UNIVERSITY OF CALIFORNIA, SAN DIEGO

Primal-Dual Interior Methods for Quadratic Programming

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Mathematics with a Specialization in Computational Science

by

Anna Shustrova

Committee in charge:

Professor Philip E. Gill, Chair
Professor Henry D. I. Abarbanel
Professor Randolph E. Bank
Professor Michael Holst
Professor Alison L. Marsden

2015

The dissertation of Anna Shustrova is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Prof. Philip Gill for his patience and support. The last months before the defense were very intense and I am very grateful for all the extra time and help on his part. His guidance and understanding made it all possible.

I also owe a great many thanks to Profs. Henry D. I. Abarbanel, Randolph E. Bank, Michael Holst and Alison L. Marsden for taking the time to be on my committee and being exceptionally accommodating with the scheduling of the defense.

I would also like to thank my family. Their encouragement and support kept me afloat.

A special thanks goes to my master's thesis advisor, Dr. Subramanian, who took me under his wing and set me on the path that led to this Ph.D.

VITA

2003	B. A. in Mathematics, University of California, Berkeley
2005–2006	Teaching Associate, California State University, Los Angeles
2006	M. S. in Mathematics, California State University, Los Angeles
2006–2011	Graduate Teaching Assistant, University of California, San Diego
2010	C. Phil. in Mathematics with a Specialization in Computational Science, University of California, San Diego
2015	Ph. D. in Mathematics with a Specialization in Computational Science

ABSTRACT OF THE DISSERTATION

Primal-Dual Interior Methods for Quadratic Programming

by

Anna Shustrova

Doctor of Philosophy in Mathematics with a Specialization in Computational Science

University of California, San Diego, 2015

Professor Philip E. Gill, Chair

Interior methods are a class of computational methods for solving a constrained optimization problem. Interior methods follow a continuous path to the solution that passes through the interior of the feasible region (i.e., the set of points that satisfy the constraints). Interior-point methods may also be viewed as methods that replace the constrained problem by a sequence of unconstrained problems in which the objective function is augmented by a weighted "barrier" term that is infinite at the boundary of the feasible region. Convergence to a solution of the constrained problem is achieved by solving a sequence of unconstrained problems in which the weight on the barrier term is steadily reduced to zero.

This thesis concerns the formulation and analysis of interior methods for

the solution of a quadratic programming (QP) problem, which is an optimization problem with a quadratic objective function and linear constraints. The linear constraints may include an arbitrary mixture of equality and inequality constraints, where the inequality constraints may be subject to lower and/or upper bounds. QP problems arise in a wide variety of applications. An important application is in sequential quadratic programming methods for nonlinear optimization, which involve minimizing a sequence of QP subproblems based on a quadratic approximation of the nonlinear objective function and a set of linearized nonlinear constraints.

Two new interior methods for QP are proposed. Each is based on the properties of a barrier function defined in terms of both the primal and dual variables. The first method is suitable for a QP with all inequality constraints. At each iteration, the Newton equations for minimizing a quadratic model of the primal-dual barrier function are reformulated in terms of a symmetric indefinite system of equations that is solved using an inertia controlling factorization. This factorization provides an effective method for the detection and convexification of nonconvex problems. The second method is intended for problems with a mixture of equality and inequality constraints. In this case, the QP constraints are converted to so-called standard form and a primal-dual augmented Lagrangian is used to ensure the feasibility of the equality constraints in the limit.

1 Introduction

1.1 Overview

Quadratic programming (QP) involves minimizing a quadratic objective function subject to linear constraints on the variables. The most general quadratic program is given by

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & q(x) = \tfrac{1}{2}x^T H x + c^T x \\ \text{subject to} \quad & \ell \leq \begin{pmatrix} Ax - b \\ x \end{pmatrix} \leq u, \end{aligned} \tag{1.1}$$

where $q(x)$ is the quadratic objective function, $c \in \mathbb{R}^n$, H is an $n \times n$ symmetric matrix, A is an $m \times n$ constraint matrix, and $\ell \leq u \in \mathbb{R}^{m+n}$. An equality constraint or fixed variable may be defined by setting $\ell_i = u_i$.

By introducing slack variables $s \in \mathbb{R}^m$, the general problem (1.1) can be written in the equivalent form

$$\begin{aligned} \underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \quad & q(x) \\ \text{subject to} \quad & Ax - s = b \\ & \ell \leq \begin{pmatrix} s \\ x \end{pmatrix} \leq u. \end{aligned} \tag{1.2}$$

An additional benefit of adding slack variables to the inequalities is that the constraint matrix $\begin{pmatrix} A & I_m \end{pmatrix}$ has full row rank. The assumption of full rank is required in many optimization methods. To simplify exposition, however, we will not explicitly include slack variables, but instead make the assumption that the constraint matrix has full row rank when required. Therefore, without loss of generality we

can write (1.2) as

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & q(x) \\ \text{subject to} & Ax = b, \qquad \ell \leq x \leq u. \end{array} \tag{1.3}$$

Although methods can be extended to both upper and lower bounds, to simplify exposition in the theoretical part of the thesis, only nonnegativity bounds on x are considered. In this case, the QP is said to be in standard form, and is written as

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & q(x) \\ \text{subject to} & Ax = b, \qquad x \geq 0. \end{array} \tag{1.4}$$

For some of the methods discussed in this thesis it is convenient to consider quadratic problems in so-called all-inequality form. To achieve this formulation one could for instance rewrite the equality constraints as

$$Ax \geq b \quad \text{and} \quad Ax \geq -b.$$

Thus, without loss of generality, we can assume that the QP in all-inequality form is written as

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & q(x) \\ \text{subject to} & Ax \geq b. \end{array} \tag{1.5}$$

The difficulty of solving a quadratic program largely depends on the convexity of the objective function q. If the Hessian H of the objective function is positive semidefinite, then the QP is said to be convex. In this case, a local solution of the QP is also a global solution. If H is negative definite or indefinite, the QP is said to be nonconvex. Nonconvex quadratic programs, are known to be NP-hard—even for the calculation of a local minimizer. Additionally, a nonconvex QP may be unbounded below.

There are many methods for finding local minimizers of a QP, but most fall into two categories: active-set methods and interior-point methods. Each has its advantages and disadvantages. Active-set methods are iterative methods that solve a sequence of equality-constrained quadratic subproblems. These constraints correspond to the current prediction of the optimal active set, which is the set of constraints that are satisfied with equality at the optimal point. In contrast,

interior methods solve the QP by converting it into a parameterized sequence of unconstrained problems whose solutions form a differentiable path that passes through a solution of the QP. In the simplest case, the path is parameterized by a positive scalar parameter μ, such that as $\mu \to 0$, a point $x(\mu)$ on the path converges to the solution of the QP. In general, interior methods require fewer iterations, but each iteration is more expensive because it requires finding the solution of a linear systems involving all the variables and constraints, whereas active-set methods some smaller systems involving only a subset of the variables and constraints.

Our interest in quadratic programming methods is primarily motivated by their usefulness in solving nonlinear optimization problems in the context of sequential quadratic programming (SQP) methods. SQP methods solve a sequence of quadratic programming subproblems in which a quadratic model of the objective function is minimized subject to a linearization of the nonlinear constraints. Active-set methods are often preferred for solving these subproblems because of their ability to take advantage of good starting points, making them particularly efficient in solving a sequence of related problems. Interior-point methods, on the other hand, are generally considered to be less effective for solving sequences of related problems because a good starting point may lie far from the path $x(\mu)$. In general, when close to an optimal point, a step towards the path is not well-defined, and it may take many iterations to move onto the path.

1.2 Contributions of this Thesis

This thesis considers a new interior-point method for solving quadratic programs. The method is based on existing barrier methods and incorporates features of both primal-dual barrier methods and modified barrier methods. In particular it inherits the effectiveness of primal-dual methods at following the barrier trajectory and, like modified barrier methods, eliminates the need for the barrier parameter to go to zero to force convergence.

Chapter 2 provides a review of some basic tools and ideas often used in interior methods, including unconstrained minimization, necessary and sufficient

optimality conditions for quadratic programs, and a brief introduction to penalty functions. We also discuss matrix factorization, specifically inertia controlling symmetric indefinite factorization, and its use in solving primal-dual KKT systems that arise within interior methods. An overview of existing barrier methods, including a more detailed look at classical and modified barrier functions as well as primal-dual methods is presented in Chapter 3. Chapter 4 introduces a new primal-dual modified barrier method for solving QPs in all-inequality form (1.5) starting from a given feasible point. Chapter 5 focuses of solving QPs in standard form (1.4). Here we take advantage of existing primal-dual augmented lagrangian methods to treat equality constraints and combine them with the barrier method proposed in Chapter 4 to deal with the nonnegativity bounds on x. In Chapter 6, we briefly discuss how to extend the method of Chapter 5 to include upper and lower bounds on the variables and provide numerical results for the proposed algorithm.

1.3 Notation

- $q(x)$, quadratic objective function evaluated at x.
- $g(x) = c + Hx$, the gradient of the quadratic objective evaluated at x.
- (x_k, z_k), approximate primal-dual solution to the QP in all-inequality form, where z_k is a vector of dual variables for the inequality constraints.
- (x_k, y_k, z_k), approximate primal-dual solution to the QP in standard form at step k, where y_k is a vector of dual variables for the equality constraints and z_k is a vector of dual variables for the nonnegativity constraints.
- $q_k = q(x_k)$, $g_k = g(x_k)$.
- $e = (1, \ldots, 1)^T$, the column vector all ones whose size depends on the context.
- I, identity matrix.
- e_j, vector representing the jth column of I.

- $X = \text{diag}(x_1, \ldots, x_n)$, diagonal matrix with the components of the vector x on the diagonal.
- If $x, y \in \mathbb{R}^n$, $x \cdot y$ and $x \cdot\!/ y$ are vectors in $\mathbb{R}^n$ such that $[x \cdot y]_i = x_i y_i$ and $[x \cdot\!/ y]_i = x_i / y_i$.
- For a symmetric matrix B, $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ represent the smallest and the largest eigenvalues of B, respectively.
- Unless otherwise specified, $\|\cdot\|$ denotes $\|\cdot\|_2$.
- $B(x, \delta) = y \in \mathbb{R}^n : \|x - y\| < \delta$ denotes an open ball around x with radius δ.
- $S(x, r) = y \in \mathbb{R}^n : \|x - y\| = r$ denotes a sphere or radius r centered at x.

1.4 Definitions

Definition 1.4.1 (Unconstrained (Strict) Local Minimizer)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$. Then x^* is a local minimizer of f if there exists a $\delta > 0$ such that $f(x) \geq f(x^*)$ for all $x \in B(x^*, \delta)$. If the inequality is strict, then it is said to be a strict minimizer.*

Definition 1.4.2 (Feasible Region)**.** *Feasible region for an optimization problem, denoted $\mathcal{F}$, is the set of all points that satisfy the constraints for that problem.*

Definition 1.4.3 (Constrained (Strict) Local Minimizer)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$. Then x^* is a constrained local minimizer of f if there exists a $\delta > 0$ such that $f(x) \geq f(x^*)$, for all $x \in B(x^*, \delta) \cap \mathcal{F}$. If the inequality is strict, then it is said to be a strict constrained minimizer.*

Definition 1.4.4 (Isolated Constrained Local Minimizer)**.** *A constrained strict local minimizer is isolated if there exists an open ball $B(x^*, \delta)$ such that x^* is the only constrained minimizer in $B(x^*, \delta)$.*

Definition 1.4.5 ((Strictly) Convex Function)**.** *A twice-continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be convex if $\nabla^2 f(x)$ is positive semidefinite for all $x \in \mathbb{R}^n$. It is strictly convex if $\nabla^2 f(x)$ is positive define for all $x \in \mathbb{R}^n$.*

Definition 1.4.6 (Strongly Convex Function)**.** *A twice-continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be strongly convex if there exists an $\omega > 0$, such that $p^T \nabla^2 f(x) p \geq \omega \|p\|^2$ for all $x \in \mathbb{R}^n$ and all nonzero $p \in \mathbb{R}^n$, i.e., $\nabla^2 f(x)$ is sufficiently positive definite for all $x \in \mathbb{R}^n$.*

Definition 1.4.7 (Descent Direction)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$. Then p is a descent direction for f at a if $\nabla f(a)^T p < 0$.*

Definition 1.4.8 (Direction of Sufficient Descent)**.** *A direction p_k is a direction of sufficient descent if p_k is bounded and $\lim_{k\to\infty} \nabla f(x_k)^T p_k = 0$ implies $\lim_{k\to\infty} \nabla f(x_k) = 0$ and $\lim_{k\to\infty} p_k = 0$.*

Definition 1.4.9 (Direction of Negative Curvature)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$. Then p is a direction of negative curvature for f at a if $p^T \nabla^2 f(a) p < 0$.*

Definition 1.4.10 (Direction of Sufficient Negative Curvature)**.** *A direction p_k is a direction of sufficient negative curvature if p_k is bounded and $\lim_{k\to\infty} p_k^T \nabla^2 f(x_k) p_k = 0$ implies $\lim_{k\to\infty} \lambda_{\min}(H(x_k)) = 0$ and $\lim_{k\to\infty} p_k = 0$.*

Definition 1.4.11 (Inertia of a Matrix)**.** *If $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix, then inertia of A, denoted $\mathrm{In}(A)$, is the integer triple (i_+, i_-, i_0) indicating the number of positive, negative and zero eigenvalues of A.*

1.5 Useful Results

Result 1.5.1 (Sylvester's Law of Inertia)**.** *If $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix and $S \in \mathbb{R}^{n\times n}$ is a nonsingular matrix, then $\mathrm{In}(S^T A S) = \mathrm{In}(A)$.*

Theorem 1.5.1. *Let $H \in \mathbb{R}^{n\times n}$ be a symmetric matrix, and $A \in \mathbb{R}^{m\times n}$. Assume A has rank $r \leq m$ and let $Z \in \mathbb{R}^{n\times r}$ be a matrix whose columns span the basis of the null space of A. If*

$$K = \begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix},$$

then

$$\mathrm{In}(K) = In(Z^T H Z) + (r, r, m - r).$$

□

Corollary 1.5.1. *Let $H \in \mathbb{R}^{n\times n}$ be a symmetric matrix, and $A \in \mathbb{R}^{m\times n}$. Assume A has full row rank and let $Z \in \mathbb{R}^{n\times m}$ be a matrix whose columns span the basis of the null space of A. If*

$$K = \begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix},$$

then

$$\text{In}(K) = In(Z^T H Z) + (m, m, 0).$$

If $Z^T H Z$ is positive definite, then

$$\text{In}(K) = (n, m, 0).$$

and we say K has correct inertia.

Theorem 1.5.2. *Let $H \in \mathbb{R}^{n\times n}$ be a symmetric matrix, $D \in \mathbb{R}^{m\times m}$ a nonsingular symmetric matrix and $A \in \mathbb{R}^{m\times n}$. If*

$$K = \begin{pmatrix} H & A^T \\ A & D \end{pmatrix},$$

then

$$\text{In}(K) = In(H - A^T D^{-1} A) + In(D).$$

Proof. Consider the nonsingular matrix

$$S = \begin{pmatrix} I_n & 0 \\ -D^{-1}A & I_m \end{pmatrix}.$$

Then

$$S^T K S = \begin{pmatrix} H - A^T D^{-1} A & 0 \\ 0 & D \end{pmatrix}.$$

Hence, Sylvester's Law gives

$$\text{In}(K) = \text{In}(S^T K S) = In(H - A^T D^{-1} A) + In(D).$$

□

Theorem 1.5.3 (Debreu's Lemma)**.** *Let $H \in \mathbb{R}^{n\times n}$ be a symmetric matrix and $A \in \mathbb{R}^{m\times n}$. Then $x^T H x > 0$ for all nonzero $x \in \mathbb{R}^n$ satisfying $Ax = 0$ if and only if the exists a $\bar{\mu} \geq 0$ such that $H + \frac{1}{\mu} A^T A$ is positive definite for all $0 < \mu \leq \bar{\mu}$.* □

Theorem 1.5.4 (Debreu's Lemma Variation)**.** *Let $H \in \mathbb{R}^{n\times n}$ be a symmetric matrix, $D \in \mathbb{R}^{m\times m}$ a symmetric positive-definite matrix and $A \in \mathbb{R}^{m\times n}$. Then $x^THx > 0$ for all nonzero $x \in \mathbb{R}^n$ satisfying $Ax = 0$ if and only if the exists a $\bar{\mu} \geq 0$ such that $H + \frac{1}{\mu}A^TDA$ is positive definite for all $0 < \mu \leq \bar{\mu}$.*

Proof. (" $\Leftarrow$ ") If $H + \frac{1}{\mu}A^TDA$ is positive definite, then for all nonzero $x \in \mathbb{R}^n$

$$x^T\Big(H + \frac{1}{\mu}A^TDA\Big)x = x^THx + \frac{1}{\mu}x^TA^TDAx > 0.$$

Hence, for all nonzero x such that $Ax = 0$,

$$x^T\Big(H + \frac{1}{\mu}A^TDA\Big)x = x^THx > 0.$$

(" $\Rightarrow$ ") This direction is proved by contradiction. Let $x^THx > 0$ for all nonzero x satisfying $Ax = 0$ and assume that there is no $\bar{\mu}$ such that $H + \frac{1}{\mu}A^TDA$ is positive definite for all $0 < \mu \leq \bar{\mu}$. To that end, define a positive sequence $\{\mu_k\}$ such that $\{\mu_k\} \to 0$ and a corresponding sequence $\{x_k\}$ with the property that

$$x_k^T\Big(H + \frac{1}{\mu_k}A^TDA\Big)x_k \leq 0. \tag{1.6}$$

Without loss of generality, assume that $\|x_k\| = 1$. Then $\{x_k\} \in S(0,1)$ and since $S(0,1)$ is a compact set, $\{x_k\}$ has a convergent subsequence $\{x_k\}_K \to \bar{x} \in S(0,1)$. Multiplying (1.6) by μ_k and taking limits on $k \in K$ gives

$$\bar{x}^TA^TDA\bar{x} = \lim_{k\in K} x_k^T(\mu_kH + A^TDA)x_k \leq 0. \tag{1.7}$$

Since D is symmetric and positive definite, it can be written as $D = R^TR$, where R is nonsingular. Hence

$$\bar{x}^TA^TDA\bar{x} = \bar{x}^TA^TR^TRA\bar{x} = \|RA\bar{x}\|^2$$

and (1.7) implies that $RA\bar{x} = 0$. Since R is nonsingular, it must be that $A\bar{x} = 0$.

Next, (1.6) implies

$$\mu_kx_k^THx_k + x_k^TA^TDAx_k = \mu_kx_k^THx_k + \|RAx_k\|^2 \leq 0.$$

As $\|RAx_k\| \geq 0$ and $\mu_k > 0$, it must follow that $x_k^THx_k < 0$ for all k. Taking limits on $k \in K$ yields

$$\bar{x}^TH\bar{x} = \lim_{k\in K} x_k^THx_k < 0.$$

So $A\bar{x} = 0$, but $\bar{x}^TH\bar{x} < 0$ which is a contradiction. □

2 Background

2.1 Unconstrained Optimization

Consider the unconstrained minimization problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice-continuously differentiable.

This section reviews the necessary and sufficient conditions for a point to be a local minimizer of f, as well as some basic methods for finding such points.

2.1.1 Optimality Conditions

Theorem 2.1.1 (First-Order Necessary Conditions)**.** *Let x^* be a local minimizer of f. Then $\nabla f(x^*) = 0$.*

This condition is not sufficient for optimality. The following theorems establish conditions on f that guarantee a minimum at x^*.

Theorem 2.1.2 (Second-Order Necessary Conditions)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice-continuously differentiable. Then x^* is a local minimizer of f only if*

1. $\nabla f(x^*) = 0$,
2. $\nabla^2 f(x^*)$ *is positive semidefinite.*

The conditions of Theorem 2.1.2 are necessary, but not sufficient for optimality. The next theorem gives sufficient conditions that guarantee that x^* is a strict local minimizer.

Theorem 2.1.3 (Second-Order Sufficient Conditions)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice-continuously differentiable. Then x^* is a strict local minimizer of f if*

1. $\nabla f(x^*) = 0$,
2. $\nabla^2 f(x^*)$ *is positive definite.*

2.1.2 Newton's method

Given a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, its minimizers are found among the solutions of the system of equations $\nabla f(x) = 0$, i.e., among the zeros of the vector-valued function $\nabla f(x)$. In general this system is nonlinear and must be solved using an iterative method. In this section we review variants of Newton's method for finding zeros of a vector-valued function.

Given a starting point x_0, Newton's method defines a sequence of points $\{x_k\}$ that, under appropriate conditions, converges to a zero of ∇f at a quadratic rate. Each new iterate x_{k+1} is defined as the zero of the linearization of $\nabla f(x)$ at the previous iterate x_k, i.e., x_{k+1} must satisfy

$$\nabla^2 f(x_k)(x_{k+1} - x_k) + \nabla f(x_k) = 0.$$

Hence, if $\nabla^2 f(x)$ is nonsingular, the next iterate is given by $x_{k+1} = x_k + p_k$, where p_k satisfies

$$\nabla^2 f(x_k) p_k = -\nabla f(x_k). \tag{2.1}$$

Equations (2.1) are known as the Newton equations.

In the context of optimization, Newton's method may be viewed as a method for finding minimizers of a sequence of quadratic models of the objective function f. A quadratic model can be obtained by taking the first three terms of the Taylor-series approximation of f about the current point x_k, i.e.,

$$f(x_k + p) \approx f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p.$$

This approximation is minimized if p minimizes the quadratic function

$$m_k(p) = \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p,$$

which models the change in the objective $f(x_k+p)-f(x_k)$. From Section 2.1.1, any minimizer of m_k must satisfy $\nabla m_k(p) = 0$, or equivalently $\nabla^2 f(x_k)p = -\nabla f(x_k)$. If $\nabla^2 f(x_k)$ is positive definite then p is the unique unconstrained minimizer of $m_k(x)$.

Algorithm 2.1.1 Pure Newton's Method

INPUT: x_0;
  Set $k = 0$;
  while $\nabla f(x_k) \neq 0$ **do**
    Solve $\nabla^2 f(x_k)p_k = -\nabla f(x_k)$;
    $x_k \leftarrow x_k + p_k$;
    $k \leftarrow k + 1$;
  end while
OUTPUT: x_k

Newton's method is attractive because of its potential quadratic rate of convergence. However, the Newton method of Algorithm 2.1.1 has drawbacks. Far from a solution the quadratic model may be a poor approximation for a general nonlinear function f. As a consequence, if the starting point is not sufficiently close to a solution, Newton's method may fail to converge. Another drawback is that if $\nabla^2 f(x_k)$ is not positive definite, the quadratic model may not have a bounded minimum (it may not even have a stationary point if $\nabla^2 f(x_k)$ has a zero eigenvalue). This implies that there is no guarantee that a stationary point produced by Newton's method is a minimizer unless the Hessian is positive definite, i.e., unless f is locally strictly convex.

