the unsymmetric matrix P in (2.12).

Symmetrizing P, as suggested in [13] and [11], produces an asymptotically ill-conditioned matrix

$$\widetilde{P} = \begin{pmatrix} -W & A^T & B^T \\ A & 0 & 0 \\ B & 0 & Z^{-1}D \end{pmatrix},$$

where the ill-conditioning arises from the ultimately unbounded elements of Z^{-1} corresponding to inactive constraints. It is shown in [13] that this ill-conditioning is benign, in the sense that the solution of the symmetrized primal-dual system is well determined. Furthermore, [13] discusses how a suitable symmetric indefinite factorization can be used to solve such systems accurately and to detect indefiniteness in a matrix related to the reduced Hessian of the Lagrangian. If indefiniteness is encountered, various modification strategies can be applied; see [13] and [11] for details.

Our strategy is different: we eliminate the $(1, 3)$ block of (2.12), producing an $(n+m) \times (n+m)$ symmetric linear system in Δx and Δy,

$$\begin{pmatrix} -K & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} h \\ -c \end{pmatrix}, \tag{3.1}$$

where

$$K = W + B^T D^{-1} Z B \quad \text{and} \quad h = g - A^T y - \mu B^T D^{-1} e. \tag{3.2}$$

We call K the *condensed* primal-dual "Hessian" matrix. It is shown in [26] that, under Conditions 1–5, K becomes increasingly ill-conditioned as the iterates converge; however, the ill-conditioning is usually benign, in the sense that the absolute accuracy of the computed search direction is only marginally worse than if it were computed from the original well conditioned system (2.12). The special structure of the ill-conditioning of K will be discussed in Section 4 when we consider indefiniteness.

For simplicity of exposition, we assume here that $\hat{p} > 0$, $m > 0$, and $0 < m + \hat{p} < n$. When $A(x)$ has full row rank, (3.1) can be solved by first computing the QR factorization of A^T:

$$A^T = QR = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 \\ 0 \end{pmatrix}, \tag{3.3}$$

where Q is orthogonal and R_1 is a nonsingular $m \times m$ upper-triangular matrix. It follows that

$$AQ_1 = R_1^T \quad \text{and} \quad AQ_2 = 0, \tag{3.4}$$

so that Q_1 is an orthonormal basis for the range of A^T, and Q_2 is an orthonormal basis for the null space of A. (The first relation in (1.7) thus holds with $N_A = Q_2$.) The step Δx may be written in two orthogonal pieces as

$$\Delta x = Q_1 \Delta x_1 + Q_2 \Delta x_2, \tag{3.5}$$

where $Q_1 \Delta x_1$ and $Q_2 \Delta x_2$ are the parts of Δx in the range of A^T and null space of A.

Because $AQ_2 = 0$, the second equation in (3.1), $A\Delta x = -c$, reduces to a nonsingular lower-triangular system that can be solved for Δx_1:

$$A\Delta x = A(Q_1 \Delta x_1 + Q_2 \Delta x_2) = AQ_1 \Delta x_1 = R_1^T \Delta x_1 = -c. \tag{3.6}$$

The first equation of (3.1) may be written in terms of Δx_1 and Δx_2 as

$$-K\Delta x + A^T \Delta y = -KQ_1 \Delta x_1 - KQ_2 \Delta x_2 + A^T \Delta y = h.$$

Since $Q_2^T A^T = 0$, multiplication by Q_2^T gives a linear system that expresses Δx_2 in terms of Δx_1:

$$Q_2^T K Q_2 \Delta x_2 = -Q_2^T h - Q_2^T K Q_1 \Delta x_1 = -Q_2^T g + \mu Q_2^T B^T D^{-1} e - Q_2^T K Q_1 \Delta x_1. \tag{3.7}$$

As we shall show in Section 4, $Q_2^T K Q_2$ will be nonsingular when (x, y, z) is sufficiently close to (x^*, y^*, z^*), in which case (3.7) uniquely determines Δx_2. The full vector Δx can then be computed from (3.5) using Q_1, Q_2, Δx_1, and Δx_2.

Because $d(x) > 0$, the third equation in the primal-dual system (2.12) uniquely determines Δz:

$$D\Delta z = \mu e - Dz - ZB\Delta x, \quad \text{or} \quad \Delta z = -z + D^{-1}(\mu e - ZB\Delta x). \tag{3.8}$$

It is shown in [26] that, even though elements of D corresponding to active constraints are tiny and possibly subject to large relative errors because of cancellation, Δz can be calculated to essentially full precision.

To obtain Δy, we substitute the relation $A^T = Q_1 R_1$ in the first row of (3.1):

$$A^T \Delta y = Q_1 R_1 \Delta y = h + K\Delta x,$$

so that Δy satisfies the nonsingular upper-triangular system

$$R_1 \Delta y = Q_1^T (h + K\Delta x). \tag{3.9}$$

4. Second-Order Information

A constrained minimizer of (1.1) is distinguished from a non-minimizing stationary point by non-negativity of the eigenvalues of $N_J^T W N_J$, the Hessian of the Lagrangian projected into the null space of the Jacobian of all binding constraints; see Condition 5 of Section 1. Since there is no explicit identification of the active inequality constraints in an interior method, it is not obvious how this condition could be checked in practice. However, we shall see that the influence of the active inequalities can be detected implicitly as the iterates converge.

