PRIMAL-DUAL INTERIOR METHODS FOR NONCONVEX NONLINEAR PROGRAMMING[*]

ANDERS FORSGREN[†] AND PHILIP E. GILL[‡]

Abstract. This paper concerns large-scale general (nonconvex) nonlinear programming when first and second derivatives of the objective and constraint functions are available. A method is proposed that is based on finding an approximate solution of a sequence of unconstrained subproblems parameterized by a scalar parameter. The objective function of each unconstrained subproblem is an augmented penalty-barrier function that involves both primal and dual variables. Each subproblem is solved with a modified Newton method that generates search directions from a primal-dual system similar to that proposed for interior methods. The augmented penalty-barrier function may be interpreted as a merit function for values of the primal and dual variables.

An inertia-controlling symmetric indefinite factorization is used to provide descent directions and directions of negative curvature for the augmented penalty-barrier merit function. A method suitable for large problems can be obtained by providing a version of this factorization that will treat large sparse indefinite systems.

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

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

PII. S1052623496305560

1. Introduction. This paper concerns second-derivative line-search methods for the solution of the nonlinear programming problem

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

where $c(x)$ is an m-vector of nonlinear functions with ith component $c_i(x)$, $i = 1, \ldots, m$, and $\mathcal{E}$ and $\mathcal{I}$ are nonintersecting index sets. Throughout, it is assumed that f and c are twice-continuously differentiable, with the gradient of $f(x)$ denoted by $g(x)$ and the $m \times n$ Jacobian of $c(x)$ denoted by $J(x)$.

The methods considered in this paper are all *interior methods* in the sense that, given an initial iterate x_0 in the strict interior of the set $\{x \mid c_i(x) \geq 0, \ i \in \mathcal{I}\}$, then all subsequent iterates lie in the strict interior. In recent years, interior methods for NP have received considerable attention because of their close relationship with the "new" polynomial approaches to linear and quadratic programming. The methods of this paper exploit some of this recent research, particularly in the area of primal-dual methods for convex programming.

The format for the constraints allows a general inequality constraint to be treated either directly as an inequality $c_i(x) \geq 0$ or indirectly as an equality and nonnegative

slack variable $c_i(x) - s_i = 0$, $s_i \geq 0$. We make this distinction because the methods considered here treat equality and inequality constraints differently: inequality constraints are evaluated only at feasible points, whereas equality constraints are feasible only as a solution is approached. This implies that the equality constraint $c_i(x) - s_i = 0$ is likely to be evaluated at a point for which $c_i(x) < 0$. The only exception occurs with linear equality constraints, which can be kept feasible at every iterate (see, e.g., Gill, Murray, and Wright [19]).

In many physical and engineering applications, the constraint functions not only characterize the desired properties of the solution but also define a region in which the problem statement is meaningful (for example, $f(x)$ or some of the constraint functions may be undefined outside the feasible region). In these applications, an interior point can usually be determined trivially. Interior methods are therefore highly appropriate for this class of problem. Ironically, interior methods were originally proposed for precisely this reason, but this feature was largely overlooked during the revival of interior methods because of the emphasis on computational complexity. One aspect of our research on interior methods has concerned the benefits provided by the knowledge of strictly feasible approximate solutions. In certain applications, the exploitation of feasibility may account for the superiority of some specialized methods compared to generic NLP methods (e.g., the method of moving asymptotes in structural optimization; see Svanberg [40]). The availability of general-purpose methods that generate interior points will considerably strengthen our ability to solve practical problems efficiently.

The proposed algorithm for NP is based on solving a sequence of unconstrained minimization problems parameterized by a positive parameter μ. In the classical penalty-barrier method given by Fiacco and McCormick [13], the unconstrained function

$$(1.1) \qquad M^\mu(x) = f(x) + \frac{1}{2\mu} \sum_{i \in \mathcal{E}} c_i(x)^2 - \mu \sum_{i \in \mathcal{I}} \ln c_i(x)$$

is minimized for a sequence of decreasing values of μ. The first constraint term on the right-hand side is the usual quadratic penalty function with penalty parameter $1/(2\mu)$. The second constraint term is the logarithmic barrier function, which creates a positive singularity at the boundary of the feasible region and thereby enforces strict feasibility while approaching the solution.

The penalty-barrier methods described here involve *outer* and *inner* iterations. Each outer iteration is associated with an element of a decreasing positive sequence of parameters $\{\mu_j\}$ such that $\lim_{j \to \infty} \mu_j = 0$. The inner iterations correspond to an iterative process for the unconstrained minimization of $M^\mu(x)$ for a given μ. The first-order optimality conditions for the unconstrained problem imply that there exist vectors $(x(\mu), \lambda(\mu))$ such that

$$(1.2) \quad g(x(\mu)) - J(x(\mu))^T \lambda(\mu) = 0, \quad \text{where} \quad \lambda_i(\mu) = \begin{cases} -c_i(x(\mu))/\mu & \text{if } i \in \mathcal{E}, \\ \mu/c_i(x(\mu)) & \text{if } i \in \mathcal{I}. \end{cases}$$

The m-vector $\lambda(\mu)$ can be interpreted as an estimate of λ^*, the Lagrange multipliers of NP, and Fiacco and McCormick [13] give conditions under which local solutions $(x(\mu), \lambda(\mu))$ of $M^\mu(x)$ converge to (x^*, λ^*) as $\mu \to 0$. When regarded as a function of the parameter μ, the set of minimizers $x(\mu)$ defines a continuously differentiable path known as the *trajectory* or *penalty-barrier trajectory*.

The direct unconstrained minimization of $M^\mu(x)$ involves a number of serious difficulties that all result from the fact that the minimization problem becomes increasingly ill conditioned as μ is decreased. These difficulties have been addressed in recent papers by Gould [22, 23], Conn, Gould, and Toint [8], and Dussault [11], who propose several modifications of the classical penalty-barrier algorithm. As in the original method, these modifications emphasize the minimization of $M^\mu(x)$ with respect to the "primal" variables x and define new estimates of $\lambda(\mu)$ as a by-product of the main computation. Dussault [11] observes that these multiplier estimates can be inaccurate when x is not close to the trajectory and proposes computing least-squares multipliers from an estimate of the active constraints at x^*.

In this paper we propose a different approach in which estimates of $x(\mu)$ and $\lambda(\mu)$ are updated simultaneously. This approach has been very successful for interior methods for convex programming, but its extension to the nonconvex case has been problematic because of the lack of a suitable *merit function* or *potential function* that forces convergence from an arbitrary starting point. Here we propose a merit function based on the classical penalty-barrier function augmented by a weighted *proximity measure* that measures the distance of (x, λ) to the trajectory $(x(\mu), \lambda(\mu))$. This *augmented penalty-barrier function* is denoted by $M^{\mu,\nu}(x, \lambda)$, where ν denotes the positive weight on the proximity measure (see section 3.1). The function $M^{\mu,\nu}(x, \lambda)$ has several important properties.

- The penalty-barrier solutions $x(\mu)$ and $\lambda(\mu)$ are *local minimizers of $M^{\mu,\nu}(x, \lambda)$*.
- $M^{\mu,\nu}(x, \lambda)$ can be minimized with respect to both x and λ using a modified Newton method that generates search directions from a symmetric *primal-dual* system similar to that proposed for interior methods. These search directions can be calculated in a numerically stable way, even when the Hessian $\nabla^2 M^{\mu,\nu}(x, \lambda)$ is not positive definite.
- For all positive μ and ν, under certain assumptions, the inner iterates converge to a point $x(\mu)$ satisfying the second-order necessary conditions for a minimizer of $M^{\mu,\nu}(x, \lambda)$.

The paper is organized as follows. In section 2 we review some basic properties of the classical penalty-barrier method and consider some recent developments in primal-dual interior methods for convex programming. The unsymmetric and symmetric primal-dual equations and their properties are discussed in sections 2.3 and 2.4. In section 3 we propose a new augmented penalty-barrier merit function and provide a modified Newton method for minimizing this function with respect to both the primal and dual variables. The linear system associated with this method is shown to be equivalent to the symmetric form of the primal-dual equations. It follows that $M^{\mu,\nu}(x, \lambda)$ may be interpreted as a merit function for primal and dual variables generated by the primal-dual system. Finally, in section 4 we describe how the search direction can be calculated using a certain inertia-controlling symmetric indefinite factorization of the primal-dual system. This factorization allows the efficient calculation of descent directions and directions of negative curvature for the merit function. A method suitable for large problems can be obtained by providing a version of this factorization that will treat large sparse indefinite systems.

In practice, the penalty-barrier subproblem is never solved to completion. Instead, the subproblem is terminated early, and the next value of μ is chosen so that the sequence of aggregated inner iterates converges to x^* at a rapid rate. However, although our choice of merit function is influenced by the desire for rapid local convergence, the principal focus is on the construction of the inner iterations, i.e., on the

minimization of $M^{\mu,\nu}(x,\lambda)$ for given μ and ν. Methods that terminate the subproblems early and define an associated decrease in μ will be considered in a future paper. Such methods rely heavily on the techniques discussed here.