2.1.3 Line-Search Methods

As discussed in the previous section, Newton's Method may not converge to a solution unless the starting point is sufficiently close to a stationary point of f. Safeguards must me implemented to guarantee convergence from any starting point. A common strategy is to include a line search along the direction p_k. In this case, the new iterate is defined by $x_{k+1} = x_k + \alpha_k p_k$, where α_k is a positive

step length. The step α_k is chosen so that the function is sufficiently reduced by the move to x_{k+1}. A common condition for guaranteeing such a reduction is that the actual change in function value is at least a fraction of the change predicted by the local model $m_k(x)$, i.e., we require

$$f(x_k) - f(x_{k+1}) \geq \eta(m_k(x_k) - m_k(x_{k+1})) = \eta(m_k(x_k) - m_k(x_k + \alpha_k p_k)), \quad (2.2)$$

for some $\eta \in (0, \frac{1}{2})$.

An essential requirement for any line-search model is the existence of a positive α_k that satisfies (2.2). In particular, we require that there exists an $\bar{\alpha}$ such that

$$f(x_k + \alpha p_k) \leq f(x_k) - \eta(m_k(x_k) - m_k(x_k + \alpha p_k)), \qquad \forall \alpha \in (0, \bar{\alpha}). \quad (2.3)$$

Consider the following procedure for finding an acceptable step length.

Algorithm 2.1.2 Backtracking Line Search

INPUT: x_k, p_k, $\eta \in (0, \frac{1}{2})$, $\gamma \in (0, 1)$;

Set $\alpha = 1$;

while $f(x_k + \alpha p_k) > f(x_k) - \eta(m_k(x_k) - m_k(x_k + \alpha p_k))$ **do**

$\alpha \leftarrow \gamma\alpha$;

end while

$x_{k+1} \leftarrow x_k + \alpha p_k$;

OUTPUT: x_{k+1}

Under assumption (2.3) on the local model and given appropriate descent directions p_k, it can be shown that a backtracking line search is enough to guarantee a sufficient decrease in f.

The local model m_k used in line-search need not be the same model used to define the vector p_k. The model may be based on either first- or second-order Taylor-series approximations of f at x_k. For example, if the model is a first-order approximation, i.e., $m_k(x) = f(x_k) + \nabla f(x_k)^T(x - x_k)$, then condition (2.2) is

$$f(x_{k+1}) \leq f(x_k) + \eta\alpha\nabla f(x_k)^T p_k,$$

which is often referred to as the Armijo condition.

2.1.4 Search Directions

As mentioned in Section 2.1.2, the search direction defined by the Newton equations is appropriate only when the Hessian is positive definite. As such, including a line search with a Newton's method is not enough to force global convergence. Thus we would like to modify Newton's method in such a way that the computed direction is always a descent direction, but is identical to the Newton direction in the neighborhood of a solution. Such *modified* Newton methods take advantage of the fast local convergence of Newton's method.

One approach is to modify $\nabla^2 f(x_k)$ to make it positive definite. In this case, instead of solving the conventional Newton equations, we solve

$$B_k p_k = -\nabla f(x_k), \tag{2.4}$$

where B_k is some positive-definite approximation of $\nabla^2 f(x_k)$ that is chosen so that $B_k = \nabla^2 f(x_k)$ if $\nabla^2 f(x_k)$ is positive definite. The modified Hessian B_k is usually given by $B_k = \nabla^2 f(x_k) + E$, where E is positive semidefinite (see Section 2.4.3 for an example of how to compute E). As B_k is positive definite, the solution of the modified Newton system (2.4) is guaranteed to be a descent direction.

However the modified Newton equations (2.4) still fail to provide a descent direction if x_k happens to be a saddle point or a maximizer, i.e., $\nabla f(x_k) = 0$, but $\nabla^2 f(x_k)$ is indefinite or negative definite. In such cases the search direction p_k can be defined as a direction of negative curvature, i.e., it satisfies

$$p_k^T \nabla^2 f(x_k) p_k < 0. \tag{2.5}$$

It is not uncommon to use search directions that are a combination of a modified Newton direction and a direction of negative curvature. Consider, for example the modified Newton method in Algorithm 2.1.3. Moré and Sorensen [19] show that if s_k and d_k are directions of sufficient descent and negative curvature, respectively, and the line search ensures that $\nabla f(x_k)^T s_k \to 0$ and $d_k^T \nabla^2 f(x_k) d_k \to 0$, then every limit point of the sequence of iterates $\{x_k\}$ will satisfy the second-order necessary conditions for optimality.

Algorithm 2.1.3 A Modified Newton's Method with Backtracking Line Search

INPUT: x_0, $\eta \in (0, \frac{1}{2})$, $\gamma \in (0, 1)$;

$k = 0$;

while $\nabla f(x_k) \neq 0$ **and** $\nabla^2 f(x_k)$ not positive semidefinite **do**

if $\nabla^2 f(x_k)$ is p.d. **then**

Solve $\nabla^2 f(x_k) s_k = -\nabla f(x_k)$;

$d_k = 0$;

else

Solve $B_k p_k = -\nabla f(x_k)$;

Find d_k such that $d_k^T \nabla^2 f(x_k) d_k < 0$ and $\nabla f(x_k)^T d_k < 0$;

end if

$p_k \leftarrow s_k + d_k$;

Set $\alpha = 1$;

while $f(x_k + \alpha p_k) > f(x_k) - \eta(m_k(x_k) - m_k(x_k + \alpha p_k))$ **do**

$\alpha \leftarrow \gamma\alpha$;

end while

$x \leftarrow x_k + \alpha p_k$;

$k \leftarrow k + 1$;

end while

OUTPUT: x_k

2.2 Quadratic Programming

Consider the mixed-constraint QP

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & q(x) = \frac{1}{2} x^T H x + c^T x \\ \text{subject to} & Ax = b, \qquad Dx \geq f. \end{array} \tag{2.6}$$

In this section we discuss necessary and sufficient conditions for a point x^* to be the second-order minimizer of the mixed-constraint QP. These conditions can easily be adapted to quadratic problems in either all-inequality form or standard form.

2.2.1 Optimality Conditions

We begin by stating conditions that are necessary for a point x^* to be a first-order minimizer of the QP (2.6).

Definition 2.2.1 (KKT conditions)**.** *The vector $x^* \in \mathbb{R}^n$ is a first-order KKT point if there exist $y^* \in \mathbb{R}^m$ and $z^* \in \mathbb{R}^n$ such that*

1. *$Ax^* = b$ and $Dx^* \geq f$ (feasibility),*
2. *$g(x^*) = A^T y^* + D^T z^*$ (stationarity),*
3. *$z^* \geq 0$ (nonnegativity),*
4. *$(Dx^* - f) \cdot z^* = 0$ (complementarity).*

Unless the QP is convex, the KKT conditions are not enough to ensure optimality and second-order conditions are required. Second-order optimality conditions require the definition of the active set and inactive set of constraints at a given point.

Definition 2.2.2. *For the set of inequality constraints $Dx \geq f$, we say that a constraint i is active at a point x if $d_i^T x = f_i$, and inactive if $d_i^T x > f_i$. The active set at x, denoted as $\mathcal{A}(x)$, is the set of indices of the constraints that are active at x, i.e., $\mathcal{A}(x) = \{i : d_i^T x = f_i\}$. If the context is clear, the argument is dropped and the active set is simply referred to as $\mathcal{A}$. We also use the notation $D_{\mathcal{A}}$ to denote a submatrix consisting of rows of D whose indices are in $\mathcal{A}$.*

The following second-order sufficiency result is stated here without proof (see Borwein [1], Contesse [5], Majthay [16] and McCormick [18]).

Theorem 2.2.1 (Second-order Necessary and Sufficient Conditions)**.** *The vector $x^* \in \mathbb{R}^n$ is a (strict) local minimizer of (2.6) if and only if*

1. *x^* is a KKT point with multipliers y^* and z^*.*
2. *$p^T H p (>) \geq 0$ for all nonzero $p \in \mathbb{R}^n$ satisfying $Ap = 0$, $g(x^*)^T p = 0$ and $D_{\mathcal{A}} p \geq 0$.*

An alternative formulation of Condition 2 is

2. $p^T Hp(>) \geq 0$ *for all nonzero* $p \in \mathbb{R}^n$ *satisfying* $Ap = 0$, $D_{\mathcal{A}_+} p = 0$ *and* $D_{\mathcal{A}_0} p \geq 0$,

where $\mathcal{A}_+(x^*) = \{i \in \mathcal{A}(x^*) : z_i^* > 0\}$ and $\mathcal{A}_0(x^*) = \{i \in \mathcal{A}(x^*) : z_i^* = 0\}$. The purpose of second-order sufficient conditions is to be able to verify that a given point is indeed the minimizer. However, Condition 2 of the theorem above is not easy to verify computationally. In particular, checking that $p^T Hp \geq 0$ over all p satisfying $D_{\mathcal{A}} p \geq 0$, if H is indefinite, is NP-hard (see Pardalos and Schnitger [20] and Pardalos and Vavasis [21]). The difficulty arises in the presence of zero multipliers that correspond to the active constraints at x^*. The complementarity condition in Definition 2.2.1 forces z_i^* to be zero if constraint i is inactive yet allows the possibility that $z_i^* = 0$ when constraint i is active. If there are no zero active multipliers, Condition 2 can be written in a computationally tractable form. Thus, it may be necessary to restrict the set of possible multipliers by assuming that only strictly positive multipliers correspond to active constraints. This assumption is called strict complementarity and is formally defined below.

Definition 2.2.3 (strict complementarity)**.** *Strict complementarity holds at a KKT point* x^**, if there exists an* z^* *such that* $z_{\mathcal{A}}^* > 0$.

If we impose strict complementarity on a KKT point x^*, then Condition 2 is reduced to

2. $p^T Hp(>) \geq 0$ *for all nonzero* $p \in \mathbb{R}^n$ *satisfying* $Ap = 0$ *and* $D_{\mathcal{A}} p = 0$

and is fairly straightforward to verify.

It is possible to state computationally tractable sufficiency conditions without requiring strict complementarity by placing restrictions on H instead. This new sufficiency condition 2 can be stated as

2. $p^T Hp(>) \geq 0$ *for all nonzero* $p \in \mathbb{R}^n$ *satisfying* $Ap = 0$, $D_{\mathcal{A}_+} p = 0$.

2.3 Penalty and Augmented Lagrangian Functions

One way to solve a constrained optimization problem is to convert it into a sequence of parameterized unconstrained problems. Typically, a penalty function is used for the equality constraints and a barrier function is used for the inequality constraints. Barrier functions will be discussed in great detail in subsequent chapters. This section provides a brief survey of penalty functions. To that end consider the following equality constrained program

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & c(x) = 0, \end{aligned} \tag{2.7}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are twice differentiable functions.

One of the best-known penalty functions for treating the equality constraints of (2.7) is the classical quadratic penalty function

$$P(x;\mu) = f(x) + \frac{1}{2\mu}\|c(x)\|^2,$$

where μ is a scalar parameter. Under certain assumptions (see e.g., [6], [9]), the unconstrained minimizers of the penalty function $P(x;\mu)$ form a differentiable trajectory that approaches the solution of the constrained problem (2.7) as $\mu \to \infty$. The penalty method usually has as a two-level structure. The inner iterations minimize the penalty function for the current value of the parameter μ, usually using some form of Newton's method. The outer iterations check the optimality conditions for the original problems and decrease the penalty parameter μ if necessary.

As $\mu \to 0$, the Newton equations become increasingly ill-conditioned. However, as the following theorem shows, this ill-conditioning is largely benign.

Theorem 2.3.1. *Let $J(x)$ denote the $m \times n$ Jacobian of $c(x)$, and let $Z(x)$ be any matrix whose columns form a basis for the null space of $J(x)$. Consider the direction p satisfying the Newton equations $\nabla^2 P(x;\mu)p = -\nabla P(x;\mu)$. Then*

1. *the Newton direction p satisfies the equations*

$$\begin{pmatrix} H & J(x)^T \\ J(x) & -\mu I \end{pmatrix} \begin{pmatrix} p \\ -y \end{pmatrix} = - \begin{pmatrix} \nabla f(x) \\ c(x) \end{pmatrix}, \tag{2.8}$$

where $H = \nabla^2 f(x) + \dfrac{1}{\mu} \sum_{i=1}^{m} c_i(x) \nabla^2 c_i(x)$,

2. *if $Z(x)^T H Z(x)$ is positive definite and $\nabla P(x;\mu) \neq 0$, then there exists a finite non-negative $\bar{\mu}$ such that, for all $\bar{\mu} > \mu$, p is a descent direction for $P(x;\mu)$.*

Proof. To simplicity the notation, we write $g = \nabla f(x)$, $J = J(x)$ and $c = c(x)$.

The Newton equations for minimizing $P(x;\mu)$ are

$$\nabla^2 P(x;\mu) p = -\nabla P(x;\mu),$$

where

$$\begin{aligned} \nabla P(x;\mu) &= g + \frac{1}{\mu} J^T c, \\ \nabla^2 P(x;\mu) &= \nabla^2 f(x) + \frac{1}{\mu} \sum_{i=1}^{m} c_i \nabla^2 c_i + \frac{1}{\mu} J^T J. \end{aligned}$$

Noting that $\nabla^2 f(x) + \dfrac{1}{\mu} \sum_{i=1}^{m} c_i \nabla^2 c_i = H$, we get

$$(H + \frac{1}{\mu} J^T J) p = -(g + \frac{1}{\mu} J^T c).$$

On the other hand, equation (2.8) can be written explicitly as

$$\begin{aligned} Hp - J^T y &= -g \\ Jp + \mu y &= -c. \end{aligned}$$

By first eliminating y and obtaining a system involving only p, we obtain the equivalent systems

$$\begin{aligned} (H + \frac{1}{\mu} J^T J) p &= -(g + \frac{1}{\mu} J^T c) \\ \mu y &= -(c + Jp). \end{aligned} \tag{2.9}$$

To prove part (2), it is necessary to show that $\nabla P(x;\mu)^T p < 0$ for all μ sufficiently small. If $Z(x)^T H Z(x)$ is positive definite, then $p^T H p > 0$ for all nonzero p satisfying $Jp = 0$. Then, by Theorem 1.5.3, there exists a $\bar{\mu} > 0$ such that $H + \frac{1}{\mu} J^T J$ is positive definite for all $0 < \mu < \bar{\mu}$. Noting that the right hand side of equation (2.9) is just $-\nabla P(x;\mu)$, it follows that

$$\nabla P(x;\mu)^T p = -\nabla P(x;\mu)^T \big(H + \frac{1}{\mu} J^T J\big)^{-1} \nabla P(x;\mu) < 0$$

as long as $\nabla P(x;\mu) \neq 0$. □

Quadratic penalty methods were largely superseded by methods based on the augmented Lagrangian function, which was proposed independently by Hestenes [15] and Powell [23]. An augmented Lagrangian function has the form

$$P(x;\mu,y^a) = f(x) - c(x)^T y^a + \frac{1}{2\mu}\|c(x)\|^2,$$

where $y^a \in \mathbb{R}^m$ is an estimate of the vector of Lagrange multipliers at a solution of (2.7). The augmented Lagrangian function can be viewed as the classical quadratic penalty function applied to the following shifted version of problem (2.7)

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & c(x) - \mu y^a = 0. \end{array} \tag{2.10}$$

If $y^a = y^*$ and μ is sufficiently small, the solution x^* associated with y^* may be found with a single unconstrained minimization of $P(x;\mu,y^a)$. This result is used to formulate methods in which the inner iterations involve minimizing an unconstrained augmented Lagrangian function, and the outer iterations check for termination and define a new multiplier estimate.

Further notable advances in penalty methods can be traced to the development of efficient primal-dual path-following methods. A primal-dual penalty function can be derived by defining a primal-dual pair $(x(\mu), y(\mu))$, such that $x(\mu)$ minimizes $P(x;\mu)$, and $y(\mu)$ is the vector of Lagrange multipliers for the shifted problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & c(x) = c(x(\mu)). \end{array}$$

The primal-dual pair $(x(\mu), y(\mu))$ satisfies the equations:

$$\nabla f(x) - \sum_{i=1}^{m} y_i \nabla c_i(x) = 0 \quad \text{and} \quad y_i = -\frac{1}{\mu} c_i(x),$$

which may be rewritten as:

$$\begin{aligned} \nabla f(x) - J(x)^T y &= 0, \\ c(x) + \mu y &= 0, \end{aligned}$$

where $J(x)$ is the Jacobian of the constraints $c(x)$. The Forsgren-Gill primal-dual function for the equality constrained problem (2.7) is given by

$$P^{\nu}(x, y; \mu) = f(x) + \frac{1}{2\mu}\|c(x)\|^2 + \frac{\nu}{2\mu}\|c(x) + \mu y\|^2, \tag{2.11}$$

for a fixed positive scalar ν (see [7]). The Forsgren-Gill penalty function is the quadratic penalty function augmented by a term that penalizes deviations from the trajectory of minimizers $(x(\mu), y(\mu))$.

A primal-dual function that does not require $\mu \to 0$ and allows the use of estimates of the multipliers y, may be derived by applying the primal-dual penalty function (2.11) to the shifted problem (2.10) (see Robinson [24], Gill and Robinson [12, 13]). This gives the primal-dual augmented Lagrangian function

$$P^{\nu}(x, y; \mu, y^a) = f(x) - c(x)^T y^a + \frac{1}{2\mu}\|c(x)\|^2 + \frac{\nu}{2\mu}\|c(x) + \mu(y - y^a)\|^2.$$

This function will be discussed in more detail in Chapter 5 in the context of solving quadratic programs in standard form.

2.4 Convexification Methods

Convexification is a process for defining a local convex approximation of a nonconvex problem. This approximation may be defined on the full space of variables or just on some subset. Many model-based optimization methods use some form of convexification. For example, line-search methods for unconstrained and linearly-constrained optimization define a convex local quadratic model in which the Hessian $H(x_k)$ is replaced by a positive-definite matrix $H(x_k) + E_k$ (see,

e.g., Greenstadt [14], Gill and Murray [11], Schnabel and Eskow [25], and Forsgren and Murray [10]). All of these methods are based on convexifying an unconstrained or equality-constrained local model.

Ideally, any algorithm for computing E_k should be "minimally invasive", i.e., if Hessian is positive definite, then the modification should be zero.

2.4.1 Symmetric Indefinite Factorization

Each iteration of a primal-dual method requires the solution of a Newton type system of equations involving a symmetric primal-dual KKT matrix

$$K = \begin{pmatrix} H & A^T \\ A & -D \end{pmatrix}, \tag{2.12}$$

where H is an $n \times n$ symmetric matrix, A is an $m \times n$ matrix and D is an $m \times m$ positive-definite diagonal matrix. For systems with large dimensions this computation dominates the cost of an iteration. Many methods, both iterative and direct, have been proposed for solving these equations. Some of the most successful methods rely on factorizing K using a symmetric indefinite factorization (see Bunch and Parlett [4] and Bunch and Kaufman [2]). If K is defined as above, then there exists a permutation P, a block-diagonal $B = \text{diag}(B_1, B_2, \ldots, B_s)$, and a unit lower-triangular L, such that $P^T K P = LBL^T$. The diagonal blocks B_j are either one-by-one or two-by-two and are nonsingular if $\begin{pmatrix} H & A^T \end{pmatrix}$ has full rank, otherwise some trailing blocks B_j may be zero.

Before we can describe LBL^T factorization, the following definition is needed.

Definition 2.4.1. *Consider the following partition of a symmetric matrix K.*

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

where $K_{12} = K_{21}^T$. The Schur complement of K_{11} in K, denoted K/K_{11}, is defined as

$$K/K_{11} = K_{22} - K_{21}K_{11}^{-1}K_{12}.$$

With this definition, K may be written in the form

$$\begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} = \begin{pmatrix} I & 0 \\ K_{21}K_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} K_{11} & 0 \\ 0 & K/K_{11} \end{pmatrix} \begin{pmatrix} I & K_{11}^{-1}K_{12} \\ 0 & I \end{pmatrix}.$$

To start the factorization, a symmetric 1×1 or 2×2 pivot is chosen and brought to the leading position using symmetric row and column interchanges. An elimination step is then applied that, depending on the size of the pivot, yields the first one or two columns of L. The pivot K_{11}, (say) becomes the first diagonal block of B and the associated Schur complement $P_1^T K P_1/K_{11}$ remains to be factorized. This first elimination step may be written in the form

$$P_1^T K P_1 = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} = L_1 B_{11} L_1^T + \begin{pmatrix} 0 & 0 \\ 0 & S \end{pmatrix},$$

where K_{11} is 1×1 or 2×2,

$$L_1 = \begin{pmatrix} I \\ K_{21}K_{11}^{-1} \end{pmatrix}, \quad B_{11} = K_{11}, \quad \text{and} \quad S = K_{22} - K_{21}K_{11}^{-1}K_{12}.$$

In subsequent steps this procedure is repeated until the Schur complement has dimension zero and complete LBL^T factorization is obtained.

The numerical stability of the LBL^T factorization depends of the choice of pivots, and a number of strategies have been proposed (see, e.g., Bunch, Kaufman, and Parlett [3]). In the next section a strategy is defined that controls the inertia of the block diagonal matrix B.

2.4.2 The Inertia Controlling LBL^T Factorization

An inertia controlling factorization may be used in conjunction with a convexification method described in the next section, to define a modified Hessian for the primal-dual KKT matrix (2.12). The factorization defines a pivot order that facilitates the definition of a modified Hessian $H + E$ giving a KKT matrix of correct inertia (see Forsgren and Murray [10], Forsgren and Gill [7]).

If K is the primal-dual matrix (2.12) defined in the previous section, 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 pivots. We can further distinguish between pivot types based on, whether the pivot has positive, negative, zero or mixed eigenvalues. Let the superscripts $+$ and $-$ denote the sign of the eigenvalue. For 1×1 pivots, the possible pivot types are H^+, H^-, and D^-. For 2×2 pivots, the possibilities are HH^{++}, HH^{+-}, HH^{--}, HD^{+-} and DD^{--}.

The inertia-controlling factorization consists of two phases. To ensure a certain inertia of the B matrix, only pivots of type H^+, D^-, HH^{++}, HD^{+-} and D^{--} are allowed in the first phase. The first phase continues until no D pivots or HD pivots remain in the Schur complement. For example consider the following partition of K defined in (2.12)

$$K = \begin{pmatrix} H_{11} & H_{12} & A_1^T \\ H_{21} & H_{22} & A_2^T \\ A_1 & A_2 & -D \end{pmatrix}.$$

For the purpose of this discussion and without loss of generality, assume that the order of pivots in the first phase doesn't matter and all pivots come form the rows corresponding to the H_{11} and D blocks. Let $\widehat{P}$ be the permutation such that at at the end of phase one

$$\widehat{P}^T K \widehat{P} = \begin{pmatrix} H_{11} & A_1^T & H_{12} \\ A_1 & -D & A_2 \\ H_{21} & A_2^T & H_{22} \end{pmatrix}.$$

Thus, at the end of the first phase, we end up with

$$\widehat{P}^T K \widehat{P} = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} = \begin{pmatrix} L_{11} \\ L_{21} \end{pmatrix} B_{11} \begin{pmatrix} L_{11}^T & L_{21}^T \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & S \end{pmatrix},$$

where

$$K_{11} = \begin{pmatrix} H_{11} & A_1^T \\ A_1 & -D \end{pmatrix},$$

$K_{21} = K_{12}^T = \begin{pmatrix} H_{21} & A_2^T \end{pmatrix}$, $K_{22} = H_{22}$ and $S = K_{22} - K_{21}K_{11}^{-1}K_{12}$ is the Schur complement left to be factorized. The proposed pivot strategy ensures that the inertia of B_{11} block diagonal matrix is given by $\text{In}(B_{11}) = (n_1, m, 0)$, where $n_1 \leq n$ depends of the number of H rows used in the first phase.