This section first explains how information about positive-definiteness of the reduced Hessian of the Lagrangian function is obtained without predicting the active set, and then briefly discusses the strategies used when indefiniteness is encountered. Only line-search methods are considered here; trust region methods are a possible alternative (see [2], [3], and [6]).

4.1. Matrix relationships

The matrix containing second-order information in our formulation of the primal-dual equations is $Q_2^T K Q_2$ (see (3.2) and (3.7)), which we write here as $N_A^T K N_A$ since the columns of Q_2 form a basis for the null space of A. We now indicate why the matrix $N_A^T K N_A$ contains the information needed to check positive-definiteness of $N_J^T W N_J$. Unless stated otherwise, it is assumed in this section that δ, the distance from v to v^* in (2.13), is small enough so that asymptotic properties apply. In particular, smoothness implies that $A(x)$, $\widehat{B}(x)$, and $J(x)$ have full rank, and that their smallest singular values are bounded away from zero.

We may assume without loss of generality that the null-space basis matrices N_A and N_J are chosen so that $N_J(x)$ constitutes the last $n-m-\hat{p}$ columns of $N_A(x)$ (see (1.6)):

$$N_A(x) = \begin{pmatrix} U(x) & N_J(x) \end{pmatrix}. \tag{4.1}$$

The orthonormal columns of the $n \times \hat{p}$ matrix U, i.e., the first $\hat{p}$ columns of N_A, span the part of the null space of A that does not lie in the null space of $\widehat{B}$, which means that $\widehat{B}U$ is nonsingular. Furthermore, because the smallest singular value of $\widehat{B}$ is bounded away from zero, the smallest singular value of $\widehat{B}U$ is also bounded away from zero.

Separating the active and inactive inequality constraints, we write the condensed primal-dual Hessian K of (3.2) as

$$K(x,y,z) = W(x,y,z) + \widehat{B}^T \widehat{D}^{-1} \widehat{Z} \widehat{B} + \bar{B}^T \bar{D}^{-1} \bar{Z} \bar{B}. \tag{4.2}$$

(Although the active set is not known during the solution process, this representation allows us to analyze the structure of K.) Using the forms (4.1) and (4.2), and the relations $\widehat{B}N_J = 0$ (see (1.8)), $\bar{z} \to 0$, $\bar{d} \to \bar{d}* > 0$, $\hat{z} \to \hat{z}* > 0$, and $\hat{d} \to 0$, the matrix $N_A^T K N_A$ may be written as

$$N_A^T K N_A = \begin{pmatrix} U^T \\ N_J^T \end{pmatrix} K \begin{pmatrix} U & N_J \end{pmatrix}, \quad \text{or}$$

$$N_A^T K N_A = \begin{pmatrix} U^T \widehat{B}^T \widehat{D}^{-1} \widehat{Z} \widehat{B} U & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} U^T W U & U^T W N_J \\ N_J^T W U & N_J^T W N_J \end{pmatrix} + O(\|\bar{z}\|). \tag{4.3}$$

We first consider the matrix $U^T \widehat{B}^T \widehat{D}^{-1} \widehat{Z} \widehat{B} U$. All elements of $\hat{d}$ and $\hat{z}$ are positive, with $\hat{d} \to 0$ and $\hat{z} \to \hat{z}*$. As v converges to v^*, the elements of $\widehat{D}^{-1}$ are thus unbounded while the diagonal elements of $\widehat{Z}$ converge to finite positive values. Combined with the fact that the smallest singular value of $\widehat{B}U$ is bounded away from zero, these properties show that $U^T \widehat{B}^T \widehat{D}^{-1} \widehat{Z} \widehat{B} U$ is positive definite, and that all of its $\hat{p}$ eigenvalues are unbounded as $v \to v^*$.

Let $\bar{U}$ denote the first term on the right-hand side of (4.3), i.e.,

$$\bar{U} = \begin{pmatrix} U^T \widehat{B}^T \widehat{D}^{-1} \widehat{Z} \widehat{B} U & 0 \\ 0 & 0 \end{pmatrix}.$$

It is obvious that $\bar{U}$ has $\hat{p}$ positive eigenvalues equal to those of $U^T \widehat{B}^T \widehat{D}^{-1} \widehat{Z} \widehat{B} U$ and $n-m-\hat{p}$ zero eigenvalues. Furthermore, the first $\hat{p}$ columns of the $(n-m)$-dimensional identity matrix I_{n-m} define a simple invariant subspace for $\bar{U}$ corresponding to the $\hat{p}$ unbounded eigenvalues, and the last $n-m-\hat{p}$ columns of I_{n-m} form a simple invariant subspace corresponding to the $n-m-\hat{p}$ zero eigenvalues (see [22]). Based on assumptions and arguments similar to those in [26], it can be shown using (4.3) that, as $v \to v^*$, $N_A^T K N_A$ has $\hat{p}$ positive eigenvalues that are unbounded, and $n-m-\hat{p}$ eigenvalues that are within $O(\delta)$ (see (2.13)) of the eigenvalues of $N_J^T W N_J$—which is precisely the matrix that defines the second-order optimality condition.

This result is interesting for two reasons. First, the relationship between the eigenvalues of $N_A^T K N_A$ and $N_J^T W N_J$ does not depend on knowledge of the active set, but only on asymptotic properties of points near the barrier trajectory; hence the known matrix $N_A^T K N_A$ automatically provides guidance about the eigenvalues of the unknown matrix $N_J^T W N_J$ without any need to predict which inequality constraints are active. Second, it is reasonable for an algorithm to make decisions about potential indefiniteness of the reduced Lagrangian Hessian near the solution by checking positive-definiteness of $N_A^T K N_A$, although caution must be exercised in extracting information about the small eigenvalues.