2. Background.

2.1. Nomenclature. Given a real symmetric matrix A, the inertia of A—denoted by $In(A)$—is the associated integer triple (a_+, a_-, a_0) indicating the number of positive, negative, and zero eigenvalues of A. The largest and smallest eigenvalues of A are denoted by $\eta_{\max}(A)$ and $\eta_{\min}(A)$. The expression $A \succ 0$ means that A is positive definite. A member of a sequence of real symmetric matrices $\{A_k\}$ is said to be sufficiently positive definite if its eigenvalues are greater than a small positive number independent of k. The vector e_i is used to denote the ith unit vector of the appropriate dimension. The vector e denotes the vector of all ones whose dimension is defined by the context of its use.

2.2. The classical penalty-barrier function. Standard second-order linesearch or trust-region methods are easily adapted to find a local minimizer of M^μ (for details see, e.g., Dennis and Schnabel [10]). Regardless of the choice of method, as the iterates converge to an isolated local solution $x(\mu)$, the change in variables at iterate x is determined by the Newton equations $\nabla^2 M^\mu(x)\Delta x = -\nabla M^\mu(x)$, which define Δx as the step to the point that minimizes a quadratic model of M^μ. The Hessian $\nabla^2 M^\mu(x)$ and gradient $\nabla M^\mu(x)$ can be represented in terms of auxiliary quantities $\pi^\mu(x)$ and $\Omega^\mu(x)$ as follows. Let $\pi^\mu(x)$ denote the m-vector:

$$(2.1) \qquad \pi_i^\mu(x) = \begin{cases} -c_i(x)/\mu & \text{if } i \in \mathcal{E}, \\ \mu/c_i(x) & \text{if } i \in \mathcal{I}. \end{cases}$$

For any given μ and x, the vector $\pi^\mu(x)$ approximates $\lambda(\mu)$ (and hence λ^*). In particular, if $x = x(\mu)$, then $\pi^\mu(x) = \lambda(\mu)$, from (1.2).

Let $\Omega^\mu(x)$ denote the diagonal matrix $\mathrm{diag}(\omega^\mu(x))$, where

$$(2.2) \qquad \omega_i^\mu(x) = \begin{cases} 1/\mu & \text{if } i \in \mathcal{E}, \\ \mu/c_i(x)^2 & \text{if } i \in \mathcal{I}. \end{cases}$$

Given $\pi^\mu(x)$ and $\Omega^\mu(x)$, the derivatives of M^μ can be written as

$$(2.3\text{a}) \qquad \nabla M^\mu(x) = g(x) - J(x)^T \pi^\mu(x),$$
$$(2.3\text{b}) \qquad \nabla^2 M^\mu(x) = H(x, \pi^\mu(x)) + J(x)^T \Omega^\mu(x) J(x),$$

where $H(x,\lambda)$ denotes the Hessian with respect to x of the Lagrangian $f(x) - \lambda^T c(x)$.

Unconstrained methods for minimizing M^μ are well defined as long as the derivatives (2.3) are bounded. This will always be the case if c is evaluated at points that are strictly feasible with respect to the inequality constraints. For example, in a linesearch method, the initial estimate x_0 is chosen to lie in the strict interior of the region $\{x \mid c_i(x) \geq 0,\ i \in \mathcal{I}\}$, and subsequent line searches ensure that $c_i(x + \alpha \Delta x) > 0$ for all $i \in \mathcal{I}$ (see, e.g., Murray and Wright [37]).

Given the derivatives (2.3), the Newton equations can be written as

$$(2.4) \qquad \Big(H(x, \pi^\mu(x)) + J(x)^T \Omega^\mu(x) J(x)\Big)\Delta x = -(g(x) - J(x)^T \pi^\mu(x)).$$

Unfortunately, the direct minimization of M^μ cannot be recommended as a practical algorithm. As $\mu \to 0$, terms in the matrix $J(x)^T \Omega^\mu(x) J(x)$ associated with the gradients of the equality constraints and active inequality constraints can become unbounded. If the number $\bar{m}$ of unbounded terms satisfies $0 < \bar{m} < n$, then $\nabla^2 M^\mu(x)$ becomes increasingly ill conditioned as μ is reduced, and the inverse of $\nabla^2 M^\mu(x)$ is singular in the limit (see Murray [36] and Wright [41]).

The ill-conditioning of $\nabla^2 M^\mu$ not only affects the sensitivity of the numerical solution of (2.4) but also inhibits the final rate of convergence of the Newton iterates. In the next three sections we review some recent approaches that are intended to overcome these difficulties.

2.3. The primal-dual equations. The effects of inevitable ill-conditioning can be removed by exploiting some recent developments in interior methods for convex programming. These methods compute primal and dual variables that satisfy a system of nonlinear equations equivalent to a perturbed system of first-order optimality conditions.

For any value of μ, an associated point $(x(\mu), \lambda(\mu))$ on the penalty-barrier trajectory satisfies the $n + m$ equations

$$(2.5a) \qquad g(x) - J(x)^T \lambda = 0,$$

$$(2.5b) \qquad \mu \lambda_i + c_i(x) = 0, \quad i \in \mathcal{E},$$

$$(2.5c) \qquad c_i(x) \lambda_i - \mu = 0, \quad i \in \mathcal{I}.$$

These relations imply that $(x(\mu), \lambda(\mu))$ can be determined by solving $n + m$ nonlinear equations in the $n + m$ unknowns (x, λ) using an iterative method, which is usually some variant of Newton's method.

To derive a suitable form of Newton's method, it is helpful to rewrite the m relations (2.5b)–(2.5c) in vector form $W^\mu(x)(\lambda - \pi^\mu(x)) = 0$, where $\pi^\mu(x)$ is given by (2.1) and $W^\mu(x)$ is the diagonal matrix with diagonal entries

$$(2.6) \qquad w_i^\mu(x) = \begin{cases} \mu & \text{if } i \in \mathcal{E}, \\ c_i(x) & \text{if } i \in \mathcal{I}. \end{cases}$$

Conditions (2.5) imply that $(x(\mu), \lambda(\mu))$ solves the $n + m$ nonlinear equations $F^\mu(x, \lambda) = 0$, where $F^\mu(x, \lambda)$ and its Jacobian $F^{\mu\prime}(x, \lambda)$ are given by

$$F^\mu(x, \lambda) = \begin{pmatrix} g(x) - J(x)^T \lambda \\ W^\mu(x)(\lambda - \pi^\mu(x)) \end{pmatrix}, \quad F^{\mu\prime}(x, \lambda) = \begin{pmatrix} H(x, \lambda) & -J(x)^T \\ Z(\lambda) J(x) & W^\mu(x) \end{pmatrix},$$

and $Z(\lambda)$ is a diagonal matrix with diagonal entries

$$(2.7) \qquad z_i(\lambda) = \begin{cases} 1 & \text{if } i \in \mathcal{E}, \\ \lambda_i & \text{if } i \in \mathcal{I}. \end{cases}$$

Suppose that (x, λ) is an estimate of $(x(\mu), \lambda(\mu))$. Let π, g, c, J, H, and W denote the quantities $\pi^\mu(x)$, $g(x)$, $c(x)$, $J(x)$, $H(x, \lambda)$, and $W^\mu(x)$. Given (x, λ), the next iterate of Newton's method for a solution of $F^\mu = 0$ is $(x + \Delta x, \lambda + \Delta \lambda)$, where $(\Delta x, \Delta \lambda)$ satisfies the Newton equations

$$(2.8) \qquad \begin{pmatrix} H & -J^T \\ ZJ & W \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \end{pmatrix} = - \begin{pmatrix} g - J^T \lambda \\ W(\lambda - \pi) \end{pmatrix}.$$

Similar equations have been proposed for interior methods for convex programming (see, e.g., Kojima, Mizuno, and Yoshise [28], Lustig, Marsten, and Shanno [29], Megiddo [31], Mehrotra [32], Monteiro and Adler [33], and Gill et al. [18]) and non-convex programming (see, e.g., El-Bakry et al. [12] and Forsgren, Gill, and Shinnerl [16]).

The class of methods considered in this paper is characterized by the property that every iterate (x, λ) satisfies $c_i(x) > 0$ and $\lambda_i > 0$ for all $i \in \mathcal{I}$. (It is this property that has come to be associated with the idea of a "primal-dual" method.) If this property is satisfied at an initial point (x_0, λ_0), then it can be enforced at every subsequent iterate if we are prepared to restrict the step along each search direction. In this case, the next iterate will be $(x+\alpha\Delta x, \lambda+\alpha\Delta\lambda)$, where α is a step length such that $c_i(x + \alpha\Delta x) > 0$ and $\lambda_i + \alpha\Delta\lambda_i > 0$ for all $i \in \mathcal{I}$. (The primal-dual property is enforced automatically by the merit function proposed in section 3.1.) With this restriction on the choice of x and λ, the Newton equations (2.8) are known as the *unsymmetric primal-dual equations*. Note that the diagonal matrices Z and W are positive definite.

The primal-dual equations have the important property that, if the second-order sufficient conditions for optimality hold, the Jacobian $F^{\mu\prime}(x, \lambda)$ is nonsingular in a neighborhood of (x^*, λ^*) for μ sufficiently small (see Fiacco and McCormick [13]). This result has two important implications. First, it allows the application of the standard convergence theory for Newton's method, which states that the Newton iterates will converge quadratically in a sufficiently small neighborhood of $(x(\mu), \lambda(\mu))$ for μ sufficiently small.