In the second phase, the choice of type of pivot is unrestricted, and the only consideration is that $\|L\|$ is bounded. On completion of phase two, we have $K_{22} - K_{21}K_{11}^{-1}K_{12} = L_{22}B_{22}L_{22}^T$ and this gives the complete factorization

$$\begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} B_{11} & 0 \\ 0 & B_{22} \end{pmatrix} \begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22}^T \end{pmatrix}. \tag{2.13}$$

2.4.3 Computing a Modified KKT Matrix

The inertia controlling factorization allows us to easily compute a modified matrix $\bar{H} = H + E$, where E is symmetric positive semidefinite, such that the modified KKT matrix

$$\bar{K} = \begin{pmatrix} \bar{H} & A^T \\ A & -D \end{pmatrix} \tag{2.14}$$

has correct inertia. Consider the complete factorization given by (2.13). As $B_{22} = \text{diag}(B_1, \ldots, B_l)$ is block diagonal, where each B_j block is either 1×1 or symmetric 2×2, its is fairly straight forward to compute a symmetric positive definite modification $\bar{B}_{22} = B_{22} + E_{22}$. For example, consider the following scheme for computing E_{22}. Let $E_{22} = \text{diag}(E_1, \ldots, E_l)$, where each E_j has the same dimension as the corresponding B_j. Let $0 < \omega \ll 1$ be the tolerance for the smallest eigenvalue allowed. If B_j is a 1×1 block, then

$$E_j = \begin{cases} \max(\omega, -B_j) - B_j & \text{if } B_j < \omega \\ 0 & \text{otherwise.} \end{cases}$$

If B_j is a 2×2 block, then E_j can be computed using eigenvalues of B_j. Note that as B_j is symmetric, its two eigenvalues are real and can be easily computed. Let $\alpha = \max(\omega, -\lambda_{\min}(B_j)) - \lambda_{\min}(B_j))$, then define

$$E_j = \begin{cases} \alpha I & \text{if } \lambda_{\min}(B_j) < \omega \\ 0 & \text{otherwise.} \end{cases}$$

As K_{22} consists entirely of elements from rows and columns of H, any modification to B_{22} only affects H and we therefore obtain $\bar{K}$ defined in (2.14), and a symmetric positive semi-definite $E = \bar{P}^T \operatorname{diag}(0, E_{22})\bar{P}$, where $\bar{P}$ is the appropriate permutation matrix.

The following result follows.

Result 2.4.1. *Let $\bar{B}_{22}$ be a sufficiently positive definite modification of B_{22} of the completed inertia controlling factorization* (2.13)*. Then*

$$\bar{K} = \begin{pmatrix} \bar{H} & A^T \\ A & -D \end{pmatrix}$$

has correct inertia and $\bar{H} + A^T D^{-1} A$ is sufficiently positive definite.

Proof. Sylvester's Law of Inertia gives

$$\operatorname{In}(K) = \operatorname{In}(P^T K P) = \operatorname{In}(LBL^T) = \operatorname{In}(B)$$

and hence

$$\operatorname{In}(K) = \operatorname{In}(B_{11}) + \operatorname{In}(B_{22}) = (n1, m, 0) + \operatorname{In}(B_{22}).$$

So if $\bar{B}_{22}$ is sufficiently positive definite, $\operatorname{In}(\bar{K}) = (n, m, 0)$. Theorem 1.5.2 gives $\operatorname{In}(\bar{K}) = \operatorname{In}(\bar{H} + A^T D^{-1} A) + (0, m, 0)$ and so $\operatorname{In}(\bar{H} + A^T D^{-1} A) = n$. □

3 Barrier Methods

In this section we present an overview of barrier methods as they pertain to a more general nonlinear optimization problem given by

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & c(x) \geq 0, \end{aligned} \tag{3.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and c is an m-vector of constraint functions with each $c_i : \mathbb{R}^n \to \mathbb{R}$ twice continuously differentiable. We say that the nonlinear program is convex, if the objective function is convex and the inequality-constraint functions are concave (i.e., $-c_i(x)$ is a convex function).

Throughout this chapter the following notation will be used:

- $J(x)$ is the constraint Jacobian;
- $L(x, z) = f(x) - z^T c(x)$;
- $H(x, z) = L_{xx}(x, z) = \nabla^2 f(x) - \sum_{i=1}^{m} z_i \nabla^2 c_i(x)$;
- (x_k, z_k) denotes an approximate primal-dual solution to (3.1) at step k, where z_k is a vector of dual variables for the inequality constraints;
- $f_k = f(x_k)$, $\nabla f_k = \nabla f(x_k)$, $c_k = c(x_k)$, $J_k = J(x_k)$ and $H_k = H(x_k, z_k)$;
- $\mathcal{A} = \{i : c_i(x^*) = 0\}$ denotes the index set of constraints that are active at a solution x^*;
- $\mathcal{A}_+ = \{i \in \mathcal{A} : z_i^* > 0\}$ denotes the index set of constraints active at a solution for which a corresponding optimal Lagrange multiplier is strictly positive;

- $\mathcal{F} = \{x \in \mathbb{R}^n : c(x) \geq 0\}$ denotes the feasible set.

3.1 Optimality Conditions

In the convergence theorems for various barrier methods presented in this chapter it is necessary to make certain first- and second-order assumptions on a local minimizer x^* of problem (3.1).

Definition 3.1.1. *A point x^* is a first-order KKT point, if there exists a $z^* \in \mathbb{R}^m$, such that*

1. $c(x^*) \geq 0$,
2. $\nabla f(x^*) - J(x^*)^T z^* = 0$,
3. $z^* \geq 0$, *and*
4. $c(x^*) \cdot z^* = 0$.

In general, the multipliers at a KKT point are not unique. The following definition characterizes the set of acceptable multipliers at a KKT point x^*.

Definition 3.1.2. *Given a KKT point x^* for problem* (3.1)*, the set of acceptable multiplies is defined as*

$$\mathcal{M}_z(x^*) = \{z : \nabla f(x^*) = J(x^*)^T z, \;\; z \geq 0, \;\; c(x^*) \cdot z = 0\}. \tag{3.2}$$

In the case of quadratic programming, where the constraints are linear, an optimal solution is a first-order KKT point. However, this is not the case for a problem with nonlinear constraints unless certain constraint qualifications or regularity conditions hold at x^*. There are many constraint qualifications that guarantee a local minimizer is a KKT point. Here we give the two most common ones.

Definition 3.1.3 (Linear Independence Constraint Qualification (LICQ))**.** *For problem* (3.1)*, the linear independence constraint qualification holds at the feasible point x, if x is strictly feasible, i.e., there are no constraints active at x, or if*

$J_{\mathcal{A}}(x)$ has full row rank, i.e., if the gradients of the active constraints are linearly independent.

Another constraint regularity is due to Mangasarian and Fromovitz [17].

Definition 3.1.4 (Mangasarian-Fromovitz Constraint Qualification(MFCQ))**.** *For problem* (3.1)*, the Mangasarian-Fromovitz constraint qualification holds at the feasible point x, if x is strictly feasible or if there exists a vector p such that $J_{\mathcal{A}}(x)p > 0$.*

The main properties of these constraint qualifications may be summarized as follows:

1. The MFCQ is a weaker condition then LICQ, in the sense that if LICQ holds, then so does MFCQ, but not the other way around.

2. If LICQ holds at a KKT point x^*, then the Lagrange multiplier vector z^* is unique.

3. If MFCQ holds at a KKT point x^*, the set of acceptable Lagrange multipliers $\mathcal{M}_z(x^*)$ is bounded.

4. LICQ is more practical in the sense that is is computationally easier to verify.

We can now state first-order necessary conditions for a local minimizer.

Theorem 3.1.1 (First-order Necessary Conditions)**.** *For problem* (3.1)*, if x^* is a local minimizer for which MFCQ holds, then x^* is a KKT point.* □

As with quadratic programming, unless the nonlinear problem is convex, first-order conditions are not enough to ensure optimality and second-order conditions are required. We first state the necessary second-order conditions for a local minimizer

Theorem 3.1.2 (Second-order Necessary Conditions)**.** *Let x^* be a local minimizer of problem* (3.1) *at which LICQ holds. Then x^* is a KKT point and, further, it must be true that $p^T H(x^*, z^*)p \geq 0$ for all p satisfying $J_{\mathcal{A}}(x^*)p = 0$.* □

Necessary conditions let us check if a point is not a local minimizer. We need sufficient conditions to conclude that x^* is indeed a local minimizer of the inequality constrained nonlinear program. Several second-order conditions exist that guarantee a minimum.

Definition 3.1.5 (First Second-order Sufficient Condition (SSC1))**.** *Let x^* be a KKT point of problem* (3.1)*. Then SSC1 holds for (x^*, z^*) if there exists $\omega > 0$ such that $p^T H(x^*, z^*)p \geq \omega\|p\|^2$ for all nonzero p satisfying $g(x^*)^T p = 0$ and $J_{\mathcal{A}}(x^*)p \geq 0$.*

The SSC1 requires that the condition holds for just one z^* in the set of acceptable multipliers at x^*. The next sufficiency condition is more demanding in that it places the requirement on all acceptable multipliers.

Definition 3.1.6 (Second Second-order Sufficient Condition (SSC2))**.** *Let x^* be a KKT point of problem* (3.1)*. Then SSC2 holds for x^* if for all $z^* \in \mathcal{M}_z(x^*)$, there exists $\omega > 0$ such that $p^T H(x^*, z^*)p \geq \omega\|p\|^2$ for all nonzero p satisfying $g(x^*)^T p = 0$ and $J_{\mathcal{A}}(x^*)p \geq 0$.*

The following theorem gives the sufficiency conditions that allow us to verify that a certain point is a strict local constrained minimizer.

Theorem 3.1.3 (Second-order Sufficient Conditions for Strict Minimizer)**.** *The point x^* is a strict local constrained minimizer of problem* (3.1) *if*

1. *x^* is a KKT point,*
2. *SSC1 holds at (x^*, z^*) for some $z^* \in \mathcal{M}_z(x^*)$.* □

Verifying SSC1 for all p that satisfy $g(x^*)^T p = 0$ and $J_{\mathcal{A}}(x^*)p \geq 0$ requires finding the global minimizer of a possibly indefinite quadratic form over a cone, which is an NP-hard problem. As in the QP case, computationally tractable sufficient conditions may be stated by enlarging the set of all p for which $p^T H(x^*, z^*)p$ must be sufficiently positive. This leads to the following more strict result.

Theorem 3.1.4 (Computational Second-order Sufficient Conditions for Strict Minimizer)**.** *The point x^* is a strict local constrained minimizer of problem* (3.1) *if*

1. x^* *is a KKT point,*

2. *For* $z^* \in \mathcal{M}_z(x^*)$ *from part* (1) *and all nonzero* p *satisfying* $g(x^*)^T p = 0$ *and* $J_{\mathcal{A}_+}(x^*)p \geq 0$*, there exists* $\omega > 0$ *such that* $p^T H(x^*, z^*)p \geq \omega \|p\|^2$. □

As strict as these conditions are they are still not sufficient to guarantee that the constrained minimizer is isolated. Constrained qualifications are needed to prove that x^* is an isolated local constrained minimizer of problem (3.1). We also require the more strict SSC2 to holds.

Theorem 3.1.5 (Second-order Sufficient Conditions for Isolated Minimizer)**.** *The point* x^* *is an isolated local constrained minimizer of problem* (3.1) *if*

1. x^* *is a KKT point, i.e.,* $c(x^*) \geq 0$ *and there exists a Lagrange multiplier vector* $z^* \in \mathcal{M}_z(x^*) = \{z : \nabla f(x^*) = J(x^*)^T z,\ z \geq 0,\ c(x^*) \cdot z = 0\}$,

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

3. *the SSC2 holds, i.e., for all* $z^* \in \mathcal{M}_z(x^*)$ *and all* $p \neq 0$ *satisfying* $g(x^*)^T p = 0$ *and* $J_{\mathcal{A}}(x^*)p \geq 0$*, there exists* $\omega > 0$ *such that* $p^T H(x^*, z^*)p \geq \omega \|p\|^2$. □

As in Theorem 3.1.3, verifying condition (3) is NP hard. In addition, confirming that MFCQ holds requires solving a linear program. The following result gives sufficient conditions that are simpler to verify computationally.

Theorem 3.1.6 (Strong Second-order Sufficient Conditions for Isolated Minimizer)**.** *The point* x^* *is an isolated local constrained minimizer of problem* (3.1) *if*

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

2. x^* *is a KKT point and strict complementarity holds, i.e., the unique Lagrange multiplier vector satisfies* $z_i > 0$ *for all* $i \in \mathcal{A}(x^*)$,

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

3.2 The Classical Barrier Function

Classical barrier functions were developed as a way of solving nonlinearly constrained optimization problems by converting them into a sequence of parameterized unconstrained problems. Important features of classical barrier functions are that they preserve the continuity properties of the constraints on the set of strictly feasible points and they are infinite on the boundary of the feasible region. The most common classical barrier function (CBF) is the logarithmic function.

$$B(x; w) = f(x) - \sum_{i=1}^{m} w_i \ln c_i(x),$$

where w_i are positive barrier parameters chosen so that $\|w\| \to 0$. If we let $w_i = \mu$ for all i, then the barrier function takes the form

$$B(x; \mu) = f(x) - \mu \sum_{i=1}^{m} \ln c_i(x).$$

Under appropriate conditions, minimizing $B(x; \mu)$ for a sequence of positive μ converging to zero forces the sequence of unconstrained minimizers of $B(x; \mu)$ to converge to the minimizer of the original problem (3.1).

The following theorem on the convergence of minimizers $x(\mu)$ of the parameterized sequence of barrier functions is stated without proof. The theorem also gives conditions under which the vector $x(\mu)$, when regarded as a function of μ, forms a differentiable path to a constrained minimizer x^*. This path is known as the barrier trajectory, The proof and further discussion can be found, for example, in Forsgren, Gill and Wright [9].

Theorem 3.2.1 (Barrier trajectory)**.** *Assume the set* $\{x : c_i(x) > 0,\ i = 1, \ldots, m\}$ *of strictly feasible points is nonempty. Let* x^* *be a local constrained minimizer for which sufficient optimality conditions stated in Theorem* 3.1.5 *hold. Assume a classical barrier method is applied so that* μ_k *converges monotonically to zero as* $k \to \infty$*. Then*

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

2. *for such a subsequence* $\{x_k\}$*, the sequence of barrier multipliers* $\{z_k\}$ *is bounded, where* $z_k = \mu \,/\!\cdot\, c(x_k)$*,*

3. $\lim_{k\to\infty} z_k = \bar{y}_k \in \mathcal{M}_z(x^*)$.

If, in addition, strict complementarity holds, then

4. $\bar{y}_{\mathcal{A}} > 0$,

5. *for sufficiently large* k, $\nabla^2 B(x_k;\mu_k)$ *is positive definite;*

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

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

8. $\|x_k - x^*\| = \Theta(\mu_k)$. □

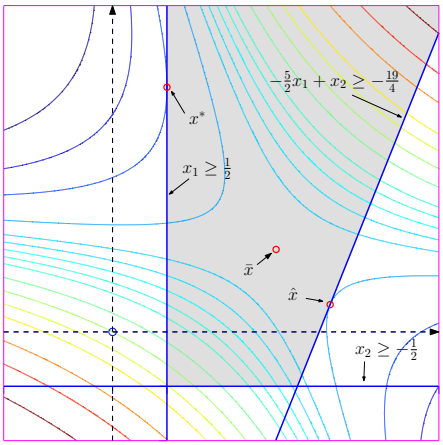
The following example illustrates the properties of the logarithmic barrier function. Consider the QP

$$\begin{array}{ll} \underset{x}{\text{minimize}} & q(x_1,x_2) = \frac{1}{4}x_1^2 + 3x_1x_2 + x_2^2 - 3x_1 - 6x_2 \\ \text{subject to} & -\frac{5}{2}x_1 + x_2 \geq -\frac{19}{4}, \quad x_1 \geq \frac{1}{2}, \quad x_2 \geq -\frac{1}{2}. \end{array} \tag{3.3}$$

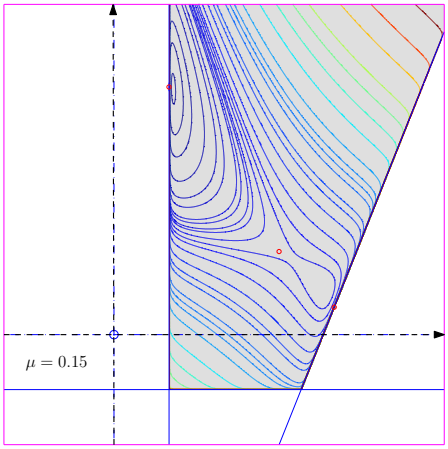
For this problem the Hessian H, constraint matrix A and lower-bound vector b are given by

$$H = \begin{pmatrix} \frac{1}{2} & 3 \\ 3 & 2 \end{pmatrix}, \quad A = \begin{pmatrix} -\frac{5}{2} & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} -\frac{19}{4} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}.$$

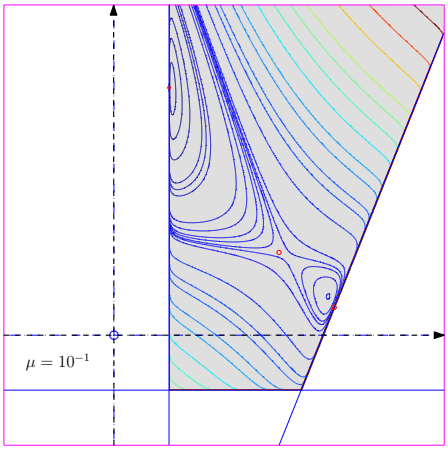
Figure 3.1a shows the level curves of the objective function and the constraint functions. The feasible region is shown in grey. There are three stationary points, two of which are local constrained minimizers. Figures 3.1b–3.1e show the effect of the choice of μ on the level curves of the barrier function. As μ gets smaller the contours start to resemble those of the original objective function, except on the boundary, where the barrier function is infinite and the level curves become



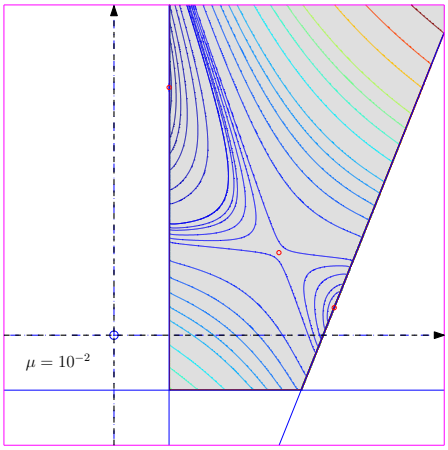
(a) Three Solutions.



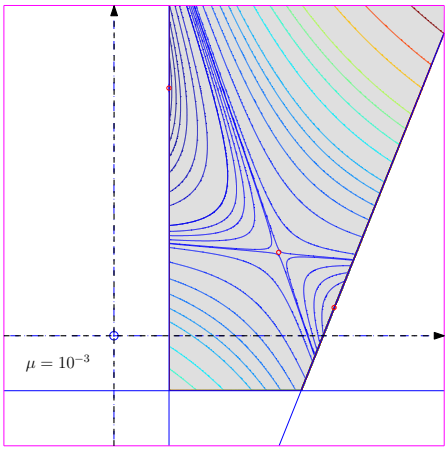
(b) $\mu = .15$.



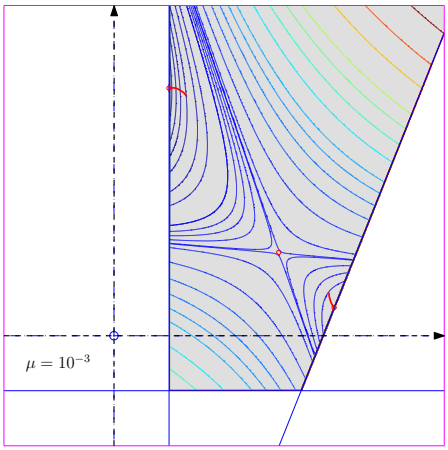
(c) $\mu = .1$.



(d) $\mu = .01$.



(e) $\mu = .001$.



(f) Barrier Trajectory.

Figure 3.1: Properties of Classical Barrier Function.

bunched together. Figure 3.1f shows the trajectories of local unconstrained minimizers of the logarithmic barrier function converging to the two local minimizers x^* and $\widehat{x}$ on the boundary.

The barrier trajectory theorem suggests an algorithm based on computing the minimizers $x(\mu)$ of parameterized barrier functions for a sequence of barrier parameters μ converging to zero. Analogous to the penalty-type methods discussed in Chapter 2.3, this barrier algorithm is a two-tier method with inner and outer iterations. In the inner iterations an approximate minimizer $x(\mu)$ of $B(x, \mu)$ is computed using some variant of Newton's method. In the outer iterations, a test for convergence to a local constrained minimizer of problem (3.1) is performed and parameter μ is reduced if necessary.

The classical Newton's barrier equations are given by

$$\nabla^2 B(x;\mu)\Delta x = -\nabla B(x;\mu),$$

where

$$\nabla B(x;\mu) = \nabla f(x) - \sum_{i=1}^{m} \frac{\mu}{c_i(x)} \nabla c_i(x)$$

and

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

The derivatives of $B(x;\mu)$ may be simplified by defining a vector

$$\pi(x;\mu) = \mu \,./\, c(x) = \mu C(x)^{-1} e. \tag{3.4}$$

This vector, referred to as the vector of primal multipliers, may be interpreted as an estimate of the Lagrange multipliers in the sense that, as the sequence $\{x_k\} \to x(\mu)$, it holds that $\{\pi(x_k;\mu)\} \to z(\mu)$. Expressed in terms of π, the gradient and the Hessian are

$$\nabla B(x;\mu) = \nabla f(x) - J(x)^T \pi(x;\mu)$$

and

$$\nabla^2 B(x;\mu) = H(x, \pi(x;\mu)) + J(x)^T \Pi(x;\mu) C(x)^{-1} J(x).$$

Algorithm 3.2.1 summarizes the classical Newton barrier method. The algorithm uses a backtracking line search. While most common line-search methods

use polynomial local models of the barrier function, such models may be inadequate for modeling the sharp increase of the barrier function that occurs near the boundary.

Both the gradient and the Hessian of the barrier function are undefined on the boundary. Moreover, if the optimal solution lies on the boundary, the Hessian of $B(x;\mu)$ is increasingly ill-conditioned as $\mu \to 0$. This property may be characterized more precisely as follows. Suppose that the barrier Hessian is evaluated at a strictly feasible point that is close to the minimizer x^* and relatively near the the barrier trajectory (i.e., the smallest active constraint value is not too small compared to μ, or more formally $\min_{i\in\mathcal{A}} c(x) = \Omega(\mu)$, so that $\max_{i\in\mathcal{A}} \pi(x;\mu) = O(1)$). At such a point, the barrier Hessian has two widely separated sets of eigenvalues. In the first set the eigenvalues become unbounded as $\mu \to 0$. The second set of eigenvalues are $\Theta(1)$ as $\mu \to 0$. The subspace that corresponds to the large eigenvalues is close to the range of $J_{\mathcal{A}}^T$, while the subspace that corresponds to the small eigenvalues is close to the null space of $J_{\mathcal{A}}$. However, when evaluated at "bad" points, that are not close to the trajectory (i.e., the constrained values are very small compared to μ) the Hessian becomes arbitrarily ill-conditioned. In practice, ill-conditioning is not as harmful as one might think (see, e.g., Wright [27]). When computed at "nice" points using backward stable methods and assuming error cancelation, the search directions are relatively accurate because of the special structure of the Hessian at these points and the fact that the gradient lies almost entirely in the range space of $J_{\mathcal{A}}^T$. This result does not hold for "bad" points, thus for the barrier method to be effective, the solution must be approached along the barrier trajectory.