4.2. Treatment of indefiniteness

Our treatment of an indefinite $N_A^T K N_A$ involves "modification" techniques from unconstrained optimization that are typically based on the Cholesky or symmetric indefinite factorizations; see [5] for a recent survey. If $N_A^T K N_A$ is positive definite, Δx_2 satisfies (3.7). If $N_A^T K N_A$ is not sufficiently positive definite, we first define a positive definite matrix $\widetilde{K}$ related to $N_A^T K N_A$, where the exact relationship between $N_A^T K N_A$ and $\widetilde{K}$ depends on the modification technique. The vector Δx_2 is then taken as the solution of

$$\widetilde{K} \Delta x_2 = -Q_2^T g + \mu Q_2^T B^T D^{-1} e - Q_2^T K Q_1 \Delta x_1. \tag{4.4}$$

Note that the right-hand side of this equation is zero when $c = 0$ and $N_A^T \nabla \beta = 0$, i.e., at a constrained stationary point for the subproblem (2.3). For all of the modification methods used, a direction of negative curvature $\widehat{\Delta x}$ satisfying $\widehat{\Delta x}^T N_A^T K N_A \widehat{\Delta x} < 0$ can also be calculated; details will appear in a forthcoming report [14]. Since $N_A^T K N_A$ ultimately becomes positive definite close to a minimizer, the expectation (confirmed in our numerical tests) is that no modification should be needed during the final iterations.

Our investigation of strategies dealing with indefiniteness in line-search primal-dual methods will appear separately [14]. Its three major conclusions are:

1. In solving nonconvex problems with a primal-dual line search method, possible indefiniteness in $N_A^T K N_A$ should be treated explicitly. If it is not, the iterates are likely to converge to a non-minimizing stationary point, especially if the merit function does not retain the element of minimization in the original problem. (See Section 5.) In solving problem HS33 [18], for example, a primal-dual method that does not treat indefiniteness becomes trapped near a non-minimizing stationary point.

2. In primal-dual methods that use the matrix K of (3.2) in a Newton-like setting, matrix modification techniques should preserve, as much as possible, the structure of the "large space" of K, i.e., the subspace associated with its large (ultimately unbounded) positive eigenvalues. Unless a good approximation to the portion of Δx lying in the large space is obtained, the computed search direction is likely to be poor.

3. The original Schnabel-Eskow modified Cholesky method [21] often produces a much larger modification than alternative methods when applied to a matrix with the special structure of $N_A^T K N_A$, namely the sum of a very large (in norm) rank-deficient positive semidefinite matrix and a much smaller unstructured symmetric matrix (see (4.3)).

5. The Line Search

In a line-search method, calculation of a search direction is followed by the line search itself, which determines a positive scalar α (the *steplength*) such that the current iterate is incremented by α times the search direction. (In Newton-based methods, $\alpha = 1$ is almost always the "ideal" value.) In order to choose α, we need to measure progress during a given iteration. A standard approach is to define a suitable *merit function* with the property that an unconstrained minimizer of the merit function is a solution of the desired problem, and then to require each new iterate to produce a reduction in the merit function. Our primal-dual method uses a variation of this strategy.

5.1. Merit functions and the watchdog strategy

The choice of merit function in nonlinearly constrained optimization is particularly complicated, and has been the subject of research for many years (see [10] and [4]). Definition of a suitable merit function for primal-dual interior methods has recently been considered by several researchers. For example, a merit function is proposed in [11] that uses the logarithmic barrier function for the inequalities and a classical quadratic penalty function for the equalities. A merit function based on the "quasi-central path" is given in [1].

As we have seen, the original problem (1.1), which involves minimization, is first transformed to the subproblem (2.3), which also involves minimization, and then to the nonlinear primal-dual equations $F_\mu(v) = 0$ of (2.11). If our intent were solely to find a solution of $F_\mu(v) = 0$, an obvious merit function would be the two-norm of F_μ (see [8]). However, a solution of $F_\mu(v) = 0$ is not necessarily a minimizer of the subproblem (2.3), so that reducing $\|F_\mu(v)\|$ does not express the important property of minimization with respect to the original problem.

Our strategy for defining a merit function is a variation of the "watchdog" technique first suggested in [4] in the context of nonlinearly constrained optimization. The idea is that there are two merit functions: a primary merit function whose decrease signifies almost-certain progress toward the solution; and a less demanding secondary merit function, here $\|F_\mu\|$, with the property that it will almost always be reduced if the method is taking full Newton steps (i.e., the steplength is unity) near the solution. An upper bound (a positive integer `watchmax`) is imposed on the number of iterations permitted without a reduction in the primary merit function.

During a given line search, each trial steplength is first tested to see whether the primary merit function has decreased. If so, the iteration is a success and the steplength is accepted. In addition, the steplength is accepted as long as the secondary merit function has decreased, except when `watchmax` iterations have occurred without a reduction in the primary merit function. If the primary merit function has not been decreased after `watchmax` iterations, "the watchdog barks", and the algorithm enforces a decrease in the primary merit function before continuing to the next iterate. In our experience, this approach seems to improve reliability at little additional cost because the "backtracking" safeguard is almost never invoked.