The second benefit of a nonsingular primal-dual Jacobian concerns the quality of the first Newton iterate immediately after μ is reduced (to $\bar{\mu}$, say). In practice, each unconstrained minimization is done only approximately, with a termination condition that determines the accuracy of the unconstrained minimizers. An important feature of any implementation of a penalty-barrier method is that, in the limit, this termination condition should allow only one inner iteration for each value of μ. If this is to be the case, the solution of the first primal-dual system associated with $\bar{\mu}$ must provide an increasingly good estimate of $(x(\bar{\mu}), \lambda(\bar{\mu}))$. If second-order sufficiency holds at (x^*, λ^*), and μ is reduced to $\bar{\mu}$ at a point $(x(\mu), \lambda(\mu))$ on the trajectory, then the new primal-dual nonlinear system will be "close" to the old one for μ sufficiently small. This implies that the aggregated set of inner iterates will converge very rapidly, given a suitable scheme for reducing μ.

Alternative methods for $(x(\mu), \lambda(\mu))$ can be derived by linearizing equivalent forms of the perturbed condition (2.5c) (see, e.g., Ye [45], Gill et al. [18], and Gonzaga [21]). However, it must be emphasized that the favorable property mentioned above does not hold for these alternatives. For example, suppose that (2.5c) is replaced by the condition $\lambda_i - \mu/c_i(x) = 0$ for $i \in \mathcal{I}$; or, equivalently, $\lambda_i - \pi^\mu(x) = 0$ for $i \in \mathcal{I}$. The associated Newton equations for the combined constraints are

$$(2.9) \qquad \begin{pmatrix} H & -J^T \\ \Omega J & I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} g - J^T\lambda \\ \lambda - \pi \end{pmatrix},$$

where Ω is the diagonal matrix $\Omega^\mu(x)$ of (2.2). These equations are known as the *unsymmetric primal barrier equations* for (x, λ). If these equations are defined with $\lambda = \pi^\mu(x)$, and $\Delta\lambda$ is eliminated from the system, we obtain the classical penalty-barrier equations (2.4) for Δx. This implies that the solution of (2.9) can be used

to find a solution of (2.4) in a numerically stable way (see Gould [22], who first proposed this calculation for the corresponding row-scaled symmetric system). However, although the Newton direction can be calculated accurately, the Jacobian associated with the primal system (2.9) is unbounded as $\mu \to 0$, and the region of quadratic convergence shrinks with μ. Moreover, reducing μ may give a new Δx that differs substantially from the step to the minimizer on the trajectory (see Wright [42] for a discussion of this point in the context of the primal barrier method). For a discussion of the asymptotic rate of convergence of the classical penalty-barrier method, see Gould [23] and Dussault [11]. They show that for certain termination criteria, an initial inner iterate can be calculated so that the aggregated sequence of inner iterates converges at a two-step Q-quadratic rate. Strategies for modifying the classical barrier method are considered by Conn, Gould, and Toint [8].

2.4. The symmetric primal-dual system. An important implication of solving nonconvex problems is the prominent role of *symmetric* systems in the definition of the search direction. If the objective is not convex, even the verification of optimality requires knowledge of the inertia (i.e., the number of positive, negative, and zero eigenvalues) of the Hessian in the subspace orthogonal to the active constraint gradients. It is not at all obvious how the inertia of a matrix can be estimated efficiently without utilizing its symmetry.

The unsymmetric primal-dual equations can be symmetrized by premultiplying the last m rows by Z^{-1} and changing the sign of $\Delta\lambda$. This gives

$$(2.10) \qquad \begin{pmatrix} H & J^T \\ J & -D \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta\lambda \end{pmatrix} = - \begin{pmatrix} g - J^T\lambda \\ D(\lambda - \pi) \end{pmatrix},$$

where $D = Z^{-1}W$. Following the convention used in convex programming, we call these equations the *symmetric primal-dual equations*.

The diagonal matrix D in (2.10) plays a crucial role in the analysis. Formally, we write $D = D^\mu(x, \lambda)$, with

$$(2.11) \qquad d_i^\mu(x, \lambda) = \begin{cases} \mu & \text{if } i \in \mathcal{E}, \\ c_i(x)/\lambda_i & \text{if } i \in \mathcal{I}. \end{cases}$$

If the second-order sufficient conditions for optimality hold at (x^*, λ^*), then the elements of D^μ either go to zero or go to infinity as $\mu \to 0$. For equality constraints and active inequality constraints, $d_i^\mu(x(\mu), \lambda(\mu)) \to 0$, and for inactive inequality constraints, $d_i^\mu(x(\mu), \lambda(\mu)) \to \infty$. This implies that the symmetric system (2.10) is ill conditioned, with some diagonal elements becoming unbounded as $\mu \to 0$. Forsgren, Gill, and Shinnerl [16] show that this ill-conditioning is artificial in the sense that the true sensitivity of the solution is independent of the size of the large diagonals. In particular, they show that the primal-dual solution can be found accurately using a certain symmetric indefinite factorization of the primal-dual system (see section 4.1). Related stability discussions can also be found in Ponceleón [38] and Wright [43, 44].

2.5. Global convergence. Until now we have considered the *local* properties of the primal-dual method. In doing so, we have made two assumptions: (i) the iterates (x, λ) lie in a sufficiently small neighborhood of $(x(\mu), \lambda(\mu))$ for μ sufficiently small; and (ii) $(x(\mu), \lambda(\mu))$ lies on a trajectory of *local minimizers* of M^μ. In order to make the method suitable for general problems, it is necessary to use the primal-dual direction $(\Delta x, \Delta\lambda)$ in conjunction with a merit function that forces convergence from

an arbitrary starting point. Moreover, since second derivatives of f and c are being calculated at each iteration, it seems reasonable to expect that every limit point of the sequence of iterates should satisfy the *second-order* necessary conditions for optimality.

In the case of primal-dual *interior* methods (where it is assumed that all constraints are inequalities), various merit functions have been suggested. One particularly popular merit function is

$$M(x, \lambda) = \frac{1}{2} F^\mu(x, \lambda)^T F^\mu(x, \lambda),$$

where $F^\mu(x, \lambda)$ is the residual vector associated with the primal-dual equations (2.10). El-Bakry et al. [12] propose this merit function for general nonlinear programming and show that, under certain assumptions, it provides convergence to a point that satisfies the first-order conditions for optimality. In section 3.1, we propose a merit function with the potential of giving limit points that satisfy the second-order necessary optimality conditions.

3. Second-order penalty-barrier methods.

3.1. An augmented penalty-barrier merit function. For positive parameters μ and ν, consider the combined *augmented penalty and barrier function* $M^{\mu,\nu}(x, \lambda)$ such that

$$M^{\mu,\nu}(x, \lambda) = f(x) + \frac{1}{2\mu} \sum_{i \in \mathcal{E}} \left(c_i(x)^2 + \nu(c_i(x) + \mu\lambda_i)^2 \right)$$

(3.1)
$$- \mu \sum_{i \in \mathcal{I}} \left(\ln c_i(x) + \nu \left(\ln \left(\frac{c_i(x)\lambda_i}{\mu} \right) + 1 - \frac{c_i(x)\lambda_i}{\mu} \right) \right).$$

This function is well defined for all (x, λ) such that $c_i(x) > 0$ and $\lambda_i > 0$ for $i \in \mathcal{I}$.

For fixed x, the vector $\lambda = \pi^\mu(x)$ minimizes $M^{\mu,\nu}(x, \lambda)$ with respect to λ, where $\pi^\mu(x)$ is defined in (2.1). Moreover, $M^{\mu,\nu}(x, \pi^\mu(x)) = M^\mu(x)$, where $M^\mu(x)$ denotes the classical penalty-barrier function (1.1). This result implies that a point $(x(\mu), \lambda(\mu))$ on the penalty-barrier trajectory can be found by minimizing $M^{\mu,\nu}(x, \lambda)$ with respect to both x and λ. The following lemma makes this precise.

LEMMA 3.1. *Let $M^{\mu,\nu}(x, \lambda)$ be the augmented penalty-barrier function (3.1) defined with any positive μ and ν. A point (x, λ) such that $c_i(x) > 0$ and $\lambda_i > 0$ for $i \in \mathcal{I}$ is an unconstrained local minimizer of $M^{\mu,\nu}(x, \lambda)$ if and only if $(x, \lambda) = (x(\mu), \lambda(\mu))$. Furthermore, $\min_\lambda M^{\mu,\nu}(x, \lambda) = M^{\mu,\nu}(x, \pi^\mu(x)) = M^\mu(x)$, where $M^\mu(x)$ is the classical combined penalty-barrier function (1.1) and $\pi^\mu(x)$ is defined in (2.1).*

Proof. For x fixed, $M^{\mu,\nu}(x, \lambda)$ is a strictly convex separable function of λ for $\lambda > 0$. It is straightforward to verify that the minimizing λ is given by $\pi^\mu(x)$, where $\pi^\mu(x)$ is given by (2.1). Substituting these values in (3.1) gives $M^{\mu,\nu}(x, \pi^\mu(x)) = M^\mu(x)$, where $M^\mu(x)$ is given by (1.1). The definition of $x(\mu)$ as a minimizer of $M^\mu(x)$ in conjunction with the definition $\lambda(\mu) = \pi^\mu(x(\mu))$ gives the required result. □

The augmented penalty-barrier function $M^{\mu,\nu}(x, \lambda)$ can be interpreted as the classical combined penalty-barrier function augmented by a function that measures the proximity of (x, λ) to the penalty-barrier trajectory. The parameter ν defines the relative weight of the proximity measure.