However, using the classical Newton barrier method is still inefficient, in large part because of the properties of the exact Newton step (see, e.g., Wright [26]). Immediately following the reduction of the barrier parameter μ, the Newton step Δx violates the constraints. In particular, when the current iterate x is close to the solution x^* and lies near the trajectory (i.e., $\max_{i\in\mathcal{A}} \pi(x;\mu) = O(1)$), then reducing the barrier parameter to $\bar{\mu}$ by a large enough factor may cause the next x to be far from the trajectory (i.e, $\max_{i\in\mathcal{A}} \pi(x;\mu) = \bar{\mu}/\mu$).

Algorithm 3.2.1 A Classical Newton's Barrier Method

INPUT: $x_0 \in \{x : c(x) > 0\}$, μ_0;

Set constants $\gamma \in (0,1)$, $0 < \tau \ll 1$;

$k = 0$;

while not *converged* **do**

$j = 1$;

$x_j \leftarrow x_k$;

while $\|\nabla B(x_j; \mu_k\|) > \tau$ **do**

Compute $B(x_j; \mu_k)$, $\nabla B(x_j; \mu_k)$, $\nabla^2 B(x_j; \mu_k)$;

Compute E such that $(\nabla^2 B(x_j; \mu_k) + E)$ is p.d.;

Solve $(\nabla^2 B(x_j; \mu_k) + E)\Delta x_j = -\nabla B(x_j; \mu_k)$;

Perform backtracking line search to find α_j;

Update $x_{j+1} = x_j + \alpha_j \Delta x_j$;

$j \leftarrow j + 1$;

end while

$x_{k+1} \leftarrow x_j$;

if optimality conditions hold **then**

$converged \leftarrow true$

else

$\mu_{k+1} = \gamma \mu_k$;

end if

$k \leftarrow k + 1$;

end while

OUTPUT: x_k

Various improvements for the classical Newton's barrier method exist that address the ill-conditioning of the Hessian as $\mu \to 0$ as well as inefficiencies of the exact Newton step each time μ is reduced. Specialized line-search methods suitable for barrier functions have also been proposed. On the other hand, the method presented in the next section completely eliminates the need to reduce μ beyond a certain threshold.

3.3 The Modified Barrier Function

The modified barrier function (MBF) was originally developed to avoid the need for μ to go to zero, keeping MBF and its derivatives well-defined on the boundary. Most of the discussion and theorems in this section are based on the work of Polyak [22].

The idea behind the MBF comes from noting that, for a fixed parameter μ, the feasibility region $\mathcal{F}$ of the original nonlinear program (3.1) is identical to $\{x \in \mathbb{R}^n : \mu \ln(c_i(x)/\mu + 1) \geq 0,\ i = 1, \ldots, m\}$. Moreover, any KKT point of the original problem is also a KKT point of the modified problem

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & \mu \ln\Big(\dfrac{c_i(x)}{\mu} + 1\Big) \geq 0, \quad i = 1, \ldots, m, \end{array} \tag{3.5}$$

where the multipliers $\widehat{z}_i^*$ for problem (3.5) are given by

$$\widehat{z}_i^* = \Big(\frac{c_i(x^*)}{\mu} + 1\Big) z_i^*,$$

where z_i^* are multipliers associated with the KKT point x^* of the original problem.

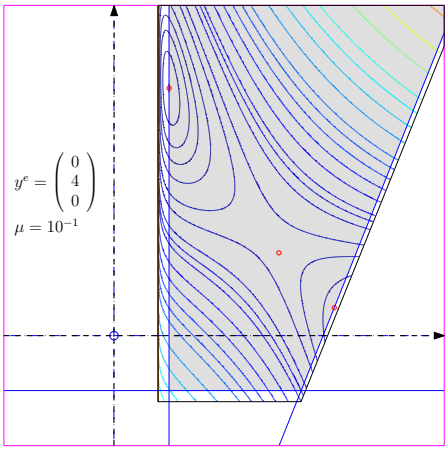
The modified barrier function is then defined as

$$B(x; \mu, z) = f(x) - \mu \sum_{i=1}^{m} z_i \ln\Big(\frac{c_i(x)}{\mu} + 1\Big), \tag{3.6}$$

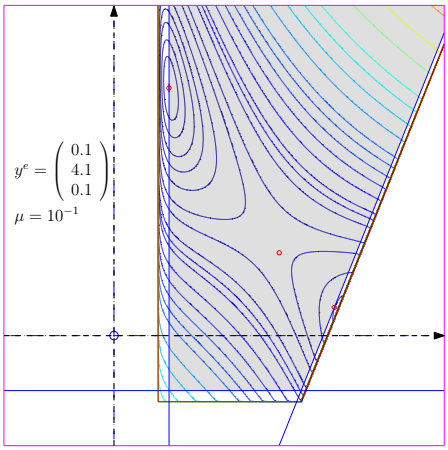
which may be viewed as the classical Lagrangian for the modified problem (3.5).

For an alternative motivation for the modified barrier function, consider a shifted problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & c(x) + \mu e \geq 0. \end{array} \tag{3.7}$$



(a) Optimal multipliers at x^*, $\mu = 0.1$



(b) Slightly perturbed multipliers at x^*, $\mu = 0.1$.

Figure 3.2: Properties of Modified Barrier Function.

The weighted classical barrier function for this problem is

$$B(x; w) = f(x) - \sum_{i=1}^{m} w_i \ln(c_i(x) + \mu).$$

Let $w_i = \mu z_i^a$, where z_i^a are multiplier estimates, then the weighted barrier function for (3.7) is given by

$$B(x; w) = f(x) - \mu \sum_{i=1}^{m} z_i^a \ln(c_i(x) + \mu)$$

and differs from the modified barrier function (3.6) by a constant.

Consider Example 3.3 from the previous section. Figures 3.2a–3.2b show that the modified barrier function is well-defined at all feasible points, including the boundary of $\mathcal{F}$. Moreover, the function has the same smoothness as the objective and constraint functions in the neighborhood of the boundary of the feasible set $\mathcal{F}$. It can also be seen that given the exact z^* and μ sufficiently small, x^* is a minimizer of the modified barrier function. It is clear that μ does not need to go to zero for the unconstrained minimizers to converge to the solution.

Result 3.3.1 (Properties of the MBF)**.** *Consider the modified barrier function* $B(x; \mu, z^a)$ *defined in* (3.6) *and assume that* x^* *is a local minimizer of the nonlinear*

program (3.1) *for which the sufficient conditions of Theorem* 3.1.6 *hold. Then the following properties follow*

1. $B(x^*; \mu, z^*) = f(x^*)$ *for all* $\mu > 0$,
2. *there exists a* $\bar{\mu} > 0$ *such that for all* $\mu \leq \bar{\mu}$, x^* *satisfies second-order sufficiency conditions for an unconstrained minimizer of* $B(x; \mu, z^*)$.

Proof. The sufficient conditions of Theorem 3.1.6 imply that x^* is a KKT point with unique Lagrange multiplier z^* with $z_i^* > 0$ for all $i \in \mathcal{A}$ and zero otherwise. It follows that

$$B(x; \mu, z^*) = f(x) - \mu \sum_{i \in \mathcal{A}} z_i^* \ln\left(\frac{1}{\mu} c_i(x) + 1\right)$$

and

$$\nabla B(x; \mu, z^*) = \nabla f(x) - \mu \sum_{i \in \mathcal{A}} \frac{z_i^*}{c_i(x) + \mu} \nabla c_i(x).$$

Similarly,

$$\begin{aligned}
\nabla^2 B(x; \mu, z^*) = \nabla^2 f(x) &- \mu \sum_{i \in \mathcal{A}} \frac{z_i^*}{c_i(x) + \mu} \nabla^2 c_i(x) \\
&+ \mu \sum_{i \in \mathcal{A}} \frac{z_i^*}{(c_i(x) + \mu)^2} \nabla c_i(x)^T \nabla c_i(x) \\
&= H(x, \mu(C_{\mathcal{A}}(x) + \mu I)^{-1} z_{\mathcal{A}}^*) + \mu J_{\mathcal{A}}^T(x)(C_{\mathcal{A}}(x) + \mu I)^{-2} Z_{\mathcal{A}}^* J_{\mathcal{A}}(x).
\end{aligned}$$

To verify part (1), note that because $c_i(x) = 0$ for all $i \in \mathcal{A}$, it must hold that

$$\ln\left(\frac{1}{\mu} c_i(x^*) + 1\right) = 0,$$

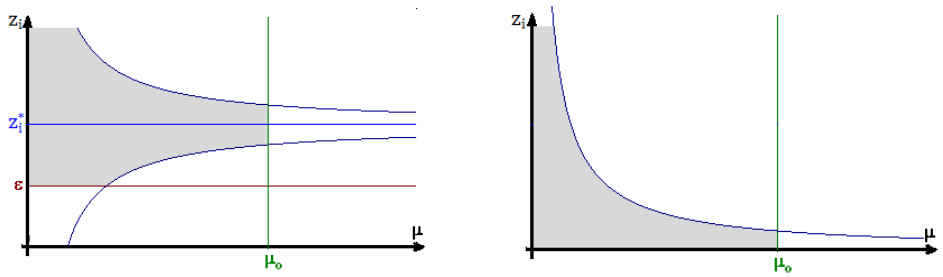
and hence $B(x^*; \mu, z^*) = f(x^*)$.

For part (2) it must be shown that that $\nabla B(x^*; \mu, z^*) = 0$ and $\nabla^2 B(x^*; \mu, z^*)$ is positive definite for all $\mu \leq \bar{\mu}$. As x^* is a KKT point, stationarity at (x^*, z^*) must hold and thus

$$\nabla B(x^*; \mu, z^*) = \nabla L(x^*, z^*) = 0.$$

Evaluated at x^*,

$$\nabla^2 B(x^*; \mu, z^*) = H(x^*, z_{\mathcal{A}}^*) + \frac{1}{\mu} J_{\mathcal{A}}^T(x^*) Z_{\mathcal{A}}^* J_{\mathcal{A}}(x^*).$$



(a) $i \in \mathcal{A}$ (b) $i \notin \mathcal{A}$

Figure 3.3: Graphical Representation of the Set $\mathcal{D}_i$.

As the second-order condition of Theorem 3.1.6 requires the existence of $\omega > 0$ such that $p^T H(x^*, z^*)p \geq \omega\|p\|^2$ for all nonzero p satisfying $J_{\mathcal{A}}(x^*)p = 0$, the variation of Debreu's lemma (see Theorem 1.5.4) implies that there must exist a $\bar{\mu} > 0$ such that for all $\mu \leq \bar{\mu}$, $\nabla^2 B(x^*; \mu, z^*)$ is positive definite. □

The main global convergence result for modified barrier functions involves the definition of a set $\mathcal{D}$ of "acceptable" multiplier estimates and barrier parameters. The result requires the assumption of strict complementarity. Let $\delta > 0$ and $0 < \epsilon < \min_{i\in\mathcal{A}} \ z_i^*$. Define $\mathcal{D}_i = \{z_i : \ z_i \geq \epsilon, \ |z_i - z_i^*| \leq \delta/\mu, \ \mu \leq \mu_0\}$ for $i \in \mathcal{A}$ and $\mathcal{D}_i = \{z_i : \ 0 \leq z_i \leq \delta/\mu, \ \mu \leq \mu_0\}$ for $i \notin \mathcal{A}$, (see Figure 3.3). Then define

$$\mathcal{D}(\mu_0, z^*, \delta, \epsilon) = \mathcal{D}_1 \otimes \cdots \otimes \mathcal{D}_r \otimes \cdots \otimes \mathcal{D}_m. \tag{3.8}$$

It is also necessary to impose a certain "niceness" condition on the shifted feasible set

$$\mathcal{F}_\mu = \{x : c_i(x) + \mu \geq 0, i = 1, \ldots, m\}$$

Observe that $\mathcal{F} \subset \mathcal{F}_\mu$. If problem (3.1) is convex, i.e., $c_i(x)$ are all concave, then compactness of $\mathcal{F}$ implies compactness of $\mathcal{F}_\mu$. If the nonlinear program is not convex, no such implication holds and the following growth condition is required:

$$\exists \mu_0 > 0 \ \text{ and } \ \tau > 0 : \quad \max \left\{ \max_{1\leq i\leq m} c_i(x) : x \in \mathcal{F}_{\mu_0} \right\} = \theta(\mu_0) \leq \tau. \tag{3.9}$$

It is clear that $\theta(\mu)$ is a decreasing function of μ. Thus if the growth condition holds for some μ_0, then $\theta(\mu) \leq \tau$ for all $\mu \leq \mu_0$. We are now ready to state the main theorem for modified barrier functions.

Theorem 3.3.1 (Local MBF convergence)**.** *Assume the set of strictly feasible points* $\{x : c_i(x) > 0, \quad i = 1, \ldots, m\}$ *is nonempty. Let* x^* *be a local constrained minimizer for which the sufficient optimality conditions stated in Theorem* 3.1.6 *hold. In addition, assume that the growth condition* (3.9) *holds. Then there exist* μ_0 *and a sufficiently small* $\delta > 0$ *such that for any* $0 < \epsilon < \min_{i \in \mathcal{A}} \ z_i^*$ *and any* (μ, z^a) *in* $\mathcal{D}(z^*, \mu_0, \delta, \epsilon)$*, the following statements hold:*

1. *there exists a vector* $\widehat{x} = \widehat{x}(\mu, z^a) = \operatorname{argmin} \{B(x; \mu, z^a) : x \in \mathbb{R}^n\}$ *such that* $\nabla B(\widehat{x}; \mu, z^a) = 0$*,*

2. *for a pair of vectors* $\widehat{x}$ *and* $\widehat{\pi} = \widehat{\pi}(\mu, z^a) = \mu(C(\widehat{x}(\mu, z^a)) + \mu I)^{-1} z^a$*, the estimate*

$$\max\{\|\widehat{x} - x^*\|, \|\widehat{\pi} - z^*\|\} \leq K\mu \|z^a - z^*\|$$

   holds, with constant K *independent of* μ*,*

3. $\widehat{x}(\mu, z^*) = x^*$ *and* $\widehat{\pi}(\mu, z^*) = z^*$*,*

4. $B(x; \mu, z^a)$ *is strongly convex in a neighborhood of* $\widehat{x} = \widehat{x}(\mu, z^a)$*.* □

In order to apply the results of Theorem 3.3.1 it is necessary to know the quantities $(\mu, z^a) \in \mathcal{D}(\mu, z^*, \delta, \epsilon)$. The following result shows how such a pair may be found.

Theorem 3.3.2 (Global MBF Theorem)**.** *Assume the set of strictly feasible points* $\{x : c_i(x) > 0, \quad i = 1, \ldots, m\}$ *is nonempty. Let* x^* *be a local constrained minimizer for which the sufficient optimality conditions of Theorem* 3.1.6 *hold. In addition, assume that the the growth condition* (3.9) *holds and that there exists a positive* μ_0 *such that* $\mathcal{F}_{\mu_0}$ *is compact. Then*

1. *for all* $\mu \leq \mu_0$ *there exists* $x(\mu) = \operatorname{argmin} \{B(x; \mu, e) : x \in \mathbb{R}^n\}$ *such that* $\nabla B(x(\mu); \mu, e) = 0$ *and* $\lim_{\mu \to 0} f(x(\mu)) = \lim_{\mu \to 0} B(x(\mu); \mu, e) = f(x^*)$*,*

2. *there exists a positive* $\bar{\mu}$ *such that* $x(\mu)$ *exists for all* $\mu \leq \bar{\mu}$*, and the estimate*

$$\max\left\{\|x(\mu) - x^*\|, \, \|\pi(\mu) - z^*\|\right\} \leq K\mu$$

   holds for some $K > 0$ *independent of* μ*, with* $\pi(\mu) = \mu(C(x(\mu)) + \mu I)^{-1} e$*.*

3. *$B(x;\mu,e)$ is strongly convex in the neighborhood of $x(\mu)$.* □

Making use of the theory above, we can now develop a modified barrier method for finding constrained minimizers of the inequality-constraint nonlinear program. In order to begin, it is necessary to find a strictly feasible point x_0. This may be done, for example, by solving argmax $\{\min_i c_i(x)\}$.

If the threshold for $\bar{\mu}$ is known *a priori*, then Theorem 3.3.2 suggests a simple procedure such that, given a primal-dual iterate (x_k, z_k^a), the next iterate (x_{k+1}, z_{k+1}^a) is defined by setting $x_{k+1} = \text{argmin}\{B(x;\bar{\mu}, z_k^a)\}$ and updating the multiplier estimates as $z_{k+1}^a = \mu(C(x_{k+1}) + \mu I)^{-1} z_k^a$. In general, $\bar{\mu}$ is not known in advance, which suggests the possibility of reducing μ to achieve convergence to the solution of the nonlinear program (3.1). As a consequence, once (x_{k+1}, z_{k+1}^a) is found, it may be necessary to define μ_{k+1} such that $\mu_{k+1} < \mu_k$. However, in this case, x_{k+1} may not be in the interior of $\mathcal{F}_{\mu_{k+1}}$, which implies that the modified barrier function may not be well-defined. In this case x_{k+1} is reset to a known strictly feasible point, or some iterative procedure is used to determine a feasible point for the new set $\mathcal{F}_{\mu_{k+1}}$.

As with the classical barrier method, each modified barrier function is minimized approximately using a form of Newton's method. The Newton directions are given by the solution of the Newton equations $\nabla^2 B(x;\mu,z^a)\Delta x = -\nabla B(x;\mu,z^a)$, where

$$\nabla B(x;\mu,z^a) = \nabla f(x) - J(x)^T (C(x) + \mu I)^{-1} z_i^a$$

and

$$\nabla^2 B(x;\mu,z^a) = H(x, \mu(C(x) + \mu I)^{-1} z^a) + \mu J(x)^T (C(x) + \mu I)^{-2} Z^a J(x).$$

Given the vector of primal multipliers $\pi(x) = \mu(C(x) + \mu I)^{-1} z^a$, the derivatives may be simplified to give

$$\nabla B(x;\mu,z^a) = \nabla f(x) - J(x)^T \pi(x)$$

and

$$\nabla^2 B(x;\mu,z^a) = H(x,\pi(x)) + J(x)^T (C(x) + \mu I)^{-1} \Pi(x) J(x).$$

Algorithm 3.3.1 A Newton's Modified Barrier Method

INPUT: $x_0 \in \mathcal{F}^+_{\mu 0}$, μ_0, z^a_0;

Set constants $\gamma \in (0, .5]$, $0 < \tau \ll 1$;

$k = 0$, $d = 1$;

while not *converged* **do**

$j = 1$; $x_j \leftarrow x_k$;

while not min_found **do**

Compute $B(x_j; \mu_k, z^a_k)$, $\nabla B(x_j; \mu_k, z^a_k)$, $\nabla^2 B(x_j; \mu_k, z^a_k)$;

Solve $(\nabla^2 B(x_j; \mu_k, z^a_k) + E)\Delta x_j = -B(x_j; \mu_k, z^a_k)$;

Perform backtracking line search to find α_j;

Update $x_{j+1} = x_j + \alpha_j \Delta x_j$;

if $\alpha_j = 1$ **and** $\|\Delta x_j\| \leq \tau$ **then**

$min\_found \leftarrow true$;

if $\nu(x_{j+1}, \pi(x_{j+1}; \mu_k, z^a_k), \mu_k) \leq \gamma^d$ **then**

Update $x_{k+1} = x_{j+1}$; $z^a_{k+1} = \pi(x_{j+1}; \mu_k, z^a_k)$;

$d \leftarrow d + 1$; $\tau \leftarrow \tau\gamma$;

else

Set $\mu_{k+1} < \mu_k$; $\tau \leftarrow \mu\tau$; $d \leftarrow 1$;

Reset $z^a_{k+1} \leftarrow z^a_0$;

Adjust $x_{k+1} \in \mathcal{F}_{\mu_{k+1}}$;

end if

end if

$j \leftarrow j + 1$;

end while

if optimality conditions hold **then**

$converged \leftarrow true$;

end if

$k \leftarrow k + 1$;

end while

OUTPUT: x_k

Algorithm 3.3.1 gives the Newton modified barrier method proposed in [22]. The algorithm starts with an arbitrary parameter μ_0 and $z^a = e$, but requires an initial iterate $x_0 \in \mathcal{F}^+_{\mu_0} = \{x : c_i(x) > \mu_0,\ i = 1,\ \ldots,\ m\} \subset \mathcal{F}$. The test for optimality involves the quantity

$$\nu(x, z^a, \mu) = \max\Big\{ -\min_{1\le i\le m} c_i(x), \|\nabla B(x;\mu, z^a)\|, \sum_{i=1}^{m} z_i^a |c_i(x)| \Big\}.$$

It should be emphasized that the barrier parameter μ in the MBF may be viewed as a regularization parameter for the constraint functions $c_i(x)$.

3.4 A Primal-Dual Barrier Function

Primal-dual interior methods are another class of methods that attempt to alleviate the difficulties associated with the classical barrier method. The methods are based on the properties of the barrier trajectory $x(\mu)$. In particular, note that the gradient of the classical barrier function vanishes at $x(\mu)$, i.e., $\nabla B(x(\mu);\mu) = 0$. Substituting the definition of $\pi(x(\mu);\mu)$ in the identity $\nabla B(x(\mu);\mu) = 0$ yields

$$\begin{aligned} g(x(\mu)) - J(x(\mu))^T z(\mu) &= 0, \\ z(\mu) - \mu C(x(\mu))^{-1} e &= 0, \end{aligned}$$

where the vector $z(\mu) = \pi(x(\mu);\mu)$ may be interpreted as a vector of approximate multiplies. Unlike primal methods, which find primal minimizers $x(\mu)$ and compute $z(\mu)$ as a by-product of the main computation, primal-dual methods treat z as independent variables and define a pair (x, z) that satisfies the primal-dual equations that hold at $x(\mu)$. That is, they seek a feasible solution $(x(\mu), z(\mu))$ (i.e., $c(x(\mu)) > 0$ and $z(\mu) > 0$) that satisfy the system

$$\begin{aligned} g(x) - J(x)^T z &= 0, \\ C(x)z - \mu e &= 0. \end{aligned}$$

These equations corresponds to perturbed first-order optimality conditions of the original problem (3.1), in the sense that, as $\mu \to 0$, the equations $c(x(\mu)) \cdot z(\mu) =$

μe define an increasingly accurate approximation of the complementarity condition $c(x^*) \cdot z^* = 0$.