The primary merit function M is a combination of the logarithmic barrier function and an augmented Lagrangian term:

$$M(x; y, \rho) = f(x) - \mu \sum_{j=1}^{p} \ln d_j(x) - y^T c(x) + \tfrac{1}{2}\rho c(x)^T c(x). \tag{5.1}$$

This merit function reflects the goal that each iterate should be an improved estimate of the solution of subproblem (2.3). Descent is sought only with respect to x, in the spirit of maintaining a connection with the original problem.

Our numerical results indicate that the watchdog rarely affects the accepted steplength, but that presence of the watchdog improves reliability. Although no "barks" occurred for the problems and algorithm choices reported in Section 9, they remain a useful safeguard for certain treatments of indefiniteness; details will appear in [14].

5.2. The penalty parameter

The direction $(\Delta x, \Delta y, \Delta z)$ defined by (2.12) is a descent direction for the secondary merit function $\|F_\mu\|$ as long as the matrix P of (2.12) is nonsingular.

We next show that Δx can be guaranteed to be a descent direction for the primary merit function M if ρ is sufficiently large. The gradient of M with respect to x is

$$\nabla M = g - \mu B^T D^{-1} e - A^T y + \rho A^T c. \tag{5.2}$$

Since $\Delta x = Q_1 \Delta x_1 + Q_2 \Delta x_2$,

$$\Delta x^T \nabla M = \Delta x_1^T Q_1^T \nabla M + \Delta x_2^T Q_2^T \nabla M.$$

The vectors Δx_1 and Δx_2 satisfy $AQ_1 \Delta x_1 = -c$ and

$$\widetilde{K} \Delta x_2 = -Q_2^T g + \mu Q_2^T B^T D^{-1} e - Q_2^T K Q_1 \Delta x_1.$$

The matrix $\widetilde{K}$ is positive definite, but note that the matrix in the third term on the right-hand side is K rather than $\widetilde{K}$; see (3.6) and (4.4). (We assume here that Δx_2 does not include a direction of negative curvature.) Suppose first that $c = 0$; then $\Delta x_1 = 0$ and $\Delta x = Q_2 \Delta x_2$. Since $AQ_2 = 0$,

$$Q_2^T \nabla M = Q_2^T g - \mu Q_2^T B^T D^{-1} e = -\widetilde{K} \Delta x_2. \tag{5.3}$$

Using (5.3) and the positive-definiteness of $\widetilde{K}$, we have

$$\Delta x^T \nabla M = \Delta x_2^T Q_2^T \nabla M = -\Delta x_2^T \widetilde{K} \Delta x_2 < 0.$$

If $c \neq 0$, (5.2) and the relation $A \Delta x = -c$ together imply that

$$\Delta x^T \nabla M = \Delta x^T g - \mu \Delta x^T B^T D^{-1} e + c^T y - \rho c^T c. \tag{5.4}$$

It is easy to see from (5.4) that, if ρ is sufficiently large, $\Delta x^T \nabla M$ must be negative. In particular, let ρ_μ be defined as

$$\rho_\mu \equiv \frac{5}{c^T c} \left(\Delta x^T (g - \mu B^T D^{-1} e) + c^T y \right). \tag{5.5}$$

If $\rho \geq \rho_\mu$, then Δx is guaranteed to be a descent direction for M.

In practice, blindly choosing ρ to satisfy $\rho \geq \rho_\mu$ is undesirable because ρ can become unacceptably large if $\|c\|$ happens to be small. Since we do not require a decrease in the merit function at every iteration, our rule for choosing ρ is the following. At the beginning of every line search, ρ remains at its current value if $\Delta x^T \nabla M < 0$. If $\Delta x^T \nabla M \geq 0$, the value of ρ is redefined as

$$\rho = \min(\rho_\mu, \ \rho_{\max}), \tag{5.6}$$

where $\rho_{\max} = 100$ in our current implementation. If Δx is not a descent direction for M, we rely on the watchdog to ensure progress.

5.3. Choosing the steplength

We follow most primal-dual methods [28] in allowing separate steplengths in x and z. All iterates must satisfy $d(x) > 0$ and $z > 0$, and so a first requirement for the line search is to maintain strict feasibility. With linear constraints, it is possible to calculate exactly the steplength from the current point to the boundary of the feasible region. For nonlinear constraints, an estimate of this steplength can be obtained using a standard linearization. Let $\tilde{\alpha}$ and $\tilde{\gamma}$ be defined as

$$\tilde{\alpha} = \max\{\, \alpha \mid d(x) + \alpha B(x) \Delta x \geq 0 \} \quad \text{and} \quad \tilde{\gamma} = \max\{\, \gamma \mid z + \gamma \Delta z \geq 0 \}, \tag{5.7}$$

so that $\tilde{\alpha}$ estimates the distance to the boundary of the nearest inequality constraint $d(x) \geq 0$, and $\tilde{\gamma}$ is the exact distance along Δz to the boundary of the nonnegative orthant.