The most important property of $M^{\mu,\nu}(x, \lambda)$ is that it is *minimized* at a point $(x(\mu), \lambda(\mu))$ on the trajectory. (We observe that this is in contrast to the saddle-point property typically associated with Lagrangians and augmented Lagrangians.)

It follows that the Fiacco–McCormick penalty-barrier method can be generalized to include dual variables by minimizing $M^{\mu,\nu}(x,\lambda)$ with respect to both x and λ using any unconstrained minimization method. In this case, a second-order method will require both the first and second derivatives of $M^{\mu,\nu}(x,\lambda)$. Differentiating $M^{\mu,\nu}(x,\lambda)$ with respect to both x and λ and using the auxiliary quantities $\pi^\mu(x)$ (2.1) and $D^\mu(x,\lambda)$ (2.11) give

$$(3.2a) \quad \nabla M^{\mu,\nu}(x,\lambda) = \begin{pmatrix} g(x) - (1+\nu)J(x)^T\pi^\mu(x) + \nu J(x)^T\lambda \\ \nu D^\mu(x,\lambda)(\lambda - \pi^\mu(x)) \end{pmatrix}, \quad \text{and}$$

$$(3.2b) \quad \nabla^2 M^{\mu,\nu}(x,\lambda) = \begin{pmatrix} \nabla^2_{xx}M^{\mu,\nu}(x,\lambda) & \nu J(x)^T \\ \nu J(x) & \nu \Gamma^\mu(\lambda) \end{pmatrix},$$

where $\Gamma^\mu(\lambda)$ is the diagonal matrix $\mathrm{diag}(\gamma^\mu(\lambda))$ such that

$$(3.3) \qquad\qquad \gamma_i^\mu(\lambda) = \begin{cases} \mu & \text{if } i \in \mathcal{E}, \\ \mu/\lambda_i^2 & \text{if } i \in \mathcal{I}. \end{cases}$$

The Hessian $\nabla^2_{xx}M^{\mu,\nu}(x,\lambda)$ is written in terms of the Lagrangian Hessian $H(x,\lambda)$. Differentiating $M^{\mu,\nu}(x,\lambda)$ twice with respect to x and collecting terms gives

$$\nabla^2_{xx}M^{\mu,\nu}(x,\lambda) = H\Big(x,(1+\nu)\pi^\mu(x) - \nu\lambda\Big) + (1+\nu)J(x)^T\Omega^\mu(x)J(x),$$

where $\Omega^\mu(x)$ is defined in (2.2).

Rather than use these derivatives directly within a "black-box" line-search or trust-region method, we propose that $\nabla^2 M^{\mu,\nu}(x,\lambda)$ be *approximated* by a symmetric matrix $S^{\mu,\nu}(x,\lambda)$ in such a way that the search direction is defined by the primal-dual equations (2.10). This strategy requires the use of a modified Newton method that can exploit an approximate Hessian. Such methods are described next.

3.2. Modified Newton methods using an approximate Hessian. We consider a certain class of modified Newton methods for the minimization of a function $M(v)$ with gradient $\nabla M(v)$ and Hessian $\nabla^2 M(v)$. For brevity, we consider a method similar to that discussed by Forsgren, Gill, and Murray [15]. However, the modified Newton methods of McCormick [30], Mukai and Polak [35], Kaniel and Dax [27], Moré and Sorensen [34], and Goldfarb [20] are all suitable alternatives.

The method generates a sequence $\{v_k\}_{k=0}^\infty$ of improving estimates of a local minimizer of M. At iteration k, a line search is performed along a nonzero direction Δv_k of the form

$$\Delta v_k = s_k + d_k,$$

where s_k and d_k are computed using a symmetric matrix $S(v)$ that approximates the Hessian $\nabla^2 M(v)$ at v_k. The crucial property of S is that the approximation is exact at any stationary point of M. In particular, we require that

$$\|S(v_k) - \nabla^2 M(v_k)\| = O(\|v_k - v^*\|),$$

where v^* is any stationary point of M. We emphasize that the exact Hessian $\nabla^2 M(v)$ is known at each step, but that we choose not to use it to compute the search direction (see section 3.1). Note that a regular modified Newton method is obtained if $S(v) = \nabla^2 M(v)$ for all v.

Although s_k and d_k are computed using $S(v_k)$, they are forced to have certain properties with respect to $\nabla M(v_k)$ and $S(v_k)$. In particular, we ensure that $\nabla M(v_k)^T s_k \leq 0$, $\nabla M(v_k)^T d_k \leq 0$, and $d_k^T S(v_k) d_k \leq 0$. Each s_k satisfies a positive-definite system of the form

$$\bar{S}(v_k) s_k = -\nabla M(v_k),$$

where $\bar{S}(v_k) = S(v_k) + E_k$ with $E_k = 0$ if $S(v_k)$ is sufficiently positive definite. The vector d_k is defined as a direction of negative curvature for $S(v_k)$ and is normalized so that $\nabla M(v_k)^T d_k \leq 0$. If $d_k^T \nabla^2 M(v_k) d_k > 0$, the vector d_k can be reset to zero. This is not essential, but it provides the convenient property that $\nabla^2 M(v_k)$ has at least one zero or negative eigenvalue if d_k is nonzero.

Once Δv_k has been determined, a line search is used to find a step length α_k such that $v_{k+1} = v_k + \alpha_k \Delta v_k$ and $M(v_k + \alpha_k \Delta v_k) < M(v_k)$. The principal role of the line search is to ensure that $\nabla M(v_k)^T s_k \to 0$ and $d_k^T \nabla^2 M(v_k) d_k \to 0$. If these conditions are satisfied, and the directions s_k and d_k are *sufficient* in the sense that $\{s_k\}$ and $\{d_k\}$ are bounded sequences that satisfy the conditions

(3.4a) $\nabla M(v_k)^T s_k \to 0 \;\Rightarrow\; \nabla M(v_k) \to 0 \quad \text{and} \quad s_k \to 0,$

(3.4b) $d_k^T \nabla^2 M(v_k) d_k \to 0 \;\Rightarrow\; \min\{\eta_{\min}(\nabla^2 M(v_k)), 0\} \to 0 \quad \text{and} \quad d_k \to 0,$

then every limit point of the sequence $\{v_k\}_{k=0}^{\infty}$ will satisfy the second-order necessary conditions for optimality (see Moré and Sorensen [34]). In our case, conditions are imposed on $d_k^T S(v_k) d_k$ rather than $d_k^T \nabla^2 M(v_k) d_k$. However, the properties of $S(v_k)$ ensure that $d_k^T S(v_k) d_k \to d_k^T \nabla^2 M(v_k) d_k$ if $\nabla M(v_k) \to 0$, and it follows that, in the limit, $S(v_k)$ is positive semidefinite if and only if $\nabla^2 M(v_k)$ is positive semidefinite.

To keep things simple, a backtracking line search is described here. (For a review of some alternatives, see Moré and Sorensen [34].) An initial step $\alpha_k = 1$ is reduced (if necessary) by a constant factor until the reduction in M is at least as large as a fixed factor τ $(0 < \tau < 1)$ of the reduction predicted by a model function formed from the first two or three terms of a Taylor-series approximation to $M(v_k + \alpha \Delta v_k)$. If $m_k(\alpha)$ denotes the univariate function $M(v_k + \alpha \Delta v_k)$, then a step length α_k is accepted if

$$m_k(\alpha) \leq m_k(0) + \tau \left(\alpha_k m_k'(0) + \frac{1}{2} \alpha_k^2 \min\{0, m_k''(0)\} \right),$$

where m' and m'' denote the first and second derivatives of $m(\alpha)$ with respect to α.

The modified Newton method can be summarized as follows.

ALGORITHM 3.1. *Modified Newton method.*
Specify τ such that $0 < \tau < \frac{1}{2}$;
Choose v_0;
$k \leftarrow 0$;
repeat
    Evaluate $M(v_k)$, $\nabla M(v_k)$;
    if $S(v_k)$ is positive semidefinite and $\nabla M(v_k) = 0$ **then**
        stop; [v_k satisfies the second-order necessary conditions for optimality.]
    else if $S(v_k)$ is positive semidefinite **then**
        Compute s_k such that $\nabla M(v_k)^T s_k < 0$; $d_k \leftarrow 0$;
    else
        [$S(v_k)$ must have at least one negative eigenvalue.]
        Compute s_k such that $\nabla M(v_k)^T s_k \leq 0$;

Compute d_k such that $\nabla M(v_k)^T d_k \leq 0$ and $d_k^T S(v_k) d_k < 0$;
end
$\Delta v_k \leftarrow s_k + d_k$;
$\alpha_k \leftarrow 1$; $m_k(\alpha_k) \leftarrow M(v_k + \alpha_k \Delta v_k)$;
do while $m_k(\alpha_k) > m_k(0) + \tau(\alpha_k m_k'(0) + \frac{1}{2}\alpha_k^2 \min\{0, m_k''(0)\})$
$\quad \alpha_k \leftarrow \alpha_k/2$; $m_k(\alpha_k) \leftarrow M(v_k + \alpha_k \Delta v_k)$;
end do
$v_{k+1} \leftarrow v_k + \alpha_k \Delta v_k$;
$k \leftarrow k + 1$;
until converged;

3.3. The primal-dual method. Our aim is to use the primal-dual equations (2.10) to define the search direction but to relate this search direction to the augmented penalty-barrier merit function (3.1).