It is convenient to describe the equations above in terms of finding zeros of the vector-valued function

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

This allows that application of Newton's method to solve the equations $F(x, z; \mu) = 0$ via Newton's method. The Newton equations for a primal-dual search direction $\Delta v = (\Delta x, \Delta z)$ are then $\nabla F(x, z; \mu)\Delta v = -F(x, z; \mu)$, which correspond to the primal-dual equations

$$\begin{pmatrix} H(x, z) & -J(x)^T \\ ZJ(x) & C(x) \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta z \end{pmatrix} = - \begin{pmatrix} g(x) - J(x)^T z \\ C(x)(z - \pi(x; \mu)) \end{pmatrix}. \tag{3.10}$$

The success of primal-dual methods is due, in part, to their effectiveness at following the barrier trajectory. In particular, the primal-dual direction $(\Delta x, \Delta z)$ computed at a point $(x, z) = (x(\mu), z(\mu))$ on the trajectory will provide a good approximation of the step towards $(x(\bar{\mu}), z(\bar{\mu}))$, the point on the trajectory corresponding to $\mu = \bar{\mu}$. This property does not hold for the the classical barrier method.

An equivalent symmetric primal-dual system can be derived by multiplying the last m rows of (3.10) by Z^{-1} and changing the sign of Δz. This gives

$$\begin{pmatrix} H(x, z) & J(x)^T \\ J(x) & Z^{-1}C(x) \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta z \end{pmatrix} = - \begin{pmatrix} g(x) - J(x)^T z \\ Z^{-1}C(x)(z - \pi(x; \mu)) \end{pmatrix}. \tag{3.11}$$

For nonconvex problems, the symmetric form is preferable because it allows the use of the symmetric indefinite factorization (see Section 2.4.1), and other methods that rely on symmetry. These methods may be used to compute an appropriate descent direction and check the second-order optimality conditions, which require knowledge of the inertia of the Hessian $H(x, z)$ in the subspace orthogonal to the active constraint gradients.

It is important to note that as $\mu \to 0$, the diagonals of Z^{-1} corresponding to the inactive constraints grow without bound, and the system (3.11) becomes

increasingly ill-conditioned. However, it has been shown (see Forsgren, Gill, and Shinnerl [8]) that this ill-conditioning is benign, and accurate primal-dual solutions can be found using symmetric indefinite factorization of the type discussed in Section 2.4.2.

To ensure global convergence of primal-dual methods from an arbitrary starting point, it is necessary to use the primal-dual direction Δv defined above in conjunction with an appropriate merit function. For nonconvex programs, the choice of an appropriate merit function is crucial. Here we consider a merit function proposed by Forsgren and Gill [7].

The Forsgren-Gill merit function for inequality constraints can be viewed as the classical barrier function augmented by a function that penalizes any deviation from the barrier trajectory $(x(\mu), z(\mu))$. Given a positive μ, the function is given by

$$B^{\nu}(x, z; \mu) = f(x) - \mu \sum_{i=1}^{m} \Big(\ln(c_i(x)) + \nu \Big(\ln \Big(\frac{c_i(x) z_i}{\mu} \Big) + 1 - \frac{c_i(x) z_i}{\mu} \Big) \Big). \quad (3.12)$$

This function is well defined for all (x, z) such that $c(x) > 0$ and $z > 0$.

The following lemma summarizes the relationship between points that minimize $B^{\nu}(x, z; \mu)$ and points that lie on the primal-dual barrier trajectory.

Lemma 3.4.1. *Let $B^{\nu}(x, z; \mu)$ be the barrier function* (3.12) *defined with any positive ν and μ. A point (x, z) such that $c(x) > 0$ and $z > 0$ is an unconstrained local minimizer of $B^{\nu}(x, z; \mu)$ if and only if $(x, z) = (x(\mu), z(\mu))$. Furthermore, $\min_z B^{\nu}(x, z; \mu) = B^{\nu}(x, \pi(x, \mu); \mu) = B^{\nu}(x; \mu)$, where $\pi(x; \mu)$ is as defined in* (3.4) *and $B^{\nu}(x; \mu)$ is the classical barrier function* (3.2). □

This result suggests that points on the barrier trajectory $(x(\mu), z(\mu))$ may be found by minimizing $B^{\nu}(x, z; \mu)$ with respect to both x and z. In order to guarantee convergence to a local minimizer of a nonconvex problems, both first and second derivatives are required. Differentiating with respect to both x and z and making use of the vector $\pi(x; \mu)$ yields

$$\nabla B^{\nu}(x, z; \mu) = \begin{pmatrix} g(x) - (1 + \nu) J(x)^T \pi(x; \mu) + \nu J(x)^T z \\ \nu Z^{-1} C(x) (z - \pi(x; \mu)) \end{pmatrix}$$

and

$$\nabla^2 B^\nu(x,z;\mu) = \begin{pmatrix} \nabla^2_{xx} B^\nu(x,z;\mu) & \nu J(x)^T \\ \nu J(x) & \nu\mu Z^{-2} \end{pmatrix} = \begin{pmatrix} \nabla^2_{xx} B^\nu(x,z;\mu) & \nu J(x)^T \\ \nu J(x) & \nu C(x)\Pi(x;\mu) Z^{-2} \end{pmatrix},$$

where $\nabla^2_{xx} B^\nu(x,z)$ is given by

$$\nabla^2_{xx} B^\nu(x,z) = H(x,(1+\nu)\pi(x;\mu) - \nu z) + (1+\nu) J(x)^T C(x)^{-1} \Pi(x;\mu) J(x).$$

For the minimization of the merit function $B^\nu(x,z;\mu)$, instead of using a conventional Newton method, which involves the equations

$$\nabla^2 B^\nu(x,z;\mu) \begin{pmatrix} \Delta x \\ \Delta z \end{pmatrix} = -\nabla B^\nu(x,z;\mu),$$

a modified Newton method is used. This method is defined by approximating the primal-dual Hessian $\nabla^2 B^\nu(x,z;\mu)$ by an approximate Hessian $S^\nu(x,z;\mu)$. This approximate Hessian is defined by replacing the approximate multipliers $\pi(x;\mu)$ by z in the matrix defining the Newton equations, i.e.,

$$S^\nu(x,z;\mu) = \begin{pmatrix} H(x,z) + (1+\nu) Z C(x)^{-1} & \nu J(x)^T \\ \nu J(x) & \nu Z^{-1} C(x) \end{pmatrix} \tag{3.13}$$

The following theorem explores the relationship between the directions defined by the modified Newton equations

$$S^\nu(x,z;\mu) \begin{pmatrix} \Delta x \\ \Delta z \end{pmatrix} = -\nabla B^\nu(x,z;\mu) \tag{3.14}$$

and the directions generated by the symmetric primal-dual equations (3.11). The result also provides a crucial property of $S^\nu(x,z;\mu)$ at stationary points of $B^\nu(x,z;\mu)$.

Theorem 3.4.1. *Assume that the barrier function $B^\nu(x,z;\mu)$ of (3.12) is defined with any positive ν and μ. Let $S^\nu(x,z;\mu)$ denote the approximate Hessian (3.13). If $c(x) > 0$ and $z > 0$, then the vector $(\Delta x, \Delta z)$ solves the modified Newton equations (3.14) if, and only if, it solves the primal-dual equations (3.10) and the symmetric primal-dual equations (3.11). Moreover, if $\nabla B^\nu(x,z;\mu) = 0$, then $S^\nu(x,z;\mu) = \nabla^2 B^\nu(x,z;\mu)$.* □

The proof, which is omitted here, is based on establishing the equivalence of the modified Newton equations (3.14) and the symmetric primal-dual equations (3.11), by premultiplying (3.11) by the nonsingular matrix

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

The proposed primal-dual method generates a sequence $\{v_k\}$ of improving estimates of a local minimizer of B^ν, by computing a search direction $\Delta v_k = (\Delta x_k, \Delta z_k)$ from the modified Newton system (3.14) and performing a line search to ensure sufficient reduction in B^ν. The search direction is computed as $\Delta v_k = s_k + d_k$, where s_k is a descent direction with respect to $B^\nu(v_k;\mu)$ and d_k is a direction of negative for the modified Hessian $S^\nu(v_k;\mu)$.

First, we state the algorithm for finding a local minimizer of $B^\nu(v;\mu)$. This is followed by a description of how the search directions are computed, and a discussion of the conditions needed to establish convergence to a second-order local minimizer.

Algorithm 3.4.1 Modified Newton Method for minimizing $B^\nu(v_k;\mu)$

INPUT: $v_0 \in \mathcal{F}$, μ, ν;

Set $\eta \in (0, .5)$;

$k = 0$;

while not *converged* **do**

Compute $B^\nu(v_k;\mu)$, $\nabla B^\nu(v_k;\mu)$, $\nabla^2 B^\nu(v_k;\mu)$, $S^\nu(v_k;\mu)$;

if $S^\nu(v_k;\mu)$ is positive definite and $\nabla B^\nu(v_k;\mu) = 0$ **then**

$converged \leftarrow true$; break;

else if $S^\nu(v_k;\mu)$ is positive definite **then**

Compute s_k; $d_k \leftarrow 0$;

else

Compute s_k and d_k;

end if

$\Delta v_k = s_k + d_k$;

$\alpha_k \leftarrow 1$;

---

while $B^\nu(v_k + \alpha_k \Delta v_k; \mu) > B^\nu(v_k; \mu) + \eta(\alpha_k \nabla B^\nu(v_k; \mu)^T \Delta v_k \ldots$
$+\frac{1}{2}\alpha^2 \min(0, \Delta v_k^T \nabla^2 B^\nu(v_k; \mu) \Delta v_k))$ **do**
$\alpha \leftarrow \alpha/2$;
end while
$v_{k+1} = v_k + \alpha_k \Delta v_k$;
$k \leftarrow k + 1$;
end while
OUTPUT: x_k;

As stated in Theorem 3.4.1, solving the modified Newton system (3.14) is equivalent to solving the symmetric primal-dual equations (3.11). Therefore, the search directions may be defined using the symmetric equations with the corresponding primal-dual matrix

$$K = \begin{pmatrix} H & J^T \\ J & -D \end{pmatrix}, \tag{3.15}$$

where $D(x, z) = Z^{-1}C(x)$. The matrix S^ν is used to measure the approximate curvature of the merit function, which requires that the inertia of S^ν must be available from the inertia of K. The following theorem gives the required relationship.

Theorem 3.4.2. *Let K be as defined in* (3.15) *with H a symmetric matrix and D a symmetric positive-definite matrix. Similarly, let S^ν be the approximate Hessian* (3.13)*, with $\nu > 0$. Then*

$$\begin{aligned} \text{In}(S^\nu) &= \text{In}(H + J^T D^{-1} J) + (m, 0, 0), \\ \text{In}(K) &= \text{In}(H + J^T D^{-1} J) + (0, m, 0). \end{aligned}$$

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

$$\lambda_{\min}(H + J^T D^{-1} J) \leq \lambda_{\min}(S^\nu). \quad \square$$

The search directions are computed using the inertia-controlling symmetric indefinite factorization of K discussed in Chapter 2.4.2. Given such a factorization,

the descent direction s is computed using a modification that guarantees the correct inertia of the modified matrix $\bar{K}$ (see Chapter 2.4.3), i.e.,

$$\begin{pmatrix} \bar{H} & J^T \\ J & -D \end{pmatrix} \begin{pmatrix} s_x \\ -s_z \end{pmatrix} = - \begin{pmatrix} g - J^T z \\ D(z - \pi) \end{pmatrix}, \tag{3.16}$$

where $\bar{H}$ is a symmetric modification such that $\bar{H} + J^T D^{-1} J$ is sufficiently positive definite. The following lemma establishes that the computed vector s is a descent direction with respect to $B^\nu(x, z; \mu)$.

Lemma 3.4.2. *Let s satisfy* (3.16). *Then*

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

Moreover, $s_x = 0$ if and only if $g - J^T \pi = 0$. Finally, if $s_x = 0$, then $s_z = \pi - z$. □

The direction of negative curvature is also computed using the inertia-controlling LBL^T. In this case

$$\begin{pmatrix} d_x \\ -d_z \end{pmatrix} = P \widetilde{d}, \quad \text{where} \quad \begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22}^T \end{pmatrix} \widetilde{d} = \pm \sigma \begin{pmatrix} 0 \\ u \end{pmatrix}. \tag{3.17}$$

In this definition, $\sigma = \sqrt{-\lambda_{\min}(B_{22})}$, and u is the normalized eigenvector associated with $\lambda_{\min}(B_{22})$. The sign of σ in (3.17) is chosen to give $d^T \nabla B^\nu(v_k; \mu) \leq 0$. The following lemma establishes that d is a direction of negative curvature for S^ν.

Lemma 3.4.3. *Let d satisfy* (3.17). *Then $J d_x + D d_z = 0$ and $d^T S^\nu(x, z; \mu) d = -\lambda_{\min}(B_{22})^2$. Furthermore, as $J d_x + D d_z = 0$, it holds that*

$$\begin{aligned} d_x^T (H + J^T D^{-1} J) d_x &= \begin{pmatrix} d_x^T & -d_z^T \end{pmatrix} K \begin{pmatrix} d_x \\ -d_z \end{pmatrix} \\ &= \begin{pmatrix} d_x^T & d_z^T \end{pmatrix} S^\nu \begin{pmatrix} d_x \\ d_z \end{pmatrix}. \quad \square \end{aligned}$$

We now turn to a brief discussion of convergence of the proposed algorithm. The principal role of the line search performed on each search direction Δv_k is to ensure that $\nabla B^\nu(v_k; \mu)^T s_k \to 0$ and $d_k^T \nabla^2 B^\nu(v_k; \mu) d_k \to 0$. If these conditions are satisfied and the directions s_k and d_k are directions of sufficient descent

and sufficient negative curvature respectively (see Definitions 1.4.8 and 1.4.10), then every limit point of the sequence $\{v_k\}$ will satisfy the second-order necessary conditions for optimality (see Moré and Sorensen [19]). Algorithm 3.4.1 imposes negative curvature conditions on $d_k^T S^\nu(v_k;\mu)d_k$ rather than on $d_k^T \nabla^2 B^\nu(v_k;\mu)d_k$. However, properties of S^ν ensure that $d_k^T S^\nu(v_k;\mu)d_k \to d_k^T \nabla^2 B^\nu(v_k;\mu)d_k$, whenever $\nabla B^\nu(v_k;\mu) \to 0$. It remains to show that directions s_k and d_k, computed as described above, are indeed sufficient. The following two theorems give the necessary results.

Theorem 3.4.3. *Assume that the sequence of directions $\{s_k\}$ satisfies* (3.16). *Furthermore assume that*

i. $\limsup_{k\to\infty} \|\bar{H}\| < \infty$,

ii. $\limsup_{k\to\infty} \|J_k\| < \infty$,

iii. $\liminf_{k\to\infty} \lambda_{\min}(D_k) > 0$,

iv. $\limsup_{k\to\infty} \lambda_{max}(D_k) < \infty$, *and*

v. $\liminf_{k\to\infty} \lambda_{\min}(\bar{H}_k + J_k^T D_k^{-1} J_k) > 0$.

Then, if

$$\lim_{k\to\infty} \nabla B^\nu(v_k;\mu)^T s_k = 0,$$

it holds that

$$\lim_{k\to\infty} s_k = 0 \quad \text{and} \quad \lim_{k\to\infty} \nabla B^\nu(v_k;\mu) = 0.$$

□

Theorem 3.4.4. *Assume that the sequence of directions $\{d_k\}$ satisfies* (3.17). *Assume that for each $L_k B_k L_k^T$ factorization, the matrices B_k have more than m negative eigenvalues. Furthermore, assume that* $\limsup_{k\to\infty} \|L_k\| < \infty$. *Then, if*

$$\lim_{k\to\infty} d_k^T S^\nu(v_k;\mu)d_k = 0,$$

it holds that

$$\lim_{k\to\infty} d_k = 0 \quad \text{and} \quad \lim_{k\to\infty} \lambda_{\min}(S^\nu(v_k;\mu)) = 0.$$

Note that the algorithm proposed here gives to scheme for reducing μ to force convergence to the constrained local minimizer of (3.1), and no proof is given for a global convergence to such a minimizer as $\mu \to 0$.

4 A Primal-Dual Modified Barrier Method

In this chapter a new primal-dual interior method is introduced for solving a quadratic program in all-inequality form. The method can be viewed as a combination of the modified barrier and the primal-dual methods described in Chapter 3.

4.1 Optimality Conditions

The problem to be solved is given by

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & q(x) = \frac{1}{2}x^T H x + c^T x \\ \text{subject to} & Ax \geq b. \end{array} \tag{4.1}$$

The second-order optimality conditions for problem (4.1) (see discussion in Chapter 2.2) are

Theorem 4.1.1 (Second-order Necessary Conditions)**.** *The vector $x^* \in \mathbb{R}^n$ is a local minimizer of* (4.1) *only if*

1. *x^* is a KKT point, i.e., there exists a z^* such that*

   (i) $Ax^* \geq b$ *(feasibility),*

   (ii) $g(x^*) = A^T z^*$ *(stationarity),*

   (iii) $z^* \geq 0$ *(nonnegativity),*

   (iv) $(Ax^* - b) \cdot z^* = 0$ *(complementarity),*

2. $p^THp \geq 0$ *for all nonzero* $p \in \mathbb{R}^n$ *satisfying* $g(x^*)^Tp = 0$ *and* $A_{\mathcal{A}}p = 0$.

Second-order sufficient conditions can be obtained with the additional assumption of strict complementarity.

4.2 The Primal-Dual Modified Barrier Function

The primal-dual modified barrier function for (4.1) is derived by applying the Forsgren-Gill function (3.12) to the shifted QP given by

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & q(x) = \frac{1}{2}x^THx + c^Tx \\ \text{subject to} & Ax - b + \mu e \geq 0. \end{array} \tag{4.2}$$

The resulting function is given by

$$q(x) - \sum_{i=1}^{m}\Big(w_i\ln(r_i+\mu) + \nu w_i\Big(\ln\Big(\frac{(r_i+\mu)z_i}{w_i}\Big) + 1 - \frac{(r_i+\mu)z_i}{w_i}\Big)\Big),$$

where $r = r(x) = Ax - b$, $\nu > 0$ and w_i are the barrier parameters associated with each constraint. This function has the same stationary points as the function

$$\begin{aligned} B^\nu(x,z;w) &= q(x) - \sum_{i=1}^{m}(w_i(\ln(r_i+\mu) + \nu(w_i\ln((r_i+\mu)z_i) + w_i - (r_i+\mu)z_i)) \\ &= q(x) - \sum_{i=1}^{m}(w_i\ln((r_i+\mu)^{\nu+1}z_i^\nu) + \nu(w_i - (r_i+\mu)z_i)). \end{aligned}$$

Let z^a be an estimate of an optimal multiplier vector z^*. Then, if w is defined as $w = \mu z^a$, the primal-dual modified barrier function is given by

$$B^\nu(x,z;\mu,z^a) = q(x) - \sum_{i=1}^{m}(\mu z_i^a\ln((r_i+\mu)^{\nu+1}z_i^\nu) + \nu\mu(z_i^a - z_i) - \nu r_i z_i).$$

With the notation $R = \operatorname{diag}(r_1(x),\ldots,r_m(x))$ and $g = g(x)$, the gradient of $B^\nu(x,z;\mu,z^a)$ may be written as

$$\begin{aligned} \nabla B^\nu(x,z;\mu,z^a) &= \begin{pmatrix} g - A^T((1+\nu)\mu(R+\mu_I)z^a - \nu z) \\ -\nu(\mu Z^{-1}z^a - (r+\mu e)) \end{pmatrix} \\ &= \begin{pmatrix} g - A^T(\pi + \nu(\pi - z)) \\ \nu\big(R+\mu I\big)Z^{-1}(z-\pi) \end{pmatrix}, \end{aligned} \tag{4.3}$$

where $\pi = \pi(x;\mu,z^a)$ denotes the vector-valued auxiliary function

$$\pi(x;\mu,z^a) = \mu\big(R(x)+\mu I\big)^{-1}z^a. \tag{4.4}$$

Similarly, the Hessian of $B^\nu(x,z;\mu,z^a)$ may be written as

$$\nabla^2 B^\nu(x,z;\mu,z^a) = \begin{pmatrix} H+(1+\nu)A^T\Pi\big(R+\mu I\big)^{-1}A & \nu A^T \\ \nu A & \nu\Pi\big(R+\mu I\big)Z^{-2} \end{pmatrix}. \tag{4.5}$$

4.3 Definition of the Search Direction

As in Chapter 3.4, at a point $v=(x,z)$ the search direction $\Delta v = (\Delta x, \Delta z)$ is defined using an approximate Hessian of $B^\nu(x,z;\mu,z^a)$ given by

$$S^\nu(x,z;\mu) = \begin{pmatrix} H+(1+\nu)A^TZ\big(R+\mu I\big)^{-1}A & \nu A^T \\ \nu A & \nu\big(R+\mu I\big)Z^{-1} \end{pmatrix}. \tag{4.6}$$

The resulting modified Newton system

$$S^\nu(x,z;\mu)\begin{pmatrix}\Delta x\\ \Delta z\end{pmatrix} = -\nabla B^\nu(x,z;\mu,z^a),$$

is equivalent to the primal-dual system

$$\begin{pmatrix} H & A^T \\ A & -\big(R+\mu I\big)Z^{-1}\end{pmatrix}\begin{pmatrix}\Delta x\\ -\Delta z\end{pmatrix} = -\begin{pmatrix} g-A^Tz \\ \big(R+\mu I\big)Z^{-1}(z-\pi)\end{pmatrix},$$

which may be verified by premultiplying the modified Newton equations $S^\nu(v;\mu)\Delta v = -\nabla B^\nu(v;\mu,z^a)$ by the nonsingular matrix

$$N = \begin{pmatrix} I & -\big(\frac{1+v}{v}\big)A^T\big(R+\mu I\big)^{-1}Z \\ 0 & \frac{1}{v}I\end{pmatrix}. \tag{4.7}$$

The following lemma provides the relationship between the inertia of S^ν and the inertia of the primal-dual matrix

$$K = \begin{pmatrix} H & A^T \\ A & -\big(R+\mu I\big)Z^{-1}\end{pmatrix}. \tag{4.8}$$

Lemma 4.3.1. *Let S^ν be as in (4.6) and let K be as in (4.8), with $\mu > 0$ and $\nu > 0$. Also, let (x, z) be such that $z > 0$ and $Ax - b + \mu e > 0$ so that $\big(R + \mu I\big)Z^{-1}$ is diagonal positive definite. Then*

$$\begin{aligned}\mathrm{In}(S^\nu) &= \mathrm{In}(H + A^T\big(R + \mu I\big)^{-1}ZA) + (m, 0, 0),\\ \mathrm{In}(K) &= \mathrm{In}(H + A^T\big(R + \mu I\big)^{-1}ZA) + (0, m, 0).\end{aligned}$$

Moreover, if $H + A^T\big(R + \mu I\big)^{-1}ZA$ has at least one negative eigenvalue, then

$$\lambda_{\min}\big(H + A^T\big(R + \mu I\big)^{-1}ZA\big) \le \lambda_{\min}(S^\nu).$$

Proof. Theorem 1.5.2 gives

$$\begin{aligned}\mathrm{In}(K) &= \mathrm{In}(H + A^T\big(R + \mu I\big)^{-1}ZA) + \mathrm{In}(-\big(R + \mu I\big)Z^{-1})\\ &= \mathrm{In}(H + A^T\big(R + \mu I\big)^{-1}ZA) + (0, m, 0).\end{aligned}$$

Similarly,

$$\begin{aligned}\mathrm{In}(S^\nu) &= \mathrm{In}(H + (1+\nu)A^TZ\big(R + \mu I\big)^{-1}A - \nu A^TZ\big(R + \mu I\big)^{-1}A)\\ &\qquad + \mathrm{In}(\nu\big(R + \mu I\big)Z^{-1})\\ &= \mathrm{In}(H + A^T\big(R + \mu I\big)^{-1}ZA) + (m, 0, 0),\end{aligned}$$

as required. For the eigenvalue result, let Q and D denote the symmetric matrices

$$Q = \begin{pmatrix} H + A^T(R + \mu I)^{-1}ZA & 0\\ 0 & 0\end{pmatrix}, \quad \text{and} \quad D = \begin{pmatrix} 0 & 0\\ 0 & Z^{-1}(R + \mu I)\end{pmatrix},$$

and let T denote the nonsingular block-triangular matrix

$$T = \begin{pmatrix} I & 0\\ Z(R + \mu I)^{-1}A & I\end{pmatrix}.$$

Consider the identity

$$S^\nu = Q + T^TDT. \tag{4.9}$$

As T is nonsingular and D is positive semidefinite, Sylvester's law of inertia implies that T^TDT is positive semidefinite. Also, as $H + A^T(R + \mu I)^{-1}ZA$ has at least

one negative eigenvalue, it holds that $\lambda_{\min}(B) = \lambda_{\min}(H + A^T(R+\mu I)^{-1}ZA)$. It follows that

$$\begin{aligned}\lambda_{\min}(S^\nu) &= \lambda_{\min}(Q + T^T D T) \\ &= \min_{u\neq 0}\Big(\frac{u^TQu}{u^Tu} + \frac{u^TT^TDTu}{u^Tu}\Big) \\ &\geq \min_{u\neq 0}\Big(\frac{u^TQu}{u^Tu}\Big) \\ &= \lambda_{\min}(H + A^T(R+\mu I)^{-1}ZA). \quad \square\end{aligned}$$

The modified Newton system is solved using an inertia controlling LBL^T factorization. At each step a new search direction $\Delta v_k = s_k + d_k$ is computed, where s_k and d_k are directions of descent and negative curvature, respectively. The descent direction s_k is computed using a symmetric modification $\bar{S}^\nu$ of S^ν, such that $\bar{H} + A^T(R+\mu I)^{-1}ZA$ is sufficiently positive definite. The required modification can be computed using inertia-controlling symmetric indefinite factorization LBL^T of K, with modifications to the block diagonal matrix B in a way that ensures that K has correct inertia (see Chapter 2.4.3). It follows that $s = (s_x; s_z)$ is given by

$$\begin{pmatrix}\bar{H} & A^T \\ A & -\big(R+\mu I\big)Z^{-1}\end{pmatrix}\begin{pmatrix}s_x \\ -s_z\end{pmatrix} = -\begin{pmatrix}g - A^Tz \\ \big(R+\mu I\big)Z^{-1}(z-\pi)\end{pmatrix}, \tag{4.10}$$

where $\bar{H}$ is a symmetric modification such that $\bar{H} + A^T\big(R+\mu I\big)^{-1}ZA$ is sufficiently positive definite. The following lemma establishes that the vector s defined in this way is a descent direction with respect to $B^\nu(x, z; \mu, z^a)$.