A frequently used technique for retaining strict feasibility in interior methods has been to choose the steplength as a fixed fraction τ (where τ is close to unity; for example, $\tau = 0.99$) of the distance to the boundary. However, the analysis of [8] shows that to preserve quadratic convergence, the steplengths taken in both x and z must converge sufficiently quickly to unity, thereby allowing full Newton steps to be taken near the solution. Hence there is a tradeoff between taking a "safe" step to ensure strict feasibility—this corresponds to a value of τ bounded away from unity—and taking the full Newton step near the solution. Following the guidelines of [8], our choice of τ depends on μ:

$$\tau = 1 - \min(0.01, \ 100\mu^2). \tag{5.8}$$

Let $\alpha^{(j)}$ and $\gamma^{(j)}$ denote the steplengths at subiteration j of the line search, with associated trial iterate $(x + \alpha^{(j)} \Delta x, \ \ y + \gamma^{(j)} \Delta y, \ \ z + \gamma^{(j)} \Delta z)$. (Note that the same steplength applies to y and z.) The initial steplengths are given by

$$\alpha^{(1)} = \min(1, \ \tau \tilde{\alpha}) \quad \text{and} \quad \gamma^{(1)} = \min(1, \ \tau \tilde{\gamma}), \tag{5.9}$$

where $\tilde{\alpha}$ and $\tilde{\gamma}$ are defined in (5.7) and τ is defined in (5.8). The initial steplength $\alpha^{(1)}$ of (5.9) may be infeasible because $\tilde{\alpha}$ of (5.7) is a poor estimate of the distance to the boundary. If $d_i(x+\alpha^{(j)}\Delta x) \le 0$ for any i, a steplength $\alpha^{(j+1)}$ that produces a strictly feasible point is found using a one-dimensional secant-like iteration. After a strictly feasible point is found, subsequent steplengths are modified using a simple Armijo-like rule; if further infeasibility is encountered, the steplength is modified accordingly.

Following a successful line search, x and z retain their final values from the line search. The new multiplier estimate for the equality constraints is then redefined as the least-squares solution of

$$y = \operatorname{argmin} \|A(x)y - g(x) + B(x)z\|. \tag{5.10}$$

6. Choosing the Barrier Parameter

An important issue in any primal-dual method is the choice of barrier parameter. According to the theory, μ must converge to zero in order for x_μ to converge to x^*. Classical barrier methods from the 1960s calculated x_μ accurately for each value of μ in a strictly decreasing sequence. In contrast, modern interior methods are based on performing an approximate calculation of x_μ, and ideally require only one Newton iteration for each μ.

The value of μ can be interpreted in two ways, as controlling the algorithm and measuring progress toward the solution. The Newton step $(\Delta x, \Delta y, \Delta z)$ is intended to be a step to (x_μ, y_μ, z_μ), and our choice of primary merit function (5.1) reflects this motivation. However, our ultimate goal is to find x^*, and the calculation of x_μ is only a means to that end. Any practical scheme for choosing μ thus involves estimating closeness to the solution and to the barrier trajectory. In order for the step to move toward the barrier trajectory and toward x_μ, μ should be *smaller* than the current distance to the solution.

The procedure given in [8] for choosing μ is based on approximate complementarity (2.10). We implemented a version of the μ-selection rule described in [8] with the added condition that μ cannot increase (which seemed to improve performance in our numerical experiments).

Procedure ETTZ:

$\sigma \leftarrow \min(0.2, 100 d^T z)$;
$\mu \leftarrow \min(\mu,\ \sigma d^T z / p)$;

An alternative strategy called "GOW" was tried, based on the idea that μ should be reduced if the algorithm has made progress during the just-completed iteration, where "progress" means that the line search took fewer than 3 subiterations. However, μ is reduced if the same value has been retained for more than 10 iterations. Close to the solution (as measured by $\|F_\mu\|$ of (2.11)), μ is reduced rapidly—by either squaring it if the current iterate seems to be very close to optimal, or dividing it by 10. If the current iterate is not viewed as close to the solution, μ is multiplied by 0.9.

Procedure GOW:

If (`lsteps` < 3) or (μ has remained constant for 10 iterations) then
  if $\|F_\mu\| < 10\mu$ then
    if $\mu < 10^{-4}$ then $\mu \leftarrow 10\mu^2$; else $\mu \leftarrow \mu/10$;
  else
    $\mu \leftarrow 0.9\mu$;

In our numerical tests, the GOW strategy often led to convergence with fewer iterations than ETTZ. For a few difficult problems, ETTZ was more efficient than GOW, but ETTZ is also much less efficient on other problems. Both approaches can fail when μ is prematurely made too small.

Let $\mu_{\rm ls}$ be defined as

$$\mu_{\rm ls} = \operatorname{argmin} \|N_A^T\big(g(x_0) - \mu B(x_0)^T D(x_0)^{-1} e\big)\|, \tag{6.1}$$

where N_A is a basis for the null space of $A(x_0)$. The initial μ is taken as

$$\min[100, \max(1, |\mu_{\mathrm{ls}}|)]. \tag{6.2}$$

7. Treatment of Infeasibility

Classical interior methods [9] assume a strictly feasible initial point x_0. Although such points are known for some problems, the default initial points for many test problems, including those reported in Section 9, are not strictly feasible. To cope with an infeasible starting guess, a standard strategy is to create an "equivalent" problem for which a strictly feasible starting point can easily be constructed.

There are many ways to define a modified problem, and none is perfect. Indeed, given two schemes for creating a modified problem, it is possible to find problems and starting points on which each outperforms the other. We describe here the approach that been most effective in our numerical experiments, but we stress that further research is needed on this aspect of primal-dual methods.