Consider the matrix $S^{\mu,\nu}(x, \lambda)$ obtained by approximating $\pi^\mu(x)$ by λ, $\Omega^\mu(x)$ by $D^\mu(x, \lambda)^{-1}$, and $\Gamma^\mu(\lambda)$ by $D^\mu(x, \lambda)$ in the matrix $\nabla^2 M^{\mu,\nu}(x, \lambda)$ of (3.2b). Then

$$(3.5) \quad S^{\mu,\nu}(x, \lambda) = \begin{pmatrix} H(x, \lambda) + (1+\nu)J(x)^T D^\mu(x, \lambda)^{-1}J(x) & \nu J(x)^T \\ \nu J(x) & \nu D^\mu(x, \lambda) \end{pmatrix},$$

and $S^{\mu,\nu}(x, \lambda) \to \nabla^2 M^{\mu,\nu}(x, \lambda)$ as $(x, \lambda) \to (x(\mu), \lambda(\mu))$. It follows that $S^{\mu,\nu}(x, \lambda)$ can be used as an approximate Hessian in a modified Newton method of the type suggested in section 3.2.

The next result provides the theoretical basis for the new method. It indicates that the modified Newton direction based on the approximate Hessian $S^{\mu,\nu}(x, \lambda)$,

$$(3.6) \quad S^{\mu,\nu}(x, \lambda)\begin{pmatrix} \Delta x \\ \Delta \lambda \end{pmatrix} = -\nabla M^{\mu,\nu}(x, \lambda),$$

satisfies the primal-dual equations (2.8) or, equivalently, the symmetric primal-dual equations (2.10). Moreover, the approximation $S^{\mu,\nu}(x, \lambda)$ is exact whenever (x, λ) is a stationary point of $M^{\mu,\nu}$, i.e., when $\nabla M^{\mu,\nu}(x, \lambda) = 0$.

THEOREM 3.2. *Let $M^{\mu,\nu}(x, \lambda)$ be the augmented penalty-barrier function (3.1) defined with any positive μ and ν. Let $S^{\mu,\nu}(x, \lambda)$ denote the approximate Hessian (3.5). If $c_i(x) > 0$ and $\lambda_i > 0$ for all $i \in \mathcal{I}$, then the vector $(\Delta x, \Delta \lambda)$ solves the modified Newton equations (3.6) if and only if it solves the primal-dual equations (2.8) and the symmetric primal-dual equations (2.10). Moreover, if $\nabla M^{\mu,\nu}(x, \lambda) = 0$, then $S^{\mu,\nu}(x, \lambda) = \nabla^2 M^{\mu,\nu}(x, \lambda)$.*

Proof. If $c_i(x) > 0$ and $\lambda_i > 0$ for $i \in \mathcal{I}$, and $\mu > 0$, the diagonal matrices $D^\mu(x, \lambda)$ (2.11) and $Z(\lambda)$ (2.7) are positive definite. The definiteness of $Z(\lambda)$ implies that the unsymmetric primal-dual equations (2.8) and the symmetric primal-dual equations (2.10) are equivalent. Hence, it suffices to consider the symmetric equations (2.10). To simplify the notation, let $H = H(x, \lambda)$, $D = D^\mu(x, \lambda)$, $J = J(x)$, $g = g(x)$, and $\pi = \pi^\mu(x, \lambda)$. As ν is nonzero and D is positive definite, the $(n+m) \times (n+m)$ block upper-triangular matrix

$$\begin{pmatrix} I & (1+\nu)J^T D^{-1} \\ 0 & \nu I \end{pmatrix}$$

is well defined, nonsingular, and can be applied to both sides of the modified Newton equation. The first part of the theorem follows directly from the identities

$$
\begin{pmatrix} I & (1+\nu)J^TD^{-1} \\ 0 & \nu I \end{pmatrix} \begin{pmatrix} H & -J^T \\ J & D \end{pmatrix} = \begin{pmatrix} H + (1+\nu)J^TD^{-1}J & \nu J^T \\ \nu J & \nu D \end{pmatrix}
$$

and

$$
\begin{pmatrix} I & (1+\nu)J^TD^{-1} \\ 0 & \nu I \end{pmatrix} \begin{pmatrix} g - J^T\lambda \\ D(\lambda - \pi) \end{pmatrix} = \begin{pmatrix} g - (1+\nu)J^T\pi + \nu J^T\lambda \\ \nu D(\lambda - \pi) \end{pmatrix},
$$

upon comparison of (2.10), (3.2a), (3.5), and (3.6).

For the second part of the proof, assume that $\nabla M^{\mu,\nu}(x, \lambda)$ is zero, with $c_i(x) > 0$ and $\lambda_i > 0$ for all $i \in \mathcal{I}$. It follows that $D^\mu(x, \lambda)$ (2.11) is nonsingular. Hence, (3.2a) implies that $\lambda = \pi^\mu(x)$, where $\pi^\mu(x)$ is defined by (2.1). It is then straightforward to verify that $\Gamma^\mu(\pi^\mu(x)) = D^\mu(x, \pi^\mu(x))$ and $\Omega^\mu(x) = D^\mu(x, \pi^\mu(x))^{-1}$.   □

It is essential that $S^{\mu,\nu}$ is an exact approximation of $\nabla^2 M^{\mu,\nu}$ at every stationary point of $M^{\mu,\nu}(x, \lambda)$. Broadly speaking, it is not necessary to compute a direction of negative curvature for $M^{\mu,\nu}$ at every point at which $\nabla^2 M^{\mu,\nu}$ is indefinite, but a direction of negative curvature must always be computed in a neighborhood of a stationary point at which the Hessian is indefinite.

4. Calculation of the search direction. Based on the discussion of the previous section, we define the search direction using the primal-dual matrix K such that

$$
(4.1) \qquad\qquad K = \begin{pmatrix} H & J^T \\ J & -D \end{pmatrix},
$$

but use the matrix S such that

$$
(4.2) \qquad\qquad S = \begin{pmatrix} H + (1+\nu)J^TD^{-1}J & \nu J^T \\ \nu J & \nu D \end{pmatrix}
$$

to measure the approximate curvature of the merit function. In the modified Newton method of section 3.2 the search direction is given by

$$
\begin{pmatrix} \Delta x \\ \Delta \lambda \end{pmatrix} = \begin{pmatrix} s_x \\ s_\lambda \end{pmatrix} + \begin{pmatrix} d_x \\ d_\lambda \end{pmatrix} = s + d,
$$

where s is a descent direction and d is a direction of negative curvature for the merit function (see section 3.2). Whenever the Hessian approximation S of (4.2) is sufficiently positive definite, s is computed from the primal-dual equations (2.10), and d is zero since S is positive definite.

Note that the inertia of S is needed at each step, yet it is the matrix K (4.1) that is being factorized. This makes it essential that the inertia of S can be deduced from the inertia of K. The following lemma gives the required relationships.

LEMMA 4.1. *Let D be a symmetric positive-definite $m \times m$ matrix, let H be a symmetric $n \times n$ matrix, let ν be positive, let K be the primal-dual KKT matrix (4.1), and let S be the approximate Hessian (4.2). Then*

$$
In(S) = In(H + J^TD^{-1}J) + (m, 0, 0),
$$
$$
In(K) = In(H + J^TD^{-1}J) + (0, m, 0).
$$

Moreover, if $H + J^T D^{-1} J$ has at least one negative eigenvalue, then

$$\eta_{\min}(H + J^T D^{-1} J) \le \eta_{\min}(S).$$

Proof. Let X be a symmetric block-partitioned matrix of the form

$$X = \begin{pmatrix} A & B^T \\ B & C \end{pmatrix}.$$

The basic properties of the inertia give $In(X) = In(C) + In(A - B^T C^{-1} B)$; see, e.g., Cottle [9, p. 197]. Furthermore, if $A - B^T C^{-1} B$ has at least one negative eigenvalue and C is positive definite, then

$$\eta_{\min}(A - B^T C^{-1} B) \le \eta_{\min}(X)$$

(see Lemma 2.1 in Forsgren, Gill, and Murray [15]).

The inertia property implies that $In(S) = In(\nu D) + In(H + J^T D^{-1} J)$. Since νD is a positive diagonal, $In(\nu D) = (m, 0, 0)$ and the first result follows. Similarly, $In(K) = In(-D) + In(H + J^T D^{-1} J)$. Since $In(-D) = (0, m, 0)$, the second result also holds. Finally, the eigenvalue relation follows from the above relation with $X = S$ and $C = \nu D$.   □

4.1. The factorization. We assume that the primal-dual KKT equations are solved using a variant of the *symmetric indefinite factorization* (see Bunch and Parlett [6], Fletcher [14], and Bunch and Kaufman [4]), which we refer to as the LBL^T *factorization*. If K denotes the particular primal-dual KKT matrix under consideration, then the LBL^T factorization defines a permutation P, a block-diagonal B, and a unit-lower-triangular L such that

(4.3) $P^T K P = L B L^T, \quad \text{where} \quad B = \text{diag}(B_1, B_2, \ldots, B_s).$

Each B_j is either one-by-one, or is two-by-two and nonsingular. The permutation P incorporates certain symmetric interchanges that are needed to preserve numerical stability.