Lemma 4.3.2. *Let s_k satisfy* (4.10)*. Then*

$$s^T\nabla B^\nu(x,z;\mu,z^a) = -s_x^T(\bar{H} + A^T\big(R+\mu I\big)^{-1}ZA)s_x - \nu(z-\pi)^T\big(R+\mu I\big)Z^{-1}(z-\pi).$$

Moreover, $s_x = 0$ if and only if $g - A^T\pi = 0$. Finally, if $s_x = 0$, then $s_z = \pi - z$.

Proof. As s_k satisfies (4.10), we can write s_z in terms of s_x

$$s_z = \pi - z - \big(R+\mu I\big)^{-1}ZAs_x. \tag{4.11}$$

The equivalence between the modified primal-dual equations (4.10) and the system $\bar{S}^\nu\Delta v = -\nabla B^\nu$ implies that s_k satisfies $\bar{S}^\nu s_k = -\nabla B^\nu$ and $s_k^T\nabla B^\nu = -s_k^T\bar{S}^\nu s_k$.

Substitution of (4.11) gives the required result. Also, if $s_x = 0$, equation (4.11) gives $s_z = \pi - z$. Finally, as s_k satisfies (4.10), the relation $\bar{H}s_x - A^T s_z = -g + A^T z$ holds, and substitution from (4.11) gives $(\bar{H} + A^T(R + \mu I)^{-1} ZA)s_x = A^T\pi - g$. The positive definiteness of $\bar{H} + A^T(R + \mu I)^{-1} ZA$ implies that $s_x = 0$ if and only if $A^T\pi - g = 0$. □

The direction of negative curvature d_k can also be computed using the inertia-controlling LBL^T. Given the factorization (2.13), if B_{22} has at least one negative eigenvalue, then write

$$\begin{pmatrix} d_x \\ -d_z \end{pmatrix} = P\widetilde{d}, \quad \text{where} \quad \begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22}^T \end{pmatrix} \widetilde{d} = \pm\sigma \begin{pmatrix} 0 \\ u \end{pmatrix}. \tag{4.12}$$

In this definition, $\sigma = \sqrt{-\lambda_{\min}(B_{22})}$, and u is the normalized eigenvector associated with $\lambda_{\min}(B_{22})$. The sign of σ in (4.12) is chosen to give $d^T \nabla B^\nu(v_k; \mu, z^a) \leq 0$. The following lemma establishes that d is a direction of negative curvature for S^ν.

Lemma 4.3.3. *Let d satisfy* (4.12). *Then d satisfies $Ad_x + \big(R + \mu I\big) Z^{-1} d_z = 0$, with $d^T S^\nu(x, z; \mu) d = -\lambda_{\min}(B_{22})^2$. Furthermore, as $Ad_x + \big(R + \mu I\big) Z^{-1} d_z = 0$, it holds that*

$$\begin{aligned} d_x^T (H + A^T \big(R + \mu I\big)^{-1} ZA) d_x &= \begin{pmatrix} d_x^T & -d_z^T \end{pmatrix} K \begin{pmatrix} d_x \\ -d_z \end{pmatrix} \\ &= \begin{pmatrix} d_x^T & d_z^T \end{pmatrix} S^\nu \begin{pmatrix} d_x \\ d_z \end{pmatrix}. \end{aligned}$$

Proof. From the definition of d_k in (4.12) we have

$$K \begin{pmatrix} d_x \\ -d_z \end{pmatrix} = PLBL^T P\widetilde{d} = \pm PLB \begin{pmatrix} 0 \\ u \end{pmatrix}.$$

The nonzero rows of $\pm PLB \begin{pmatrix} 0 \\ u \end{pmatrix}$ correspond to the H rows of K, and it follows

that $Ad_x + \big(R + \mu I\big)Z^{-1}d_z = 0$. Also

$$\begin{aligned}\widetilde{d}^T LBL^T\widetilde{d} &= \sigma^2 \begin{pmatrix} 0 & u \end{pmatrix} \begin{pmatrix} B_{11} & 0 \\ 0 & B_{22} \end{pmatrix} \begin{pmatrix} 0 \\ u \end{pmatrix} \\ &= \sigma^2 \lambda_{\min}(B_{22}) u^T u \\ &= -\lambda_{\min}(B_{22})^2.\end{aligned}$$

Forming $d_k^T S^\nu d_k$ and substituting the identity $Ad_x + \big(R + \mu I\big)Z^{-1}d_z = 0$, gives $d_x^T(H + A^T\big(R + \mu I\big)^{-1}ZA)d_x$. Similarly, from the definition of K we obtain $\begin{pmatrix} d_x & -d_z \end{pmatrix} K \begin{pmatrix} d_x \\ -d_z \end{pmatrix} = d_x^T(H + A^T\big(R + \mu I\big)^{-1}ZA)d_x$. □

Once Δv_k is computed, a backtracking line search is performed to ensure sufficient decrease. In this case a quadratic line-search model function is defined from the first three terms of the Taylor-series expansion of $B^\nu(v_k; \mu_k, z_k^a)$. An initial step $\alpha_k = 1$ is reduced by a constant factor until the reduction in $B^\nu(v_k + \alpha_k \Delta v_k; \mu_k, z_k^a)$ is at least as large as a fixed factor $\eta \in (0, \frac{1}{2})$ of the reduction predicted by the quadratic model. In particular, the new point $v_k + \alpha_k \Delta v_k$ satisfies

$$\begin{aligned}B^\nu(v_k + \alpha_k \Delta v_k; \mu_k, z_k^a) \leq B^\nu(v_k; \mu_k, z_k^a) + \eta(\alpha_k \nabla B^\nu(v_k; \mu_k, z_k^a)^T \Delta v_k \\ + \tfrac{1}{2}\alpha^2 \min(0, \Delta v_k^T \nabla^2 B^\nu(v_k; \mu_k, z_k^a)\Delta v_k), \quad (4.13)\end{aligned}$$

for some $\alpha_k \in (0, 1)$.

4.4 Multiplier and Barrier Parameter Updates and Convergence Test

We would like to update the multiplier estimates z^a as often a possible, but to ensure convergence it is necessary to impose optimality filter conditions. The proposed optimality filters are

$$\phi_L(x, z) = \psi(x, z) + \beta\omega(x, z) \quad \text{and} \quad \phi_O(x, z) = \beta\psi(x, z) + \omega(x, z),$$

where

$$\psi(x, z) = \|g(x) - A^T z\| \quad \text{and} \quad \omega(x, z) = \|\min(r(x), z)\|.$$

These filters provide a weighed measure of the accuracy of (x, z). The estimates z^a are updated if the new iterate v_k provides a sufficient decrease in either one of the two filters.

We also update the multiplier estimates if an approximate solution of the unconstrained optimization problem has been found, that is if $\nabla B^\nu(v_k; \mu_k, z_k^a)$ is sufficiently close to zero.

In general, the barrier parameter μ plays the role of a regularization parameter and we prefer not to reduce it unless the inertia of K is not correct (i.e., S^ν is not sufficiently positive definite). However, to guarantee convergence we may occasionally need to reduce it when the optimality filters fail, but the approximate solution of $\nabla B^\nu(v; \mu, z^a) = 0$ is reached.

The iterations are performed until optimality conditions for the original QP are sufficiently satisfied, that is $\|r_{opt}\|$ is near zero, where

$$r_{\text{opt}}(v) = \begin{pmatrix} g(x) - A^T z \\ \min(r(x), z) \end{pmatrix}. \tag{4.14}$$

4.5 The Algorithm

This section gives the formal statement of the proposed algorithm. It is important to note that a feasible starting point x_0 is required. If no such point is known, a separate feasibility phase is necessary.

The algorithm is similar to the inner-outer iterate algorithms, where the inner iterations minimize the penalty function for the current value of the barrier parameter, while the outer iterations check for optimality and update the parameter when necessary. The major difference here is the use of multiplier estimates as additional parameters. As such, after the search direction and appropriate step length are computed as described in Section 4.3, the multiplier estimates and the barrier parameter are updated as discussed in Section 4.4. This gives rise to the following classification of iterates. An iterate is called an L-iterate if the optimality filter ϕ_L is sufficiently reduced by the new iterate computed. Similarly, an iterate is an O-iterate if ϕ_O is sufficiently reduced. In both instances multiplier estimates z_k^a are updated to z_{k+1}. If the filters are not sufficiently reduced but the new iter-

ate is an approximate minimizer of $B^\nu(x, z; \mu_k, z_k^a)$, then the new iterate is called a B-iterate. In this case the multiplier estimates are again updated and both the barrier parameter μ_k and the subproblem tolerance τ_k are reduced. Otherwise, an iterate is called an F-iterate and the multiplier estimates z_k^a remain fixed.

Algorithm 4.5.1 Primal-dual Modified Barrier Method

INPUT: $v_0 = (x_0; z_0)$, such that $Ax_0 - b > 0$ and $z_0 > 0$;

Set control parameters $\nu > 1$, $\eta \in (0, \frac{1}{2})$, $0 < \tau_{\text{stop}} \ll 1$, $k_{\max} > 0$, $0 < \beta \ll 1$;

Set $\tau_0 > 0$, $\mu_0 > 0$;

$k = 0$;

$z_0^a \leftarrow z_0$;

Compute q_0, $r_0 = Ax_0 - b$, g_0;

for $k = 0 : \max k$ **do**

  if $\|r_{\text{opt}}(v_k)\| \le \tau_{\text{stop}}$ **then**

    exit;

  end if

  $k \leftarrow k + 1$;

  Solve for s_k and d_k according to (4.10) and (4.12);

  Set $\Delta v = s_k + d_k$;

  Set $\alpha_k = 1$;

  while Equation (4.13) not satisfied **do**

    $\alpha_k \leftarrow \gamma_\alpha \alpha_k$;

  end while

  $v_{k+1} \leftarrow v_k + \alpha_k \Delta v_k$;

  Compute q_{k+1}, r_{k+1}, g_{k+1};

  if $\phi_L(v_{k+1}) \le \frac{1}{2}\phi_L^{\max}$ **then** ▷ [L-iterate]

    $\phi_L^{\max} \leftarrow \frac{1}{2}\phi_L^{\max}$;

    $z_{k+1}^a = z_{k+1}$; $\tau_{k+1} = \tau_k$;

  else if $\phi_O(v_{k+1}) \le \frac{1}{2}\phi_O^{\max}$ **then** ▷ [O-iterate]

    $\phi_O^{\max} \leftarrow \frac{1}{2}\phi_O^{\max}$;

    $z_{k+1}^a = z_{k+1}$; $\tau_{k+1} = \tau_k$;

else if v_{k+1} satisfies $\|\nabla B(v_{k+1}, \mu_k, z_k^a)\| \leq \tau_k$ **then** ▷ [B-iterate]

$\mu_{k+1} = \gamma_\mu \mu_k$;

Reset x_k, z_k such that x_k is feasible;

Reset optimality filters;

$z_{k+1}^a = z_{k+1}$; $\tau_{k+1} = \frac{1}{2}\tau_k$;

else ▷ [F-iterate]

$z_{k+1}^a = z_k$; $\tau_{k+1} = \tau_k$;

end if

if S^ν is not sufficiently p.d. **then**

$\mu_{k+1} = \gamma_\mu \mu_k$;

Reset x_k, z_k such that x_k is feasible;

Reset optimality filters and τ;

end if

end for

OUTPUT: x_{k+1}; z_{k+1};

4.6 Convergence Results

We can now give a proof of convergence of the proposed algorithm. First, we show that for fixed μ and z^a the primal-dual modified barrier function $B^\nu(x, z; \mu z,^a)$ converges to a second-order local unconstrained minimizer.

Theorem 4.6.1. *Assume the sequence $\{v_k\}$ is contained in a compact set and (x_0, z_0) is such that $r(x_0) = Ax_0 - b > 0$ and $z_0 > 0$. Further assume that each search direction $\Delta v_k = s_k + d_k$ is computed using an inertia controlling LBL^T factorization, with $\lim_{k\to\infty} \|L_k\| < \infty$, such that s_k is a direction of descent and d_k is a direction of negative curvature and a backtracking line search is performed to ensure sufficient descent. Finally, assume that*

$$\liminf_{k\to\infty} \lambda_{\min}(R_k + \mu I)Z_k^{-1} > 0,$$
$$\limsup_{k\to\infty} \lambda_{\max}(R_k + \mu I)Z_k^{-1} < \infty \quad \textit{and}$$
$$\liminf_{k\to\infty} \lambda_{\min}(\bar{H}_k + A^T(R_k + \mu I)^{-1}ZA) > 0.$$

Then each s_k is a direction of sufficient descent, each d_k is a direction of sufficient negative curvature, and $\nabla B^\nu(x, z; \mu, z^a) \to 0$ as $k \to \infty$.

Proof. The principal role of the line search performed on each search direction Δv_k is to ensure that $\nabla B^\nu(v_k; \mu)^T s_k \to 0$ and $d_k^T \nabla^2 B^\nu(v_k; \mu) d_k \to 0$. If these conditions are satisfied and the directions s_k and d_k are directions of sufficient descent and sufficient negative curvature respectively (see Definitions 1.4.8 and 1.4.10), then every limit point of the sequence $\{v_k\}$ will satisfy the second-order necessary conditions for optimality (see Moré and Sorensen [19]). Algorithm 3.4.1 imposes negative curvature conditions on $d_k^T S^\nu(v_k; \mu) d_k$ rather than on $d_k^T \nabla^2 B^\nu(v_k; \mu) d_k$. However, properties of S^ν ensure that $d_k^T S^\nu(v_k; \mu) d_k \to d_k^T \nabla^2 B^\nu(v_k; \mu) d_k$, whenever $\nabla B^\nu(v_k; \mu) \to 0$. It remains to show that directions s_k and d_k are sufficient in the sense that if $\{\|s_k\|\}$ is bounded and

$$\lim_{k \to \infty} \nabla B^\nu(v_k; \mu, z^a)^T s_k = 0,$$

then it holds that

$$\lim_{k \to \infty} s_k = 0 \quad \text{and} \quad \lim_{k \to \infty} \nabla B^\nu(v_k; \mu, z^a) = 0.$$

In addition, if $\{\|d_k\|\}$ is bounded and

$$\lim_{k \to \infty} d_k^T S^\nu(v_k; \mu) d_k = 0,$$

then it holds that

$$\lim_{k \to \infty} d_k = 0 \quad \text{and} \quad \lim_{k \to \infty} \lambda_{\min} S^\nu(v_k; \mu) = 0.$$

First we show that the directions s_k are sufficient. Note that $\bar{S}_k^\nu$ can be written as $\bar{S}^\nu(v_k) = M_k^T W_k M_k$, where

$$W = \begin{pmatrix} \bar{H}_k + A^T (R_k + \mu I)^{-1} Z A & 0 \\ 0 & -\nu (R_k + \mu I) Z^{-1} \end{pmatrix} \quad \text{and}$$

$$M = \begin{pmatrix} I & 0 \\ (R_k + \mu I)^{-1} Z A & I \end{pmatrix}.$$

Together with the boundedness assumptions this gives $\liminf_{k\to\infty} \lambda_{\min}(\bar{S}_k^\nu) > 0$ and $\limsup_{k\to\infty} \lambda_{\max}(\bar{S}_k^\nu) < \infty$. As the sequence of directions $\{s_k\}$ satisfies (4.10), premultiplying by (4.7) implies that each s_k satisfies $\bar{S}_k^\nu s_k = -\nabla B_k^\nu$. Hence $s_k^T S_k^\nu s_k = -s_k^T (\nabla B_k^\nu)^T s_k \to 0$, and as S_k^ν is positive definite, with eigenvalues bounded away from zero, $s_k \to 0$. This in turn implies $\nabla B_k^\mu = -\bar{S}_k^\nu s_k \to 0$.

Now we show that the d_k vectors are sufficient. Assume that infinitely many of the matrices $H + A^T(R_k + \mu I)^{-1} ZA$ are not positive definite (otherwise the d_k are zero for k sufficiently large), then K has more then m negative eigenvalues and $\lambda_{\min}(H + A^T(R_k + \mu I)^{-1} ZA) < 0$. As the sequence of directions $\{d_k\}$ satisfies (4.12), it must hold that $d_k^T S_k^\nu d_k = -\lambda_{\min}(B_{22}, k)^2$. Therefore, if $\lim_{k\to\infty} d_k^T S_k^\nu d_k$, it must be the case that $\lim_{k\to\infty} \lambda_{\min}(B_{22}) = 0$. The boundedness of $\{\|L_k\|\}$ then implies that that $d_k \to 0$. The last limit can be obtained by considering the inertia controlling LBL^T factorization (2.13), with $K_{22} - K_{21}K_{11}^{-1}K_{12} = L_{22}B_{22}L_{22}^T$ and

$$\lambda_{\min}(K_{22} - K_{21}K_{11}^{-1}K_{12} = L_{22}B_{22}L_{22}^T) \leq \lambda_{\min}(H + A^T(R_k + \mu I)^{-1} ZA) < 0.$$

As $\{\|L_k\|\}$ is bounded and L_k are all unit lower triangular, $\{\|L_{22}\|\}$ and $\{\|L_{22}^{-1}\|\}$ are bounded, and it must hold that

$$\lim_{k\to\infty} \lambda_{\min}(K_{22} - K_{21}K_{11}^{-1}K_{12}) = \lim_{k\to\infty} \lambda_{\min}(L_{22}B_{22}L_{22}^T) = 0.$$

Together with Lemma 4.3.1, this implies $\lim_{k\to\infty} \lambda_{\min}(S_k^\nu) = 0$. □

The next theorem gives the main convergence result for Algorithm 4.5.1.

Theorem 4.6.2. *Let the assumptions of Theorem* 4.6.1 hold, then either

1. Algorithm 4.5.1 terminates with an approximate primal-dual first-order solution v_k satisfying $\|r_{\text{opt}}(v_k)\| \leq \tau_{\text{stop}}$, where r_{opt} is defined by (4.14), or

2. there exists a subsequence $\mathcal{S}$ such that $\lim_{k\in\mathcal{S}} \mu_k \to 0$, $\lim_{k\in\mathcal{S}} \tau_k \to 0$, and for each $k \in \mathcal{S}$ the vector v_{k+1} is an approximate unconstrained local minimizer of $B^\nu(v_k; \mu_k, z_k^a)$ and, furthermore, $\lim_{k\in\mathcal{S}} \mu_k \to 0$, forces convergence to the first-order solution of (4.1).

Proof. If there exists a subsequence of $\|r_{\text{opt}}(v_k)\|$ that converges to zero, then part 1 holds. Therefore, assume no such subsequence exists, i.e., $\|r_{\text{opt}}(v_k)\|$ is bounded away from zero.

From the definitions of the functions ϕ_L and ϕ_O and the update strategies for $\phi_L^{\max}$ and $\phi_O^{\max}$, it is clear that the number of L-iterates and O-iterates is finite. We claim that there must be an infinite number of B-iterates. To show this, assume to the contrary that there is a finite number of B-iterates. Then there must exist a $\widehat{k}$ such that all iterates such that $k > \widehat{k}$ are F-iterates. That implies that for all $k > \widehat{k}$ the barrier parameter and the multiplier estimates remain constant, i.e., $\mu_k = \mu_{\widehat{k}}$ and $z_k^a = z_{\widehat{k}}^a$ for all $k > \widehat{k}$. In addition, this implies that $\tau_k = \tau_{\widehat{k}}$ for all $k > \widehat{k}$. However, Theorem 4.6.1 ensures that with fixed barrier parameter and multiplier estimates, $\{v_k\}$ converges to an unconstrained local minimizer as $k \to \infty$, i.e., $\nabla B^\nu(v_k; \mu_{\widehat{k}}, z_{\widehat{k}}) \to 0$. This implies that $\tau_k \to 0$ and we have a contradiction. It follows that there must be an infinite number of B-iterates.