Given the initial point x_0, let $\mathcal{I}_{\mathrm{inf}}$ denote the set of indices of inequality constraints that are not strictly satisfied at x_0, and let $\mathcal{I}_+$ denote the complement of $\mathcal{I}_{\mathrm{inf}}$:

$$\mathcal{I}_{\mathrm{inf}} = \{i : d_i(x_0) \le 0\} \quad \text{and} \quad \mathcal{I}_+ = \{i : d_i(x_0) > 0\}. \tag{7.1}$$

Letting $\tilde{p} = |\mathcal{I}_{\mathrm{inf}}|$, we add $\tilde{p}$ "shift" variables s, where s is indexed corresponding to $\mathcal{I}_{\mathrm{inf}}$. The modified problem to be solved is

$$\begin{array}{ll} \underset{x,s}{\text{minimize}} & f(x) \\ \text{subject to} & c(x) = 0 \\ & s = 0 \\ & d_i(x) \ge 0, \quad i \in \mathcal{I}_+ \\ & d_i(x) + s_i \ge 0, \quad i \in \mathcal{I}_{\mathrm{inf}}. \end{array} \tag{7.2}$$

When the initial point is not strictly feasible with respect to the inequality constraints, the problem actually solved thus has $n + \tilde{p}$ variables (x and s), $m + \tilde{p}$ equality constraints ($c(x) = 0$ and $s = 0$), and p inequality constraints, where $\tilde{p}$ of the inequalities have been redefined according to (7.2).

The initial shift vector s_0 is chosen so that $d_i(x_0) + (s_0)_i > 0$. In the results reported in Section 9, the initial value of s_i is taken as $1 - 2d_i(x_0)$. Thus $d_i(x_0) + (s_0)_i = 1 - d_i(x_0)$, which must be strictly positive.

The problem transformation from (1.1) to (7.2) occurs before any primal-dual iterations are performed, and is "invisible" to the primal-dual code. Since the shift variables enter the determination of feasibility and choice of α during the line search (Section 5.3), a nice property of this treatment of infeasibility is that, after a unit value of α is accepted, the shifts become zero and the next x iterate is strictly feasible with respect to the original inequality constraints. Thereafter the shifts remain zero, with no further effect on the steps of the algorithm.

8. Pseudo-Code Summary of the Algorithm

The quantities `itmax` and `lsmax` specify limits on the number of iterations of the main loop and the line search loop respectively. The quantity `tol` is set to a small positive number.

Algorithm PD:

initialize μ according to (6.2);
choose x satisfying $d(x) > 0$ (see Section 7); $z_i \leftarrow \mu/d_i$, $i = 1,\ldots, p$; $\rho \leftarrow 10$; $y \leftarrow 0$;
`iter` $\leftarrow 0$; `watch` $\leftarrow 0$;
repeat
    `iter` $\leftarrow$ `iter` $+ 1$;
    compute the QR factorization of A^T; $N_A \leftarrow Q_2$; solve (3.6) for Δx_1;
    obtain $\widetilde{K}$ by modifying (if necessary) $N_A^T K N_A$; solve (4.4) for Δx_2;
    obtain Δx from (3.5), Δy from (3.9), and Δz from (3.8);
    find α and γ such that $d(x + \alpha\Delta x) > 0$ and $z + \gamma\Delta z > 0$ (see discussion after (5.9));
    if $\Delta x^T \nabla M(x; y, \rho) > 0$ (see (5.2)) then
        define ρ by (5.6);
    `lsteps` $\leftarrow 0$;
    if `watch` $=$ `watchmax` then
        restore $x, y, z, \Delta x, \Delta y, \Delta z, \alpha, \gamma, \rho$ and `lsteps` from `saved`;
        `watch` $\leftarrow$ `watch` $+ 1$;
        $\alpha \leftarrow \alpha/2$; $\gamma \leftarrow \gamma/2$;
    repeat
        `lsteps` $\leftarrow$ `lsteps` $+ 1$;
        if $M(x + \alpha\Delta x) < M(x)$ (see (5.1)) then
            `watch` $\leftarrow 0$;
            break out of repeat loop;
        if `watch` $<$ `watchmax` and $\|F_\mu(x + \alpha\Delta x, y + \gamma\Delta y, z + \gamma\Delta z)\| < \|F_\mu(x, y, z)\|$ then
            if `watch` $= 0$ then
                `Msave` $\leftarrow M(x; y, \rho)$;
                save $x, y, z, \Delta x, \Delta y, \Delta z, \alpha, \gamma, \rho$ and `lsteps` as `saved`;
            `watch` $\leftarrow$ `watch` $+ 1$;
            break out of repeat loop;
        $\alpha \leftarrow \alpha/2$; $\gamma \leftarrow \gamma/2$;
    until `lsteps` $>$ `lsmax`;
    if `watch` > 0 and $M(x + \alpha\Delta x; y, \rho) <$ `Msave` then
        `watch` $\leftarrow 0$;
    $x \leftarrow x + \alpha\Delta x$; $z \leftarrow z + \gamma\Delta z$; define y from (5.10);
    update μ by either procedure ETTZ or GOW (see Section 6);
until $(\|F_\mu(x, y, z)\| <$ `tol` and $\mu <$ `tol`$)$ or `iter` $>$ `itmax` or `lsteps` $>$ `lsmax`;

Note that because `watchmax` is positive, the *restore* operation in the outer loop cannot take place before the *save* operation in the inner loop. The latter belongs in the inner loop since it is possible that the secondary merit function may not satisfy the desired inequality until after one or more step length reductions have taken place. However, the *restore* operation belongs in the outer loop since `watch` cannot reach its upper limit `watchmax` without breaking out of the line search loop.