At the start of the ℓth stage ($\ell \ge 1$) of the LBL^T factorization of K we have

(4.4) $K^{(\ell)} = \left(K_{ij}^{(\ell)} \right) = P_\ell^T K P_\ell - \sum_{i=1}^{\ell-1} L_i B_i L_i^T = \begin{pmatrix} 0 & 0 \\ 0 & K_\ell \end{pmatrix},$

where P_ℓ is a permutation and L_i is either an $n \times 1$ or $n \times 2$ matrix consisting of the column(s) of L computed at the ith stage. The matrix K_ℓ is called the *Schur complement* and represents the part of K remaining to be factorized. At the ℓth stage, a one-by-one or symmetric two-by-two submatrix of K_ℓ is selected as the next pivot B_ℓ. The pivot rows and columns are brought to the leading position of K_ℓ using a symmetric interchange, which must be applied to $P_\ell^T K P_\ell$ and the rows of each L_i. The one-by-one or two-by-two pivot is then used to eliminate one or two rows and columns from the permuted Schur complement. The ℓth stage is completed by the computation of L_ℓ from B_ℓ^{-1} and the pivot columns of $K^{(\ell)}$ (for further details, see Bunch [3]).

Clearly, the permutation P depends on the sequence of pivots. Various pivoting strategies have been proposed that define a numerically stable factorization (see, e.g.,

Bunch, Kaufman, and Parlett [5]; Higham [26]). Here we consider pivot strategies that additionally ensure that B has a specific inertia.

If K is the primal-dual matrix (4.1), potential pivots in the Schur complement can be labeled to reflect the position of the pivot in the original matrix. For example, a diagonal of the Schur complement will be a "D pivot" if it is in the position occupied by an element of $-D$ in the original matrix. Similarly, a two-by-two pivot with diagonal and off-diagonal positions occupied by elements from H and D will be an "HD pivot." With this labeling, all one-by-one pivots are either H or D pivots, and two-by-two pivots are HH, DD, or HD (DH) pivots.

The key to the calculation of suitable search directions is the use of a sequence of *inertia-controlling pivots*. The resulting factorization generalizes the symmetric indefinite factorization proposed by Forsgren and Murray [17] for linear equality-constrained minimization.

The inertia-controlling factorization consists of two phases. The first phase continues until all rows and columns associated with the D block of K have been eliminated (i.e., until no D pivots or HD pivots remain in the Schur complement). During this first phase, only certain types of pivots are allowed (see below). When the first phase is complete, the pivot choice is restricted only by numerical considerations.

In order to describe the first phase it is necessary to further distinguish between pivot types. Let a superscript "$+$" or "$-$" indicate the sign of an eigenvalue of a potential pivot. For example, H^+ denotes a positive one-by-one H pivot; D^- denotes a negative one-by-one D pivot; and HD^{-+} denotes an HD pivot with one positive and one negative eigenvalue. The main feature of the first phase of the factorization is that only H^+, D^-, HH^{++}, DD^{--}, and HD^{-+} pivots are allowed.

Consider the partition of K given by

$$
\begin{pmatrix}
H_{11} & H_{12} & J_1^T \\
H_{21} & H_{22} & J_2^T \\
J_1 & J_2 & -D
\end{pmatrix},
$$

and assume that all pivots from the first phase come from rows and columns associated with H_{11}, J_1, and $-D$. Without loss of generality, we assume the permutation P to be such that

$$
P^T K P =
\begin{pmatrix}
H_{11} & J_1^T & H_{21} \\
J_1 & -D & J_2 \\
H_{21} & J_2^T & H_{22}
\end{pmatrix},
$$

i.e., the order of the pivots *within* the first phase is ignored. The matrix $P^T K P$ is partitioned as

$$
P^T K P =
\begin{pmatrix}
K_{11} & K_{12} \\
K_{21} & K_{22}
\end{pmatrix},
$$

where

$$
K_{11} =
\begin{pmatrix}
H_{11} & J_1^T \\
J_1 & -D
\end{pmatrix}, \qquad
K_{21} = K_{12}^T = \begin{pmatrix} H_{21} & J_2^T \end{pmatrix}, \quad \text{and} \quad K_{22} = H_{22}.
$$

The sequence of pivots used during the first phase fixes K_{11} as an $n_1 \times m$ matrix for some n_1 ($0 \le n_1 \le n$) and thereby identifies the blocks K_{11}, K_{12}, and K_{22} (the value

of n_1 will depend upon the relative number of one-by-one and two-by-two pivots). The restricted pivot choice ensures that $In\,(K_{11}) = (n_1, m, 0)$. Moreover, on completion of the first phase, factors L_{11} and B_{11} have been computed such that

$$K_{11} = L_{11}B_{11}L_{11}^T \quad \text{and} \quad L_{21} = K_{21}L_{11}^{-T}B_{11}^{-1}.$$

During the second phase, the pivot choice is unrestricted, which implies that the remaining Schur complement $K_{22} - K_{21}K_{11}^{-1}K_{12}$ can be factored using a standard LBL^T-factorization. The only restriction is that the choice of pivots gives a factorization for which $\|L\|$ is bounded.

On completion of the second phase we have

$$K_{22} - K_{21}K_{11}^{-1}K_{12} = L_{22}B_{22}L_{22}^T.$$

This provides the complete factorization

$$\begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} B_{11} & 0 \\ 0 & B_{22} \end{pmatrix} \begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22}^T \end{pmatrix}.$$

4.2. Calculation of the descent direction. Given the inertia-controlling LBL^T-factorization of K, the descent direction is now computed from

$$(4.5) \qquad \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} B_{11} & 0 \\ 0 & \bar{B}_{22} \end{pmatrix} \begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22}^T \end{pmatrix} \widetilde{s} = -P^T \begin{pmatrix} g - J^T\lambda \\ D(\lambda - \pi) \end{pmatrix},$$

with

$$\begin{pmatrix} s_x \\ -s_\lambda \end{pmatrix} = P\widetilde{s},$$

where $\bar{B}_{22}$ is a sufficiently positive-definite modification of B_{22}. In particular, if B_{22} is sufficiently positive definite, then $\bar{B}_{22} = B_{22}$. It is relatively straightforward to modify B_{22} since the diagonal blocks are of size one-by-one or two-by-two and the eigenvalues of each block can be found explicitly and modified if necessary. A crucial feature of the factorization is that K_{22} only includes rows and columns from H. This ensures that any modification of B_{22} affects only parts of H in K. Furthermore, (4.5) can be written as

$$(4.6) \qquad \begin{pmatrix} \bar{H} & J^T \\ J & -D \end{pmatrix} \begin{pmatrix} s_x \\ -s_\lambda \end{pmatrix} = -\begin{pmatrix} g - J^T\lambda \\ D(\lambda - \pi) \end{pmatrix},$$

where $\bar{H}$ is a symmetric modification of H such that $\bar{H} + J^TD^{-1}J$ is sufficiently positive definite. Whenever $H + J^TD^{-1}J$ is sufficiently positive definite, no modification takes place and the descent direction s is the primal-dual direction computed from (2.10). This is analogous to the method of Forsgren and Murray [17].

Alternatively, we may use Theorem 3.2 to establish that (4.6) is equivalent to

$$\bar{S}s = -\nabla M^{\mu,\nu}(x, \lambda),$$

where

$$(4.7) \qquad \bar{S} = \begin{pmatrix} \bar{H} + (1 + \nu)J^TD^{-1}J & \nu J^T \\ \nu J & \nu D \end{pmatrix}.$$

We first establish descent properties of s with respect to the merit function (3.1).

LEMMA 4.2. *Let s satisfy (4.6), where $\bar{H}$ is symmetric, D is symmetric and positive definite, and $\bar{H} + J^T D^{-1} J$ is positive definite. Then*

$$s^T \nabla M^{\mu,\nu}(x, \lambda) = -s_x^T(\bar{H} + J^T D^{-1} J)s_x - \nu(\lambda - \pi)^T D(\lambda - \pi),$$

where $\nabla M^{\mu,\nu}(x, \lambda)$ is defined by (3.2a). Moreover, $s_x = 0$ if and only if $g - J^T \pi = 0$. Finally, if $s_x = 0$, then $s_\lambda = \pi - \lambda$.

Proof. Straightforward substitution gives the first result. Elimination of s_λ in (4.6) gives

$$(\bar{H} + J^T D^{-1} J)s_x = -(g - J^T \pi).$$

Hence, since $\bar{H} + J^T D^{-1} J$ is positive definite, we have $s_x = 0$ if and only if $g - J^T \pi = 0$. Finally, if $s_x = 0$, then $g = J^T \pi$, and s_λ is given by

$$\begin{pmatrix} -J^T \\ D \end{pmatrix} s_\lambda = -\begin{pmatrix} g - J^T \lambda \\ D(\lambda - \pi) \end{pmatrix} = \begin{pmatrix} -J^T \\ D \end{pmatrix}(\pi - \lambda),$$

with unique solution $s_\lambda = \pi - \lambda$.   □

The positive definiteness of $\bar{S}$ ensures that $s^T \nabla M^{\mu,\nu}(x, \lambda) < 0$ unless $s = 0$ and $\nabla M^{\mu,\nu}(x, \lambda) = 0$. The additional information provided by Lemma 4.2 is that if $s_x = 0$, then $g - J^T \pi = 0$ and $s_\lambda = \pi - \lambda$. This means that if $x = x(\mu)$, then one further Newton step gives the λ variables as $\lambda(\mu)$.