As τ_k is reduced at each B-iterate, an infinite number of such iterates implies that $\tau_k \to 0$. It follows that the gradient of B^ν converges to the zero vector, with

$$\nabla B^\nu(v_k; \mu_k, z_k^a) = \begin{pmatrix} g_k - A^T(\pi_k + \nu(\pi_k - z_k)) \\ \nu\big(R_k + \mu_k I\big) Z_k^{-1}(z_k - \pi_k) \end{pmatrix} \to 0.$$

This implies that $\pi_k \to z_k$, i.e., $R_k z_k + \mu_k(z_k - z_k^a) \to 0$ as $k \to \infty$, and, in turn, $g_k - A^T z_k \to 0$. From the definition of a B-iterate we also have that $\mu_k \to 0$ as $k \to \infty$. Together with $R_k z_k + \mu_k(z_k - z_k^a) \to 0$ this implies that $R_k z_k \to 0$. It follows that as $k \to \infty$, v_k converges to the first-order KKT point of (4.1). □

4.7 Numerical Results

This section provides some numerical results to illustrate the properties of Algorithm 4.5.1. All test QPs are taken from the CUTEr collection of test problems. Each such test problem is given by

$$\begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{minimize}} && q(x) \\ &\text{subject to} && \begin{pmatrix} c_\ell \\ x_\ell \end{pmatrix} \le \begin{pmatrix} Ax - b \\ x \end{pmatrix} \le \begin{pmatrix} c_u \\ x_u \end{pmatrix}, \end{aligned}$$

Here we only consider a subset of quadratic test problems for which $c_\ell < c_u$ (i.e., there are no linear equality constraints) and for which a feasible x_0 is provided by CUTEr. Also, since implementation is done in MATLAB, only problems of moderate size are considered. For all problems presented here the algorithm was initialized with the parameters and initial values summarized in Table 4.7.1.

Table 4.7.1: Control Parameters and Initial Values for Algorithm 4.5.1

Parameter	Value	Parameter	Value	Parameter	Value
ν	1.0	$k_{\max}$	500	μ_0	1.0e-5
η	1.0e-2	$\tau_{\texttt{stop}}$	1.0e-6	γ_μ	0.1
β	1.0e-5	τ_0	1.0e-2	γ_α	0.5

For each test problem Table 4.7.2 specifies the name of the CUTEr problem, its size (i.e., the number of variables n and the number of linear inequality constraints m), function value `Opt` $q(x)$ at the solution, the number of iterations the algorithm took to converge to that solution and the final barrier parameter. The table also provides statistics for the number of L-, O-, B-, and F-iterates used, where LO-Itr% gives the percentage of all iterates that were L- and O-iterates, *etc.* The $d \neq 0$ indicates the number of directions of negative curvature computed. Of all the problems tested with the parameters listed in Table 4.7.1 only one, `QUDLIN`, failed to converge. However, given more iterations and varying some of the parameters this problem converges, with μ essentially converging to zero.

Table 4.7.2: Results for a Subset of CUTEr test QPs

QP Name	n	m	Opt $q(x)$	Itr	μ_F	LO-Itr%	B-Itr%	F-Itr%	$d \neq 0$
BQP1VAR	1	0	-4.8358e-10	16	1.0e-05	100	0	0	0
BQPGABIM	46	0	-3.7903e-05	23	1.0e-05	95.7	0	4.35	0
BQPGASIM	50	0	-5.5198e-05	23	1.0e-05	95.7	0	4.35	0
HS118	15	17	6.6482e+02	35	1.0e-05	88.6	0	11.4	0
HS21	2	1	4.0000e-02	43	1.0e-05	55.8	0	44.2	0
HS268	5	5	-1.4463e+04	80	1.0e-05	85	0	15	0
HS3	2	0	-2.2204e-16	1	1.0e-05	100	0	0	0
HS35	3	1	-8.8889e+00	21	1.0e-05	100	0	0	0
HS35I	3	1	-8.8889e+00	20	1.0e-05	100	0	0	0
HS35MOD	2	1	-6.2500e+00	22	1.0e-05	100	0	0	0
HS3MOD	2	0	-2.2204e-16	1	1.0e-05	100	0	0	0
HS44	4	6	-1.3000e+01	43	1.0e-05	51.2	0	48.8	8
HS44NEW	4	6	-1.5000e+01	21	1.0e-05	95.2	0	4.76	2

Table 4.7.2: Results for a Subset of CUTEr test QPs, Continued

QP Name	n	m	Opt $q(x)$	Itr	μ_F	LO%	B%	F%	$d \neq 0$
HS76	4	3	-4.6818e+00	22	1.0e-05	95.5	0	4.55	0
HS76I	4	3	-4.6818e+00	22	1.0e-05	100	0	0	0
OSLBQP	8	0	6.2500e+00	79	1.0e-05	73.4	2.53	24.1	0
PALMER1C	8	0	-1.0741e+05	1	1.0e-05	100	0	0	0
PALMER1D	7	0	-1.0741e+05	1	1.0e-05	100	0	0	0
PALMER2C	8	0	-1.6103e+04	1	1.0e-05	100	0	0	0
PALMER3C	8	0	-1.5824e+04	1	1.0e-05	100	0	0	0
PALMER4C	8	0	-1.7275e+04	1	1.0e-05	100	0	0	0
PRIMAL1	325	85	-3.5013e-02	39	1.0e-05	92.3	0	7.69	0
PRIMAL2	649	96	-3.3734e-02	34	1.0e-05	91.2	0	8.82	0
PRIMAL3	745	111	-1.3576e-01	46	1.0e-05	89.1	0	10.9	0
PRIMAL4	1489	75	-7.4609e-01	44	1.0e-05	90.9	0	9.09	0
PRIMALC1	230	9	-6.1553e+03	80	1.0e-05	68.8	0	31.2	0
PRIMALC2	231	7	-3.5513e+03	83	1.0e-05	61.4	0	38.6	0
PRIMALC5	287	8	-4.2723e+02	62	1.0e-05	80.6	0	19.4	0
PRIMALC8	520	8	-1.8309e+04	158	1.0e-05	34.8	0	65.2	0
S268	5	5	-1.4463e+04	80	1.0e-05	85	0	15	0
SIM2BQP	1	0	-5.9136e-10	25	1.0e-05	68	0	32	0
SIMBQP	2	0	-5.0354e-10	27	1.0e-05	81.5	0	18.5	0
TOINTQOR	50	0	-1.1598e+03	1	1.0e-05	100	0	0	0
ZANGWIL2	2	0	-8.4267e+01	1	1.0e-05	100	0	0	0
ZECEVIC2	2	2	-4.1250e+00	32	1.0e-05	71.9	0	28.1	0
BIGGSB1	100	0	-1.9850e+00	381	1.0e-05	13.6	0	86.4	0
CHENHARK	100	0	-2.0000e+00	37	1.0e-05	100	0	0	0
CVXBQP1	100	0	2.2725e+02	23	1.0e-05	100	0	0	0
DEGTRID	101	0	-9.9500e+01	45	1.0e-05	100	0	0	0
DEGTRID2	101	0	-9.9500e+01	191	1.0e-06	35.1	14.7	50.3	0
DIXON3DQ	100	0	-2.0000e+00	1	1.0e-05	100	0	0	0
DQDRTIC	100	0	0.0000e+00	1	1.0e-05	100	0	0	0
HARKERP2	100	0	-5.0000e-01	2	1.0e-05	100	0	0	0
HILBERTA	2	0	2.9582e-31	1	1.0e-05	100	0	0	0
HILBERTB	10	0	4.0388e-29	1	1.0e-05	100	0	0	0
JNLBRNG1	64	0	-1.7896e-01	56	1.0e-05	42.9	0	57.1	0
JNLBRNG2	64	0	-3.9528e+00	33	1.0e-05	66.7	0	33.3	0
JNLBRNGA	64	0	-3.6116e-01	73	1.0e-05	47.9	0	52.1	0
JNLBRNGB	64	0	-7.2552e+00	76	1.0e-05	48.7	0	51.3	0
LISWET1	103	100	-2.5861e+01	78	1.0e-05	48.7	2.56	48.7	0
LISWET2	103	100	-1.7360e+01	107	1.0e-05	42.1	11.2	46.7	0
LISWET3	103	100	-1.0562e+01	108	1.0e-05	40.7	12	47.2	0
LISWET4	103	100	-7.6508e+00	64	1.0e-05	48.4	0	51.6	0
LISWET5	103	100	-1.6540e+02	78	1.0e-05	55.1	3.85	41	0
LISWET6	103	100	-2.2469e+01	88	1.0e-05	50	5.68	44.3	0
LISWET7	103	100	-2.5485e+01	131	1.0e-05	29	0.763	70.2	0
LISWET8	103	100	-2.5489e+01	41	1.0e-05	68.3	0	31.7	0
LISWET9	102	100	-5.5605e+00	97	1.0e-05	34	0	66	0

Table 4.7.2: Results for a Subset of CUTEr test QPs, Continued

QP Name	n	m	Opt $q(x)$	Itr	μ_F	LO%	B%	F%	$d \neq 0$
LISWET10	103	100	-2.5997e+01	68	1.0e-05	42.6	0	57.4	0
LISWET11	103	100	-2.4788e+01	142	1.0e-05	21.8	0	78.2	0
LISWET12	103	100	-5.6574e+00	160	1.0e-05	16.9	0	83.1	0
MOSARQP1	100	10	-1.5420e+02	28	1.0e-05	100	0	0	0
MOSARQP2	100	10	-2.0652e+02	26	1.0e-05	100	0	0	0
NCVXBQP1	100	0	-1.9956e+06	253	1.0e-08	20.6	0	79.4	182
NCVXBQP2	100	0	-1.3236e+06	294	1.0e-07	16.3	0	83.7	233
NCVXBQP3	100	0	-6.6599e+05	339	1.0e-08	15.9	0	84.1	261
NOBNDTOR	64	0	-5.5211e-01	57	1.0e-05	43.9	0	56.1	0
OBSTCLAE	64	0	1.3979e+00	27	1.0e-05	100	0	0	0
OBSTCLAL	64	0	1.3979e+00	120	1.0e-05	23.3	0	76.7	0
OBSTCLBL	64	0	2.8750e+00	156	1.0e-05	14.7	0	85.3	0
OBSTCLBM	64	0	2.8750e+00	19	1.0e-05	100	0	0	0
OBSTCLBU	64	0	2.8750e+00	140	1.0e-05	15.7	0	84.3	0
PENTDI	100	0	-7.5000e-01	169	1.0e-06	56.8	10.7	32.5	0
TESTQUAD	1000	0	0.0000e+00	1	1.0e-05	100	0	0	0
TORSION1	64	0	-4.9234e-01	48	1.0e-05	54.2	0	45.8	0
TORSION2	64	0	-4.9234e-01	26	1.0e-05	100	0	0	0
TORSION3	64	0	-1.2705e+00	23	1.0e-05	73.9	0	26.1	0
TORSION4	64	0	-1.2705e+00	24	1.0e-05	100	0	0	0
TORSION5	64	0	-2.8971e+00	2	1.0e-05	100	0	0	0
TORSION6	64	0	-2.8971e+00	25	1.0e-05	100	0	0	0
TORSIONA	64	0	-4.0570e-01	49	1.0e-05	53.1	0	46.9	0
TORSIONB	64	0	-4.0570e-01	27	1.0e-05	96.3	0	3.7	0
TORSIONC	64	0	-1.1766e+00	23	1.0e-05	73.9	0	26.1	0
TORSIOND	64	0	-1.1766e+00	23	1.0e-05	100	0	0	0
TORSIONE	64	0	-2.7984e+00	2	1.0e-05	100	0	0	0
TORSIONF	64	0	-2.7984e+00	25	1.0e-05	100	0	0	0
TRIDIA	100	0	-1.0000e+00	1	1.0e-05	100	0	0	0
YAO	200	200	-7.5658e+00	50	1.0e-05	50	0	50	0

5 A Primal-Dual Modified Lagrangian-Barrier Method

The method described in Chapter 4 must have a feasible starting point to obtain a solution of (4.1). To get convergence from any starting point, one could first run a phase one algorithm to attain feasibility. Here we present another solution. By adding slack variables (see Introduction), the proposed method converts all linear inequalities into a set of linear equalities and simple bounds. The bounds are treated using the primal-dual modified barrier method, while the equality constrains are treated using a penalty type method. In particular, the algorithm proposed in this section uses the primal-dual augmented Lagrangian, briefly discussed in Section 2.3. To simplify exposition, it is assumed that the problem to be solved is in standard form.

5.1 Optimality Conditions

Consider the QP in standard form

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & c^T x + \frac{1}{2} x^T H x \\ \text{subject to} & Ax = b, \quad x \geq 0. \end{array} \tag{5.1}$$

Based on the discussion of Chapter 2.2, the second-order optimality conditions for this problem are

Theorem 5.1.1 (Second-order Necessary Conditions)**.** *The vector $x^* \in \mathbb{R}^n$ is a local minimizer of* (5.1) *only if*

1. *x^* is a KKT point, i.e., there exist y^* and z^* such that*

i. $Ax^* = 0$ *and* $x \geq 0$ *(feasibility),*

ii. $g(x^*) = A^T y^* + z^*$ *(stationarity),*

iii. $z^* \geq 0$ *(nonnegativity),*

iv. $x^* \cdot z^* = 0$ *(complementarity),*

2. $p^T H p \geq 0$ *for all nonzero* $p \in \mathbb{R}^n$ *satisfying* $g(x^*)^T p = 0$, $p_{\mathcal{A}} = 0$ *and* $Ap = 0$.

Second-order sufficient conditions are achieved if we require strict complementarity.

5.2 The Merit Function

Using a primal-dual augmented lagrangian for the equality constraints and a primal dual augmented barrier term for the nonnegativity bounds, we can convert (5.1) into the following unconstrained problem.

$$\underset{x,y,z}{\text{minimize}} \quad M^\nu(x, y, z; \mu_E, y^a, \mu_I, z^a),$$

where

$$M^\nu(x, y, z; \mu_E, y^a, \mu_I, z^a) = c^T x + \tfrac{1}{2} x^T H x - (Ax - b)^T y^a + \frac{1}{2\mu_E} \|Ax - b\|^2 \quad (5.2)$$

$$+ \frac{\nu}{2\mu_E} \|Ax - b + \mu_E (y - y^a)\|^2 \quad (5.3)$$

$$- \sum_{i=1}^{n} \mu_I z_i^a \ln \left((x_i + \mu_I)^{\nu+1} z_i^\nu \right) \quad (5.4)$$

$$- \nu \sum_{i=1}^{n} \left(\mu_I (z_i^a - z_i) - x_i z_i \right). \quad (5.5)$$

The vectors y and z are the multipliers for the equality and nonnegativity constraints, respectively.

Differentiating $M^\nu(x, y, z; \mu_E, y^a, \mu_I, z^a)$ yields

$$\nabla M^\nu(x, y, z; \mu_E, y^a, \mu_I, z^a) = \begin{pmatrix} g - A^T\big((1+\nu)\pi_E - \nu y\big) - (1+\nu)\pi_I + \nu z \\ \nu\big(Ax - b + \mu_E(y - y^a)\big) \\ \nu\big(-\mu_I Z^{-1} z^a + \mu_I e + x\big) \end{pmatrix}$$
$$= \begin{pmatrix} g - A^T y - z + (1+\nu)(A^T\big(y - \pi_E\big) + (z - \pi_I) \\ \nu\mu_E(y - \pi_E) \\ \nu\big(X + \mu_I I\big)Z^{-1}(z - \pi_I) \end{pmatrix},$$

where $\pi_E = \pi_E(x) = y^a - \dfrac{1}{\mu_E}\big(Ax - b\big)$ and $\pi_I = \pi_I(x) = \mu_I\big(X + \mu_I I\big)^{-1} z^a$ are the auxiliary vector-valued functions. Differentiating again, gives the Hessian

$$\nabla^2 M^\nu(x, y, z; \mu_E, y^a, \mu_I, z^a) = \begin{pmatrix} H_1 & \nu A^T & \nu I \\ \nu A & \nu\mu_E I & 0 \\ \nu I & 0 & \nu\Pi_I\big(X + \mu_I I\big)Z^{-2} \end{pmatrix},$$

where $H_1 = H + \frac{1}{\mu_E}(1+\nu)A^T A + (1+\nu)\Pi_I\big(X + \mu_I I\big)^{-1}$.

5.3 Search Directions

Analogous to Chapter 4, instead of solving the standard Newton equations $\nabla^2 M^\nu \Delta v = -\nabla M^\nu$, modified equations $S^\nu \Delta v = -\nabla M^\nu$ are used. The approximation S^ν is obtained by replacing π_I with z. Resulting modified Hessian is given by

$$S^\nu(x, y, y; \mu_E, \mu_I) = \begin{pmatrix} H_2 & \nu A^T & \nu I \\ \nu A & \nu\mu_E I & 0 \\ \nu I & 0 & \nu\big(X + \mu_I I\big)Z^{-1} \end{pmatrix}, \tag{5.6}$$

where $H_2 = H + \dfrac{1}{\mu_E}(1+\nu)A^T A + (1+\nu)\big(X + \mu_I I\big)^{-1} Z$.

Premultiplying $S^\nu \Delta v = -\nabla M^\nu$ by the nonsingular matrix

$$N = \begin{pmatrix} I & -\frac{1+\nu}{\nu\mu_E}A^T & -\big(X + \mu_I I\big)^{-1} Z \\ 0 & \frac{1}{\nu} I & 0 \\ 0 & 0 & \frac{1}{\nu}\big(X + \mu_I I\big)^{-1} Z \end{pmatrix} \tag{5.7}$$

gives an equivalent primal-dual system

$$\begin{pmatrix} H + \big(X + \mu_I I\big)^{-1} Z & -A^T & 0 \\ A & \mu_E I & 0 \\ \big(X + \mu_I I\big)^{-1} Z & 0 & I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = - \begin{pmatrix} g - A^T y - \pi_I \\ \mu_E (y - \pi_E) \\ z - \pi_I \end{pmatrix},$$

or

$$\begin{pmatrix} H + \big(X + \mu_I I\big)^{-1} Z & A^T \\ A & -\mu_E I \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} = - \begin{pmatrix} g - A^T (y - \pi_I) \\ \mu_E (y - \pi_E) \end{pmatrix}, \tag{5.8}$$

$$\Delta z = -(z - \pi_I + \big(X + \mu_I I\big)^{-1} Z \Delta x). \tag{5.9}$$

Therefore, we can find the search direction using inertia-controlling factorization of the matrix K discussed in Section 2.4.2, where

$$K = \begin{pmatrix} H + \big(X + \mu_I I\big)^{-1} Z & A^T \\ A & -\mu_E I \end{pmatrix}.$$

As in previous chapters, the search direction is given by $\Delta v = s + d$, where s is a descent direction and d is a direction of negative curvature. The descent direction s is obtained by solving

$$\begin{pmatrix} \bar{H} + \big(X + \mu_I I\big)^{-1} Z & A^T \\ A & -\mu_E I \end{pmatrix} \begin{pmatrix} s_x \\ -s_y \end{pmatrix} = - \begin{pmatrix} g - A^T y - \pi_I) \\ \mu_E (y - \pi_E) \end{pmatrix}, \tag{5.10}$$

$$s_z = -(z - \pi_I + \big(X + \mu_I I\big)^{-1} Z s_x), \tag{5.11}$$

where $\bar{H}$ is a symmetric modification of H such that $\bar{H} + \big(X + \mu_I I\big)^{-1} Z + \frac{1}{\mu_E} A^T A$ is sufficiently positive definite.

Note that the above system is equivalent to $\bar{S}^\nu s = -\nabla M^\nu$, where

$$\bar{S}^\nu = \begin{pmatrix} \bar{H} + \frac{1}{\mu_E}(1+\nu) A^T A + (1+\nu)\big(X + \mu_I I\big)^{-1} Z & \nu A^T & \nu I \\ \nu A & \nu \mu_E I & 0 \\ \nu I & 0 & \nu\big(X + \mu_I I\big) Z^{-1} \end{pmatrix}.$$

Lemma 5.3.1. *Let s satisfy* (5.10)*, then $s^T \nabla M^\nu \leq 0$.*

Proof. Since s satisfies (5.10) we can write s_y and s_z in terms of s_x:

$$s_y = \pi_E - y - \frac{1}{\mu_E} A s_x$$

$$s_z = \pi_I - z - \big(X + \mu_I I\big)^{-1} Z s_x,$$

and as (5.10) is equivalent to $\bar{S}^\nu s = -\nabla M^\nu$, we have

$$\begin{aligned}
s^T \nabla M^\nu &= -s^T \bar{S}^\nu s \\
&= -s_x^T (\bar{H} + \frac{1}{\mu_E}(1+\nu)A^T A + (1+\nu)\big(X + \mu_I I\big)^{-1} Z) s_x \\
&\quad - 2\nu s_x^T A^T s_y - 2\nu s_x^T s_z - \nu\mu s_y^T s_y - \nu s_z^T \big(X + \mu_I I\big) Z^{-1} s_z \\
&= -s_x^T (\bar{H} + \frac{1}{\mu_E} A^T A + \big(X + \mu_I I\big)^{-1} Z) s_x - \frac{\nu}{\mu_E} s_x^T A^T A s_x \\
&\quad - \nu s_x^T \big(X + \mu_I I\big)^{-1} Z s_x - 2\nu s_x^T A^T (\pi_E - y) + 2\frac{\nu}{\mu_E} s_x^T A^T A s_x \\
&\quad - 2\nu s_x^T (\pi_I - z) + 2\nu s_x^T \big(X + \mu_I I\big)^{-1} Z s_x - \nu\mu_E (\pi_E - y)^T (\pi_E - y) \\
&\quad + 2\nu s_x^T A^T (\pi_E - y) - \frac{\nu}{\mu_E} s_x^T A^T A s_x - \nu (\pi_I - z)^T \big(X + \mu_I I\big) Z^{-1} (\pi_I - z) \\
&\quad + 2\nu s_x^T (\pi_I - z) - \nu s_x^T \big(X + \mu_I I\big)^{-1} Z s_x \\
&= -s_x^T (\bar{H} + \frac{1}{\mu_E} A^T A + \big(X + \mu_I I\big)^{-1} Z) s_x \\
&\quad - \nu\mu_E (\pi_E - y)^T (\pi_E - y) - \nu s_x^T \big(X + \mu_I I\big)^{-1} Z s_x \\
&\leq 0.
\end{aligned}$$

□

The direction of negative curvature d can be computed from

$$\begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22}^T \end{pmatrix} \widetilde{d} = \pm\sigma \begin{pmatrix} 0 \\ u \end{pmatrix}, \tag{5.12}$$

with

$$\begin{pmatrix} d_x \\ -d_y \end{pmatrix} = P\widetilde{d} \quad \text{and} \quad d_z = -\big(X + \mu_I I\big)^{-1} Z d_x,$$

where $\sigma = \sqrt{-\lambda_{\min}(B_{22})}$ and u is an eigenvector of unit length associated with $\lambda_{\min}(B_{22})$. The sign of d is chosen such that $d^T \nabla M \leq 0$. As we solve a condensed primal-dual system and thus cannot use the gradient directly, an alternate but equivalent condition to determine the sign of d is required. The following lemma gives the necessary result.

Lemma 5.3.2. *Requiring $d^T \nabla M^\nu \leq 0$ is equivalent to $\begin{pmatrix} d_x^T & d_y^T \end{pmatrix} q \leq 0$, where*

$$q = \begin{pmatrix} \nabla_x M^\nu - \left(X + \mu_I I\right)^{-1} Z \nabla_z M \\ \nabla_y M^\nu \end{pmatrix}.$$

Proof. From definition of d, $d_z = -\left(X + \mu_I I\right)^{-1} Z d_x$. Hence

$$\begin{aligned} d^T \nabla M^\nu &= d_x^T \nabla_x M^\nu + d_y^T \nabla_y M^\nu + d_z^T \nabla_z M^\nu \\ &= d_x^T \nabla_x M^\nu + d_y^T \nabla_y M^\nu - d_x^T \left(X + \mu_I I\right)^{-1} Z \nabla_z M^\nu \\ &= d_x^T (\nabla_x M^\nu - \left(X + \mu_I I\right)^{-1} Z \nabla_z M^\nu) + d_y^T \nabla_y M^\nu. \end{aligned}$$

□

Finally, we verify that thus defined d is indeed a direction of negative curvature.