9. Numerical Results

Our primal-dual code is written in MATLAB, and all of our numerical tests were run on an SGI Challenge XL with twelve 150MHz R4400 processors, using IEEE double precision binary arithmetic, so that machine precision is approximately 1.1×10^{-16}. Results are reported for all the problems from the Hock-Schittkowski test set [18] that contain at least one nonlinear inequality constraint (in contrast to simple bounds or equalities), and for a small selection of the CUTE [7] problems that contain at least one nonlinear inequality constraint. Results are also given for three "interesting" problems—HS111 and HS118 from [18], and ORTHREGF from [7]. HS111 and ORTHREGF contain nonlinear equalities and simple bounds on the variables, and HS118 contains linear inequality constraints and simple bounds on the variables.

Table 1 summarizes the results for the algorithm with the initial μ chosen according to (6.2), the GOW rule for changing μ (see Section 6), and the Gill-Murray-Wright modified Cholesky procedure (see [16] and [14]). In all our runs, `itmax` $= 250$, `lsmax` $= 15$, `watchmax` $= 4$, and `tol` $= 10^{-8}$.

The column headings mean the following:

- Problem name.
- Starting point: "0" means that x_0 is the default provided in [18] or by CUTE [7]; "1" means that $x_0 = (1, \ldots, 1)^T$; "feas$_i$" means the ith strictly feasible point for the given problem (see the Appendix); "mu$_4$" means that x_0 is the default value, but the initial μ was set to 10^4.
- Return code: "`+*`" means a strictly feasible starting point and a successful solution; "`-*+`" means a starting point that is not strictly feasible, and a successful solution; "`--`" means a starting point that is not strictly feasible and a failure to solve the problem within 250 iterations.
- Total number of iterations. (This value is also the number of Hessian evaluations.)
- Total number of function/constraint evaluations. (This value is also the number of gradient/Jacobian evaluations.)
- Total number of times the watchdog mechanism was active, i.e., that a step was accepted that did not reduce the primary merit function; see Section 5.1.
- $\log_{10} \rho$, where ρ is the final penalty parameter; see Section 5.2.
- $||F_0||$, the norm of the right-hand side of the primal-dual equations (2.11) with $\mu = 0$.
- f, the final value of the objective function.
- the number of times the reduced Hessian K (3.2) was modified because of indefiniteness.
- $||E_{\max}||$, the largest modification, i.e., $||\widetilde{K} - K||$; see (3.2) and (4.4).
- cond(J), the condition number of J, the Jacobian of all the active constraints (1.6) at the final iteration, where an "active" constraint is defined as one whose value is less than 10^{-6}.
- cond(P), the condition number of the primal-dual matrix P (2.12) at the final iteration. This value is included solely to provide information; it is not routinely calculated as part of our algorithm.
- $\lambda_{\min}$, the smallest eigenvalue of the reduced Hessian (with respect to all the active constraints) at the final iteration; blank if the dimension of the reduced Hessian is zero.
- $\lambda_{\max}$, the largest eigenvalue of the reduced Hessian (with respect to all the active constraints) at the final iteration; blank if the dimension of the reduced Hessian is zero.

Several of the problems have special features of interest:

1. In problems HS33, HS71, HS83, and HS96 (marked with †), the reformulation associated with an infeasible starting point (see Section 7) changes the problem so that there are additional variables and equality constraints. Since the matrices J and P are different, even at the same final x, their condition numbers vary.
2. In problems HS85 and HS108 (marked with ‡), the reduced Hessian at the solution is singular; the tiny negative computed eigenvalues of the reduced Hessian shown in the tables would be zero with exact arithmetic.
3. With the given combination of algorithmic choices, problems HS88 and HS116 (marked with §) could not be solved within 250 iterations. However, solutions were obtained for these problems with other versions of our algorithm. Both problems are difficult: at the solution of HS88, cond(J) is 10^{10}; at the solution of HS116, both J and P are singular to within working accuracy.
4. In problems HS64 and HS72, the constraints include reciprocals of the independent variables and are highly nonconvex. The strategy described in Section 7 for treating infeasibility does not perform well from the default starting point. For a strictly feasible starting point, however, the algorithm converges with no difficulty on these problems.
5. Problems HS101–HS103 are essentially identical, and differ only in the values of a single constant. The constraints in these problems include fractional and negative powers of the independent variables, and cause difficulties for the default strategy for choosing μ (Section 6). We observed that forcing a large initial μ of 10^4 allowed our algorithm to converge without difficulty on these problems.

10. Summary and Conclusions

Implementation of a primal-dual method for nonconvex nonlinear programming involves several complicated algorithmic issues:

- formulation of the primal-dual equations and the primal-dual linear system;
- detection and treatment of indefiniteness in the reduced Hessian without predicting the active set;
- definition of a merit function;
- choice of the barrier parameter; and
- treatment of an infeasible starting point.

This paper has discussed in detail our motivation for decisions about these issues, defined a specific algorithm, and reported numerical results for the associated code.

All of the issues listed above deserve further research and experimentation. However, the reasonable performance of our algorithm suggests that the approaches described will be useful in developing a general-purpose primal-dual code.

References

[1] M. Argaez and R. A. Tapia (1997). On the global convergence of a modified augmented Lagrangian linesearch interior point Newton method for nonlinear programming, Report 95-38 (modified February 1997), Department of Computational and Applied Mathematics, Rice University, Houston, Texas.