Some further assumptions are required in order to make s a sufficient descent direction in the sense of (3.4a). This is summarized in the following theorem.

THEOREM 4.3. *Assume that the sequence of directions $\{s_k\}_{k=0}^\infty$ satisfies*

$$\begin{pmatrix} \bar{H}_k & J_k^T \\ J_k & -D_k \end{pmatrix} \begin{pmatrix} s_{x,k} \\ -s_{\lambda,k} \end{pmatrix} = -\begin{pmatrix} g_k - J_k^T \lambda_k \\ D_k(\lambda_k - \pi_k) \end{pmatrix},$$

where $\bar{H}_k + J_k^T D_k^{-1} J_k \succ 0$ and $D_k \succ 0$ for all k. Then, $s_k^T \nabla M^{\mu,\nu}(x_k, \lambda_k) < 0$ unless $s_k = 0$. Furthermore, assume that $\limsup_{k\to\infty} \|\bar{H}_k\| < \infty$, $\limsup_{k\to\infty} \|J_k\| < \infty$, $\liminf_{k\to\infty} \eta_{\min}(D_k) > 0$, $\limsup_{k\to\infty} \eta_{\max}(D_k) < \infty$, and $\liminf_{k\to\infty} \eta_{\min}(\bar{H}_k + J_k^T D_k^{-1} J_k) > 0$. Then, if

$$\lim_{k\to\infty} s_k^T \nabla M^{\mu,\nu}(x_k, \lambda_k) = 0,$$

it holds that

$$\lim_{k\to\infty} s_k = 0 \quad and \quad \lim_{k\to\infty} \nabla M^{\mu,\nu}(x_k, \lambda_k) = 0.$$

Proof. Theorem 3.2 gives $\bar{S}_k s_k = -\nabla M^{\mu,\nu}(x_k, \lambda_k)$, where

$$\bar{S}_k = \begin{pmatrix} \bar{H}_k + (1+\nu)J_k^T D_k^{-1} J_k & \nu J_k^T \\ \nu J_k & \nu D_k \end{pmatrix}.$$

If $\bar{H}_k + J_k^T D_k^{-1} J_k \succ 0$, $D_k \succ 0$, and $\nu > 0$, Lemma 4.1 shows that $\bar{S}_k \succ 0$. Hence, the identity $s_k^T \nabla M^{\mu,\nu}(x_k, \lambda_k) = -s_k^T \bar{S}_k s_k$ ensures that $s_k^T \nabla M^{\mu,\nu}(x_k, \lambda_k) < 0$ unless $s_k = 0$.

To prove the second part of the theorem, $\bar{S}_k$ is written as

$$(4.8) \qquad \bar{S}_k = \begin{pmatrix} I & J_k^T D_k^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \bar{H}_k + J_k^T D_k^{-1} J_k & 0 \\ 0 & \nu D_k \end{pmatrix} \begin{pmatrix} I & 0 \\ D_k^{-1} J_k & I \end{pmatrix}.$$

It follows from (4.8) and the boundedness assumptions that $\liminf_{k\to\infty} \eta_{\min}(\bar{S}_k) > 0$ and $\limsup_{k\to\infty} \eta_{\max}(\bar{S}_k) < \infty$ (see Lemma 3.1 in Forsgren and Murray [17]). The proof is completed by combining these results with the modified Newton equation $\bar{S}_k s_k = -\nabla M^{\mu,\nu}(x_k, \lambda_k)$. □

4.3. Calculation of the direction of negative curvature. Given the inertia-controlling LBL^T-factorization of K, if B_{22} has at least one negative eigenvalue, the direction of negative curvature is computed from

$$(4.9) \qquad \begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22}^T \end{pmatrix} \tilde{d} = \pm\sigma \begin{pmatrix} 0 \\ u \end{pmatrix},$$

with

$$\begin{pmatrix} d_x \\ -d_\lambda \end{pmatrix} = P\tilde{d},$$

where $\sigma = \sqrt{-\eta_{\min}(B_{22})}$ and u is an associated eigenvector of unit length. (Note that σ and u are easily obtained since B_{22} is symmetric block-diagonal with diagonal blocks of dimension one-by-one or two-by-two.) The sign of d is chosen such that $d^T \nabla M^{\mu,\nu}(x, \lambda) \le 0$. Since $K_{22} = H_{22}$ it follows from (4.9) that $Jd_x + Dd_\lambda = 0$. Also, from the definition of d it follows that $d^T S d = -\eta_{\min}(B_{22})^2$.

The following lemma establishes the properties of the curvature along a vector d such that $Jd_x + Dd_\lambda = 0$.

LEMMA 4.4. *If $Jd_x + Dd_\lambda = 0$ and D is symmetric and nonsingular, then*

$$d_x^T(H + J^T D^{-1} J)d_x = \begin{pmatrix} d_x^T & -d_\lambda^T \end{pmatrix} \begin{pmatrix} H & J^T \\ J & -D \end{pmatrix} \begin{pmatrix} d_x \\ -d_\lambda \end{pmatrix}$$

$$= \begin{pmatrix} d_x^T & d_\lambda^T \end{pmatrix} \begin{pmatrix} H + (1+\nu)J^T D^{-1} J & \nu J^T \\ \nu J & \nu D \end{pmatrix} \begin{pmatrix} d_x \\ d_\lambda \end{pmatrix}.$$

Proof. Straightforward substitution using $Jd_x + Dd_\lambda = 0$ gives the result. □

Finally, we give a sufficient curvature result for a sequence of iterates.

THEOREM 4.5. *Assume that each member of the sequence $\{d_k\}_{k=0}^{\infty}$ is computed using the system (4.9). Assume that this system is associated with the LBL^T-factors of a sequence of B_k-matrices such that for each k, B_k has more than m negative eigenvalues. Furthermore, assume that the factorization is inertia controlling, with $\limsup_{k\to\infty} \|L_k\| < \infty$. Then, if*

$$\lim_{k\to\infty} d_k^T S_k d_k = 0,$$

it holds that

$$\lim_{k\to\infty} d_k = 0 \quad and \quad \lim_{k\to\infty} \eta_{\min}(S_k) = 0.$$

Proof. It follows from (4.9) that $d_k^T S_k d_k = -\eta_{\min}(B_{22,k})^2$. Hence, if d_k satisfies $\lim_{k\to\infty} d_k^T S_k d_k = 0$, then it must be the case that $\lim_{k\to\infty} \eta_{\min}(B_{22,k}) = 0$. Now, the definition (4.9) of d_k and the assumption that $\limsup_{k\to\infty} \|L_k\| < \infty$ ensure that $\lim_{k\to\infty} d_k = 0$. Since $In(K_{11,k}) = (n_{1,k}, m, 0)$, we may view $K_{22,k} - K_{21,k} K_{11,k}^{-1} K_{12,k}$ as being formed by first eliminating $-D_k$, giving the m negative eigenvalues. The Schur complement obtained this way is $H_k + J_k^T D_k^{-1} J_k$. Hence, elimination of the remaining $n_{1,k} - m$ pivots of $K_{11,k}$ must have been done by positive-definite pivots. Since K_k has more than m negative eigenvalues, we must have $\eta_{\min}(H_k + J_k^T D_k^{-1} J_k) < 0$. We conclude that

$$(4.10) \qquad \eta_{\min}(K_{22,k} - K_{21,k} K_{11,k}^{-1} K_{12,k}) \leq \eta_{\min}(H_k + J_k^T D_k^{-1} J_k) < 0;$$

see Lemma 2.1 in Forsgren, Gill, and Murray [15]. By assumption, the factorization guarantees that $\limsup_{k\to\infty} \|L_k\| < \infty$. Since L_k is unit-lower-triangular, this implies that $\limsup_{k\to\infty} \|L_{22,k}\| < \infty$ and $\limsup_{k\to\infty} \|L_{22,k}^{-1}\| < \infty$. Hence, if $\lim_{k\to\infty} \eta_{\min}(B_{22,k}) = 0$, we must have

$$(4.11) \qquad \lim_{k\to\infty} \eta_{\min}(K_{22,k} - K_{21,k} K_{11,k}^{-1} K_{12,k}) = 0;$$

see Lemma 3.1 in Forsgren and Murray [17]. A combination of (4.10), (4.11), and Lemma 4.1 gives $\lim_{k\to\infty} \eta_{\min}(S_k) = 0$, as required. $\square$

4.4. Properties of the search direction. The sufficiency properties of Theorems 4.3 and 4.5 for the descent direction and direction of negative curvature require L and D to be bounded and D to be bounded away from a singular matrix. In addition, we have to assume that the iterates remain in a compact set for convergence to be assured. In theory, this can be handled by adding lower and upper bounds on the variables since the algorithm stays feasible with respect to all inequalities. However, this is not completely satisfactory. In this paper, the discussion on convergence is limited to the results of the previous section. A formal convergence proof will be considered in a future paper.