[2] R. H. Byrd, J. C. Gilbert, and J. Nocedal (1996). A trust region method based on interior point techniques for nonlinear programming, Technical Report OTC 96/02, Argonne National Laboratory, Argonne, Illinois.

[3] R. H. Byrd, M. E. Hribar, and J. Nocedal (1997). An interior point algorithm for large-scale nonlinear programming, Technical Report, Northwestern University, Evanston, Illinois.

[4] R. M. Chamberlain, C. Lemaréchal, H. C. Pedersen, and M. J. D. Powell (1982). The watchdog technique for forcing convergence in algorithms for constrained optimization, *Math. Prog. Study* 16, 1–17.

[5] S. H. Cheng and N. J. Higham (1996). A modified Cholesky algorithm based on a symmetric indefinite factorization, Report 289, Department of Mathematics, University of Manchester, Manchester, United Kingdom.

[6] A. R. Conn, N. I. M. Gould, and P. L. Toint (1996). A primal-dual algorithm for minimizing a non-convex function subject to bound and linear equality constraints, Report RC 20639, IBM T. J. Watson Research Center, Yorktown Heights, New York.

[7] I. Bongartz, A. R. Conn, N. Gould, and Ph. Toint (1995). CUTE: Constrained and unconstrained testing environment, *ACM Trans. Math. Software* 21, 123–160.

[8] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang (1996). On the formulation and theory of the Newton interior-point method for nonlinear programming, *J. Opt. Theory Appl.* 89, 507–541.

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

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

[11] A. Forsgren and P. E. Gill (1996). Primal-dual interior methods for nonconvex nonlinear programming, Report NA 96-3, Department of Mathematics, University of California, San Diego.

[12] A. Forsgren, P. E. Gill, and W. Murray (1995). Computing modified Newton directions using a partial Cholesky factorization, *SIAM J. Sci. Comp.* 16, 139–150.

[13] A. Forsgren, P. E. Gill, and J. R. Shinnerl (1996). Stability of symmetric ill-conditioned systems arising in interior methods for constrained optimization, *SIAM J. Matrix Anal. Appl.* 17, 187–211.

[14] D. M. Gay, M. L. Overton, and M. H. Wright. Strategies for treating indefiniteness in a primal-dual method for nonconvex optimization, Technical Report, Bell Laboratories, Murray Hill, New Jersey, to appear.

[15] P. E. Gill and W. Murray (1974). Newton-type methods for unconstrained and linearly constrained optimization. *Mathematical Programming* 28, 311-350.

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

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

[18] W. Hock and K. Schittkowski (1981). *Test Examples for Nonlinear Programming Codes*, Springer-Verlag, Berlin.

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

[20] S. M. Robinson (1974). Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms, *Mathematical Programming* 7, 1–16.

[21] R. B. Schnabel and E. Eskow (1990). A new modified Cholesky factorization, *SIAM J. Sci. Stat. Comp.* 11, 1136–1158.

[22] G. W. Stewart and J. Sun (1990). *Matrix Perturbation Theory*, Academic Press, Boston.

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

[24] M. H. Wright (1994). Some properties of the Hessian of the logarithmic barrier function, *Mathematical Programming* 67, 265–295.

[25] M. H. Wright (1995). Why a pure primal Newton barrier step may be infeasible, *SIAM Journal on Optimization* 5, 1–12.

[26] M. H. Wright (1997). Ill-conditioning and computational error in primal-dual methods for nonlinear programming, Technical Report 97-4-04, Computing Sciences Research Center, Bell Laboratories, Murray Hill, New Jersey.

[27] S. J. Wright (1996). Modifying the Cholesky factorization in interior-point methods. Technical Report ANL/MCS-P600-0596, Argonne National Laboratory, Argonne, Illinois.

[28] S. J. Wright (1997). *Primal-Dual Interior-Point Methods*, Society for Industrial and Applied Mathematics, Philadelphia.

Appendix

The strictly feasible starting points indicated in Table 1 are the following:

HS34. feas$_1$: $x_0 = [\,10^{-4}, 1.05, 2.9\,]^T$

HS59. feas$_1$: $x_0 = [\,20, 60\,]^T$;

feas$_2$: $x_0 = [\,35, 60\,]^T$;

HS64. feas$_1$: $x_0 = [\,500, 500, 500\,]^T$;

feas$_2$: $x_0 = [\,2000, 2000, 2000\,]^T$;

HS65. feas$_1$: $x_0 = [\,0, 0, 0\,]^T$;

HS66. feas$_1$: $x_0 = [\,10^{-4}, 1.05, 2.9\,]^T$;

HS71. feas$_1$: $x_0 = [\,3, 3, 3, 3\,]^T$;

HS72. feas$_1$: $x_0 = [\,500, 500, 500, 500\,]^T$;

HS83. feas$_1$: $x_0 = [\,80, 34, 34, 34, 34\,]^T$;

HS95. feas$_1$: $x_0 = [\,.02, .02, .02, .02, .02, .01\,]^T$;

HS96. feas$_1$: $x_0 = [\,.02, .02, .02, .02, .02, .01\,]^T$;

HS105. feas$_1$: $x_0 = [\,0.1, 0.1, 110, 135, 180, 10, 10, 10\,]^T$;

HS108. feas$_1$: $x_0 = [\,0.2, 0.1, 0.1, 0.1, -0.1, 0.1, -0.2, 0.1, 0.1\,]^T$.