At any iteration, an alternative to performing a line search with respect to x and λ is to reset λ to $\pi^\mu(x)$ and perform the line search with respect to x only. In this case, the directions defined by the primal-dual equations (2.8) and the primal equations (2.9) are identical. This strategy allows the use of a convergence proof for the pure primal penalty-barrier method. Unfortunately, we know of no published proof for the nonconvex case that gives limit points that satisfy the second-order necessary optimality conditions. For line-search barrier-penalty-type methods, Benchakroun, Dussault, and Mansouri [2], Gould [23], and Dussault [11] discuss convergence to a first-order point. In a trust-region setting, Byrd, Gilbert, and Nocedal [7] propose a method that converges to a first-order point.

The properties required of D for the analysis are typical. For an equality constraint, $d_i^\mu(x, \lambda)$ is fixed at μ, which is trivially bounded away from zero for μ *fixed*. However, this lower bound obviously deteriorates as μ gets small. For inequality constraints, $d_i^\mu(x, \lambda)$ is $c_i(x)/\lambda$, and on the trajectory we have $d_i^\mu(x(\mu), \lambda(\mu)) = c_i(x(\mu))^2/\mu$. Again, for a *fixed* μ, we expect $d_i^\mu(x(\mu), \lambda(\mu))$ to be both bounded away from zero and bounded from above. However, further research is required to show this rigorously and to incorporate a suitable scheme for changing μ.

5. Discussion and further research. We have demonstrated how the primal-dual search directions (2.8) can be interpreted as modified Newton directions for minimizing an augmented penalty-barrier merit function (3.1). The main benefits of this

approach are twofold: (i) the equations are well conditioned in the limit when strict complementarity holds; and (ii) the merit function is minimized with respect to both x and λ. The inertia-controlling factorization provides descent directions for $M^{\mu,\nu}(x,\lambda)$ and directions of negative curvature for the approximate Hessian $S^{\mu,\nu}(x,\lambda)$. We observe that the ill-conditioning has been removed from the equations determining the search direction and is now only present in the merit function.

If the proposed method is to work well, suitable termination criteria must be determined for the outer iterations. The convergence analysis must be extended to an algorithm that includes a strategy for reducing μ. In the limit, this algorithm should require only one inner iteration for each outer iteration. However, although preliminary numerical results have been encouraging, it is not known at this stage whether or not the merit function can induce the "Maratos effect," whereby the merit function does not accept the unit step as μ is decreased.

Finally, we note that the proposed method is applicable to sparse problems. The key is the calculation of an efficient sparse inertia-controlling factorization of K (4.1). Up to now, our experiments have been limited to the use of a dense factorization. (Implementations of other sparse inertia-controlling factorizations are discussed by Arioli et al. [1] and Gould [24, 25] in the context of linear equality constrained optimization.)

Other interesting variations of the proposed method remain to be investigated. For example, if (2.5b) is replaced by $\mu + c_i(x)/\lambda_i$ for $i \in \mathcal{E}$, then the Jacobian of the resulting system is well conditioned in the limit if strict complementarity holds. An advantage of this formulation is that the Jacobian is independent of μ. However, we have not yet been able to associate these equations with a suitable merit function. Another possibility is to use different values of μ and ν for each constraint. A zero value of μ can be assigned to any linear equality constraint once it is satisfied. This would give a method equivalent to that of Forsgren and Murray [17]. If only equality constraints are present, a zero value of μ gives the standard sequential quadratic programming search direction. However, if a constraint is not satisfied exactly, a zero value of μ is not treated by the analysis of this paper. The introduction of shifts on the constraints may be helpful in this case (see Powell [39]).

REFERENCES

[1] M. ARIOLI, T. F. CHAN, I. S. DUFF, N. I. M. GOULD, AND J. K. REID, *Computing a Search Direction for Large-Scale Linearly Constrained Nonlinear Optimization Calculations*, Tech. Rep. RAL-93-066, Rutherford Appleton Laboratory, Chilton, England, 1993.

[2] A. BENCHAKROUN, J. DUSSAULT, AND A. MANSOURI, *A two parameter mixed interior-exterior penalty algorithm*, ZOR—Math. Methods Oper. Res., 41 (1995), pp. 25–55.

[3] J. R. BUNCH, *Partial pivoting strategies for symmetric matrices*, SIAM J. Numer. Anal., 11 (1974), pp. 521–528.

[4] J. R. BUNCH AND L. KAUFMAN, *Some stable methods for calculating inertia and solving symmetric linear systems*, Math. Comp., 31 (1977), pp. 163–179.

[5] J. R. BUNCH, L. KAUFMAN, AND B. N. PARLETT, *Decomposition of a symmetric matrix*, Numer. Math., 27 (1976), pp. 95–109.

[6] J. R. BUNCH AND B. N. PARLETT, *Direct methods for solving symmetric indefinite systems of linear equations*, SIAM J. Numer. Anal., 8 (1971), pp. 639–655.

[7] R. H. Byrd, J. C. Gilbert, and J. Nocedal, *A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming*, Tech. Rep. OTC-96/2, Optimization Technology Center, Northwestern University, Evanston, IL, 1996.

[8] 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.

[9] R. W. Cottle, *Manifestations of the Schur complement*, Linear Algebra Appl., 8 (1974), pp. 189–211.

[10] J. E. Dennis, Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice–Hall, Englewood Cliffs, NJ, 1983.

[11] J.-P. Dussault, *Numerical stability and efficiency of penalty algorithms*, SIAM J. Numer. Anal., 32 (1995), pp. 296–317.

[12] A. S. El-Bakry, R. A. Tapia, Y. Zhang, and T. Tsuchiya, *On the Formulation and Theory of the Newton Interior Point Method for Nonlinear Programming*, Tech. Report 92-40, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1992. Revised April 1995.

[13] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, New York, London, Sydney, Toronto, 1968.

[14] R. Fletcher, *Factorizing symmetric indefinite matrices*, Linear Algebra Appl., 14 (1976), pp. 257–272.

[15] 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.

[16] 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.

[17] A. Forsgren and W. Murray, *Newton methods for large-scale linear equality-constrained minimization*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 560–587.

[18] P. E. Gill, W. Murray, D. B. Ponceleón, and M. A. Saunders, *Primal-dual methods for linear programming*, Math. Programming, 70 (1995), pp. 251–277.

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

[20] D. Goldfarb, *Curvilinear path steplength algorithms for minimization which use directions of negative curvature*, Math. Programming, 18 (1980), pp. 31–40.

[21] C. C. Gonzaga, *Path-following methods for linear programming*, SIAM Rev., 34 (1992), pp. 167–224.

[22] 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.

[23] 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.

[24] N. I. M. Gould, *Constructing Appropriate Models for Large-Scale, Linearly Constrained, Nonconvex, Nonlinear Optimization Algorithms*, Tech. Rep. RAL-TR-95-037, Rutherford Appleton Laboratory, Chilton, England, 1995.

[25] N. I. M. Gould, *Constructing appropriate models for large-scale, linearly-constrained, nonconvex, nonlinear optimization algorithms*, SIAM J. Optim., submitted.

[26] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, PA, 1996.

[27] S. Kaniel and A. Dax, *A modified Newton's method for unconstrained minimization*, SIAM J. Numer. Anal., 16 (1979), pp. 324–331.

[28] M. Kojima, S. Mizuno, and A. Yoshise, *A primal-dual interior point algorithm for linear programming*, in Progress in Mathematical Programming: Interior Point and Related Methods, N. Megiddo, ed., Springer-Verlag, New York, 1989, pp. 29–47.

[29] I. J. Lustig, R. E. Marsten, and D. F. Shanno, *Computational experience with a primal–dual interior point method for linear programming*, Linear Algebra Appl., 152 (1991), pp. 191–222.

[30] G. P. McCormick, *A modification of Armijo's step-size rule for negative curvature*, Math. Programming, 13 (1977), pp. 111–115.

[31] 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.

[32] S. Mehrotra, *On the implementation of a primal-dual interior point method*, SIAM J. Optim., 2 (1992), pp. 575–601.

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

[34] J. J. Moré and D. C. Sorensen, *On the use of directions of negative curvature in a modified Newton method*, Math. Programming, 16 (1979), pp. 1–20.

[35] H. Mukai and E. Polak, *A second-order method for unconstrained optimization*, J. Optim. Theory Appl., 26 (1978), pp. 501–513.

[36] 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.

[37] W. Murray and M. H. Wright, *Line search procedures for the logarithmic barrier function*, SIAM J. Optim., 4 (1994), pp. 229–246.

[38] D. B. Ponceleón, *Barrier Methods for Large-Scale Quadratic Programming*, Ph.D. thesis, Report SOL 91-2, Department of Operations Research, Stanford University, Stanford, CA, 1991.

[39] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298.

[40] K. Svanberg, *Method of moving asymptotes—a new method for structural optimization*, Internat. J. Numer. Methods Engrg., 24 (1987), pp. 359–373.

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

[42] M. H. Wright, *Why a pure primal Newton barrier step may be infeasible*, SIAM J. Optim., 5 (1995), pp. 1–12.

[43] S. J. Wright, *Stability of linear equations solvers in interior-point methods*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1287–1307.

[44] S. Wright, *Stability of augmented system factorizations in interior-point methods*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 191–222.

[45] Y. Ye, *Line Searches in Potential Reduction Algorithm for Linear Programming*. manuscript, 1989.