Lemma 5.3.3. *Let d satisfy* (5.12)*. Then d is a direction of negative curvature with respect to S^ν, i.e. $d^T S^\nu d \leq 0$.*

Proof. From definition of d we have

$$\begin{aligned} \begin{pmatrix} H + (X + \mu_I I)^{-1} Z & A^T \\ A & -\mu_E I \end{pmatrix} \begin{pmatrix} d_x \\ -d_y \end{pmatrix} &= PLBL^T P^T P \widetilde{d} \\ &= \pm \sigma PLB \begin{pmatrix} 0 \\ u \end{pmatrix} \\ &= \pm \sigma P \begin{pmatrix} 0 \\ L_{22} B_{22} u \end{pmatrix} \\ &= \begin{pmatrix} * \\ 0 \end{pmatrix}, \end{aligned}$$

where * is of length n and contains $\pm\sigma L_{22}B_{22}u$. Hence $Ad_x + \mu d_y = 0$. Consider

$$\begin{aligned}
\begin{pmatrix} d_x^T & -d_y^T \end{pmatrix} \begin{pmatrix} H + \big(X + \mu_I I\big)^{-1} Z & A^T \\ A & -\mu_E I \end{pmatrix} \begin{pmatrix} d_x \\ -d_y \end{pmatrix} & \\
&= \widetilde{d} P^T P L B L^T P^T P \widetilde{d} \\
&= \sigma^2 u^T B_{22} u \\
&= \sigma^2 \lambda_{\min}(B_{22}) u^T u \\
&= -\lambda_{\min}(B_{22})^2,
\end{aligned}$$

also

$$\begin{aligned}
\begin{pmatrix} d_x^T & -d_y^T \end{pmatrix} [K] \begin{pmatrix} d_x \\ -d_y \end{pmatrix} &= d_x^T (H + \big(X + \mu_I I\big)^{-1} Z) d_x - 2 d_x^T A^T d_y - \mu_E d_y^T d_y \\
&= d_x^T (H + \big(X + \mu_I I\big)^{-1} Z) d_x + \frac{2}{\mu_E} d_x^T A^T A d_x \\
&\quad - \frac{1}{\mu_E} d_x^T A^T A d_x \\
&= d_x^T (H + \frac{1}{\mu_E} A^T A + \big(X + \mu_I I\big)^{-1} Z) d_x,
\end{aligned}$$

and finally,

$$\begin{aligned}
d^T S^\nu d =& d_x^T (H + \frac{1}{\mu_E}(1+\nu) A^T A + (1+\nu)\big(X + \mu_I I\big)^{-1} Z) d_x \\
&\quad - 2\nu d_x^T A^T d_y - 2\nu d_x^T d_z - \nu \mu_E d_y^T d_y - \nu d_z^T \big(X + \mu_I I\big) Z^{-1} d_z \\
&= d_x^T (H + \frac{1}{\mu_E}(1+\nu) A^T A + (1+\nu)\big(X + \mu_I I\big)^{-1} Z) d_x + \frac{2\nu}{\mu_E} d_x^T A^T A d_x \\
&\quad + 2\nu d_x^T \big(X + \mu_I I\big)^{-1} Z d_x - \frac{\nu}{\mu_E} d_x^T A^T A d_x - \nu d_x^T \big(X + \mu_I I\big)^{-1} Z d_x \\
&= d_x^T (H + \frac{1}{\mu_E} A^T A + \big(X + \mu_I I\big)^{-1} Z) d_x \\
&= -\lambda_{\min}(B_{22})^2 \\
&\leq 0.
\end{aligned}$$

□

Once Δv_k is computed, a backtracking line search is performed to ensure

sufficient decrease. Here we require that $v_k + \alpha_k \Delta v_k$ satisfies

$$M^\nu(v_k + \alpha_k \Delta v_k) \leq M^\nu(v_k) + \eta(\alpha_k \nabla M^\nu(v_k)^T \Delta v_k + \tfrac{1}{2}\alpha^2 \Delta v_k^T \min(0, \nabla^2 M^\nu(v_k) \Delta v_k). \tag{5.13}$$

5.4 Updates to the Multiplier Estimates, Barrier Parameter and Convergence Test

As in the previous chapter, multiplier estimates y^a, z^a are updated when certain optimality conditions are met. For the primal-dual modified Langrangian-Barrier method the optimality filters are

$$\begin{aligned}
\phi_V(x, z) &= \eta(x, y) + \beta\omega(x, y, z) + \beta\psi(x, z), \\
\phi_L(x, y, z) &= \beta\eta(x, y) + \omega(x, y, z) + \beta\psi(x, z) \text{ and} \\
\phi_O(x, y, z) &= \beta\eta(x, y) + \beta\omega(x, y, z) + \psi(x, z),
\end{aligned}$$

where

$$\eta(x, y) = \|Ax - b\|, \quad \omega(x, y, z) = \|g(x) - A^T y - z\| \quad \text{and} \quad \psi(x, z) = \|\min(x, z)\|.$$

These filters provide a weighed measure of the optimality of a given triple (x, y, z). The multiplier estimates are updated if the new iterate v_k provides a sufficient decrease in either one of the three filters.

We also update the multiplier estimates if an approximate solution of the unconstrained optimization problem has been found, that is if $\nabla M^\nu(v_k; \mu_k, z_k^a)$ is sufficiently close to zero.

In general, the barrier parameter and the Lagrangian parameter are used for constraint regularization and thus it is preferable not to reduce them unless necessary for convergence. Usually both are reduced when the optimality filters fail, but the approximate solution of $\nabla M^\nu(v_k; \mu_E, y^a, \mu_I z^a) = 0$ is reached.

The iterations are performed until optimality conditions for the original QP

are sufficiently satisfied, that is $\|r_{opt}\|$ is near zero, where

$$r_{\text{opt}}(v) = \begin{pmatrix} Ax - b \\ g(x) - A^T y - z \\ \min(x, z) \end{pmatrix}. \tag{5.14}$$

5.5 The Algorithm

This section gives the formal statement of the proposed algorithm. As in the previous chapter, each new iterate is classified based on weather the optimality filters are sufficiently reduced or an approximate subproblem minimizer is found. Here however, there are three filters to consider. Thus we have L-, V-, and O-iterates for ϕ_L, ϕ_V and ϕ_O, respectively. In all three cases both y_k^a and z_k^a are updated. If the new iterate is an approximate minimizer of $M^\nu(x, y, z; \mu_k^E, y_k^a, \mu_k^I, z_k^a)$, then it is called an M-iterate. In this case, both y_k^a and z_k^a are updated and μ_k is reduced. Otherwise, an iterate is called an F-iterate and all multiplier estimates remain fixed.

Algorithm 5.5.1 Primal-dual Modified Lagrangian-Barrier Method

INPUT: $v_0 = (x_0, y_0, z_0)$.

If necessary move x_0 with in bounds and $z_0 > 0$.

Set control parameters $\nu > 1$, $\eta \in (0, \frac{1}{2})$, $0 < \tau_{\text{stop}} \ll 1$, $k_{\max} > 0$, $0 < \beta \ll 1$;

Set $\tau_0 > 0$, $\mu_0^E > 0$, μ_0^I;

$k = 0$;

$y_0^a \leftarrow y_0$, $z_0^a \leftarrow z_0$;

Compute q_0, g_0;

for $k = 0 : \max k$ **do**

  if $\|r_{\text{opt}}(v_k)\| \leq \tau_{\text{stop}}$ **then**

    exit;

  end if

  $k \leftarrow k + 1$;

  Solve for s_k and d_k according to (5.10) and (5.12);

  Set $\Delta v = s_k + d_k$; $\alpha_k = 1$;

while Equation (5.13) not satisfied **do**

$\alpha_k \leftarrow \gamma\alpha_k$;

end while

$v_{k+1} \leftarrow v_k + \alpha_k \Delta v_k$;

Compute q_{k+1}, g_{k+1};

if $\phi_V(v_{k+1}) \leq \frac{1}{2}\phi_V^{\max}$ **then** ▷ [V-iterate]

$\phi_V^{\max} \leftarrow \frac{1}{2}\phi_V^{\max}$;

$y_{k+1}^a = y_{k+1}$; $z_{k+1}^a = z_{k+1}$; $\tau_{k+1} = \tau_k$;

else if $\phi_L(v_{k+1}) \leq \frac{1}{2}\phi_L^{\max}$ **then** ▷ [L-iterate]

$\phi_L^{\max} \leftarrow \frac{1}{2}\phi_L^{\max}$;

$y_{k+1}^a = y_{k+1}$; $z_{k+1}^a = z_{k+1}$; $\tau_{k+1} = \tau_k$;

else if $\phi_O(v_{k+1}) \leq \frac{1}{2}\phi_O^{\max}$ **then** ▷ [O-iterate]

$\phi_O^{\max} \leftarrow \frac{1}{2}\phi_O^{\max}$;

$y_{k+1}^a = y_{k+1}$; $z_{k+1}^a = z_{k+1}$; $\tau_{k+1} = \tau_k$;

else if v_{k+1} satisfies $\|\nabla M^\nu(v_{k+1}, \mu_k, z_k^a)\| \leq \tau_k$ **then** ▷ [M-iterate]

$\mu_{k+1}^E = \frac{1}{10}\mu_k$;

$\mu_{k+1}^I = \frac{1}{10}\mu_k$;

Move x_{k+1} within bounds (if necessary), reset z_{k+1} away from zero;

Reset optimality filters;

$y_{k+1}^a = \max(-y_{max}e, \min(y_{k+1}, y_{\max}))$; $z_{k+1}^a = z_{k+1}$; $\tau_{k+1} = \frac{1}{2}\tau_k$;

else ▷ [F-iterate]

$z_{k+1}^a = z_k$; $\tau_{k+1} = \tau_k$;

end if

if S^ν is not sufficiently p.d. **then**

$\mu_{k+1}^I = \frac{1}{10}\mu_k$;

Reset x_k, z_k as before;

Reset optimality filters and τ;

end if

end for

OUTPUT: x_{k+1}; y_{k+1}; z_{k+1};

5.6 Convergence Results

The results are very similar to the ones in Chapter 4. First convergence of the merit function (5.2) for fixed parameters is shown.

Theorem 5.6.1. *Assume the sequence* $\{v_k\}$ *is contained in a compact set and* (x_0, z_0) *is such that* $x_0 > 0$ *and* $z_0 > 0$. *Further assume that each search direction* $\Delta v_k = s_k + d_k$ *is computed using inertia controlling* LBL^T, *with* $\lim_{k\to\infty} \|L_k\| < \infty$, *such that* s_k *is a direction of descent and* d_k *is a direction of negative curvature and a backtracking line search is performed to ensure sufficient descent. Finally, assume that*

$$\begin{aligned} \liminf_{k\to\infty} \lambda_{\min}(X_k + \mu I) Z_k^{-1} &> 0, \\ \limsup_{k\to\infty} \lambda_{\max}(X_k + \mu I) Z_k^{-1} &< \infty \text{ and} \\ \liminf_{k\to\infty} \lambda_{\min}(\bar{H}_k + (X_k + \mu I)^{-1} Z A) &> 0. \end{aligned}$$

Then each s_k *is a direction of sufficient descent, each* d_k *is a direction of sufficient negative curvature and* $\nabla M^\nu(v_k; \mu_E, y^a, \mu_I, z^a) \to 0$ *as* $k \to \infty$.

Now we can state the general convergence theorem.

Theorem 5.6.2. *Let the assumptions of Theorem* 5.6.1 hold, then either

1. Algorithm 5.5.1 terminates with an approximate primal-dual first-order solution v_k satisfying $\|r_{\text{opt}}(v_k)\| \le \tau_{\text{stop}}$, where r_{opt} is defined by (5.14), or

2. there exists a subsequence $\mathcal{S}$ such that $\lim_{k\in\mathcal{S}} \mu_k^E \to 0$, $\lim_{k\in\mathcal{S}} \mu_k^I \to 0$, $\lim_{k\in\mathcal{S}} \tau_k \to 0$, and for each $k \in \mathcal{S}$ the vector v_{k+1} is an approximate unconstrained local minimizer of $M^\nu(v_k; \mu_k^E, y_k^a, \mu_k^I, z_k^a)$ and, furthermore, since barrier and penalty parameters converge to zero, convergence to the first-order solution of (5.1) is achieved.

6 Numerical Results

In this chapter we describe numerical experiments for the algorithm proposed in Chapter 5. The implementation was done in MATLAB and run on an iMac machine. The numerical results are intended to show that the proposed method is reasonable and illustrate its properties.

The quadratic programs tested here all come from the CUTEr collection of test problems. Both convex and nonconvex problems of moderate size are considered. The CUTEr problems are all defined with the format

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & q(x) \\ \text{subject to} & \begin{pmatrix} c_\ell \\ x_\ell \end{pmatrix} \leq \begin{pmatrix} Ax-b \\ x \end{pmatrix} \leq \begin{pmatrix} c_u \\ x_u \end{pmatrix}. \end{array}$$

With addition of slack variables (see Introduction) it is converted into an equivalent form

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & c^Tx + \frac{1}{2}x^THx \\ \text{subject to} & Ax = b, \quad \ell \leq x \leq u. \end{array} . \tag{6.1}$$

We can extend the Algorithm 5.5.1 with upper and lower bounds on the variables instead of just nonnegativity constraints.

6.1 Extension to Upper and Lower Bounds

Here we give a brief discussion of how the method of Chapter 5 is implemented to deal with upper and lower bounds.

As in the previous chapter, we use a primal dual augmented lagrangian for the equalities and primal dual augmented barrier terms for the lower and upper

bounds to convert this QP into the following unconstrained problem.

$$\underset{x,y,z_1,z_2}{\text{minimize}} \quad M^\nu(x,y,z_1,z_2;\mu_E,y^a,\mu_I,z_\ell^a,z_u^a),$$

where

$$\begin{aligned} M^\nu(x,y,z_\ell,z_u;\mu_E,y^a,\mu_I,z_\ell^a,z_u^a) \;=\; & q(x)-(Ax-b)^Ty^a+\frac{1}{2\mu_E}\|Ax-b\|^2 \\ & +\frac{\nu}{2\mu_E}\|Ax-b+\mu_E(y-y^a)\|^2 \\ & -\sum_{i=1}^n \mu_I[z_\ell^a]_i \ln\big((x_i-\ell_i+\mu_I)^{\nu+1}[z_\ell]_i^\nu\big) \\ & -\nu\sum_{i=1}^n \big(\mu_I([z_\ell^a]_i-[z_\ell]_i)-(x_i-\ell_i)z_{\ell i}\big) \\ & -\sum_{i=1}^n \mu_I[z_u^a]_i \ln\big((u-x_i+\mu_I)^{\nu+1}[z_u]_i^\nu\big) \\ & -\nu\sum_{i=1}^n \big(\mu_I([z_u^a]_i-[z_u]_i)-(u_i-x_i)[z_u]_i\big), \end{aligned}$$

where y, z_ℓ, z_u are the multipliers for the equalities, lower and upper bounds, respectively.

The gradient $\nabla M^\nu(x,y,z_\ell,z_u;\mu_E,y^a,\mu_I,z_\ell^a,z_u^a)$ is given by

$$\begin{aligned} \nabla M^\nu &= \begin{pmatrix} g-A^T\big((1+\nu)\pi_E-\nu y\big)-(1+\nu)\pi_\ell+\nu z_\ell+(1+\nu)\pi_u-\nu z_u \\ \nu\big(Ax-b+\mu_E(y-y^a)\big) \\ \nu\big(-\mu_I Z_\ell^{-1}z_\ell^a+\mu_I e+x-l\big) \\ \nu\big(-\mu_I Z_u^{-1}z_u^a+\mu_I e+u-x\big) \end{pmatrix} \\ &= \begin{pmatrix} g-A^T\big((1+\nu)\pi_E-\nu y\big)-(1+\nu)\pi_\ell+\nu z_\ell+(1+\nu)\pi_u-\nu z_u \\ \nu\mu_E(y-\pi_E) \\ \nu\big(X-L+\mu_I I\big)Z_\ell^{-1}(z_\ell-\pi_\ell) \\ \nu\big(U-X+\mu_I I\big)Z_u^{-1}(z_u-\pi_u) \end{pmatrix}, \end{aligned}$$

where $\pi_E=\pi_E(x)\triangleq y^a-\dfrac{1}{\mu_E}\big(Ax-b\big)$, $\pi_\ell=\pi_\ell(x)\triangleq \mu_I\big(X-L+\mu_I I\big)^{-1}z_\ell^a$ and $\pi_u=\pi_u(x)\triangleq \mu_I\big(U-X+\mu_I I\big)^{-1}z_u^a$.

The Hessian $\nabla^2 M^\nu(x, y, z_\ell, z_u; \mu_E, y^a, \mu_I, z_\ell^a, z_u^a)$ is then given by

$$\nabla^2 M^\nu = \begin{pmatrix} H_1 & \nu A^T & \nu I & -\nu I \\ \nu A & \nu\mu_E I & 0 & 0 \\ \nu I & 0 & \nu\Pi_\ell\big(X - L + \mu_I I\big) Z_\ell^{-2} & 0 \\ -\nu I & 0 & 0 & \nu\Pi_u\big(U - X + \mu I\big) Z_u^{-2} \end{pmatrix},$$

where $H_1 = H + \frac{1}{\mu_E}(1+\nu)A^T A + (1+\nu)\Pi_\ell\big(X - L + \mu_I I\big)^{-1} + (1+\nu)\Pi_u\big(U - X + \mu_I I\big)^{-1}$.

Again, we solve $\nabla^2 M^\nu \Delta v = -\nabla M^\nu$ using a modified Newton method, with $S^\nu(x, y, z_\ell, z_u, \mu_E, \mu_I)$. Specifically, we consider the approximate Hessian obtained by replacing π_ℓ and π_u with z_ℓ and z_u, respectively. This gives us

$$S^\nu = \begin{pmatrix} \bar{H}_1 & \nu A^T & \nu I & -\nu I \\ \nu A & \nu\mu_E I & 0 & 0 \\ \nu I & 0 & \nu\big(X - L + \mu_I I\big) Z_\ell^{-1} & 0 \\ -\nu I & 0 & 0 & \nu\big(U - X + \mu I\big) Z_u^{-1} \end{pmatrix},$$

where $\bar{H}_1 = H + \frac{1}{\mu_E}(1+\nu)A^T A + (1+\nu)\big(X - L + \mu_I I\big)^{-1} Z_\ell + (1+\nu)\big(U - X + \mu_I I\big)^{-1} Z_u$.

After left multiplying $S^\nu \Delta v = -\nabla B^\nu$ by the nonsingular matrix

$$\begin{pmatrix} I & -\frac{1+\nu}{\nu\mu_E} A^T & -\big(X - L + \mu_I\big)^{-1} Z_\ell & \big(U - X + \mu_I\big)^{-1} Z_u \\ 0 & \frac{1}{\nu} I & 0 & 0 \\ 0 & 0 & \frac{1}{\nu}\big(X - L + \mu_I I\big)^{-1} Z_\ell & 0 \\ 0 & 0 & 0 & \frac{1}{\nu}\big(U - X + \mu_I I\big)^{-1} Z_u \end{pmatrix},$$

we get the equivalent system

$$\begin{pmatrix} H_2 & -A^T & 0 & 0 \\ A & \mu_E I & 0 & 0 \\ \big(X - L + \mu_I I\big)^{-1} Z_\ell & 0 & I & 0 \\ -\big(U - X + \mu_I I\big)^{-1} Z_u & 0 & 0 & I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z_\ell \\ \Delta z_u \end{pmatrix} = - \begin{pmatrix} r_2 \\ \mu_E(y - \pi_E) \\ z_\ell - \pi_\ell \\ z_u - \pi_u \end{pmatrix},$$

where $H_2 = H + \big(X - L + \mu_I I\big)^{-1} Z_\ell + \big(U - X + \mu_I I\big)^{-1} Z_u$
and $r_2 = c + Hx - A^T y - \pi_\ell + \pi_u$,or

$$\begin{aligned} \begin{pmatrix} H_2 & A^T \\ A & -\mu_E I \end{pmatrix} \begin{pmatrix} \Delta x \\ -\Delta y \end{pmatrix} &= - \begin{pmatrix} r_2 \\ \mu_E(y - \pi_E) \end{pmatrix}, \\ \Delta z_\ell &= -z_\ell + \pi_\ell - \big(X - L + \mu_I I\big)^{-1} Z_\ell \Delta x \\ \Delta z_u &= -z_u + \pi_u + \big(U - X + \mu_I I\big)^{-1} Z_u \Delta x. \end{aligned}$$

The descent direction s in then computed is by solving

$$\begin{aligned}\begin{pmatrix}\bar{H}_2 & A^T\\ A & -\mu_E I\end{pmatrix}\begin{pmatrix}s_x\\ -s_y\end{pmatrix} &= -\begin{pmatrix}r_2\\ \mu_E(y-\pi_E)\end{pmatrix},\\ s_{z_\ell} &= -z_\ell+\pi_\ell-\big(X-L+\mu_I I\big)^{-1}Z_\ell s_x\\ s_{z_u} &= -z_u+\pi_u+\big(U-X+\mu_I I\big)^{-1}Z_u s_x,\end{aligned}$$

where $\bar{H}_2=\bar{H}+\big(X-L+\mu_I I\big)^{-1}Z_\ell+\big(U-X+\mu_I I\big)^{-1}Z_u$ and $\bar{H}$ is a symmetric modification of H such that $\bar{H}_2+\frac{1}{\mu_E}A^TA$ is sufficiently positive definite. While the direction of negative curvature d is computed via

$$\begin{pmatrix}L_{11}^T & L_{21}^T\\ 0 & L_{22}^T\end{pmatrix}\widetilde{d}=\pm\sigma\begin{pmatrix}0\\ u\end{pmatrix},$$

with

$$\begin{pmatrix}d_x\\ -d_y\end{pmatrix}=P\widetilde{d},\quad d_{z_\ell}=-\big(X-L+\mu_I I\big)^{-1}Z_\ell d_x \text{ and } d_{z_u}=\big(U-X+\mu_I I\big)^{-1}Z_u d_x,$$

where $\sigma=\sqrt{-\eta_{\min}(B_{22})}$, and u is an eigenvector of unit length associated with $\eta_{\min}(B_{22})$. The sign of d is chosen such that $d^T\nabla M^\nu\le 0$.

Lemma 6.1.1. *The condition $d^T\nabla M^\nu\le 0$ is equivalent to $\begin{pmatrix}d_x^T & d_y^T\end{pmatrix}q\le 0$, where*

$$q=\begin{pmatrix}\nabla_x M^\nu-\big(X-L+\mu_I I\big)^{-1}Z_\ell\nabla_{z_\ell}M^\nu+\big(U-X+\mu_I I\big)^{-1}Z_u\nabla_{z_u}M^\nu\\ \nabla_y M^\nu\end{pmatrix}.\quad\square$$

6.2 Results

For all problems presented here the algorithm was initialized with the parameters and initial values summarized in Table 6.2.1.

For each test problem the results Table 6.2.2 specifies the name of the CUTEr problem, its size (i.e, the number of variables n and the number of linear inequality constraints m), function value `Opt` $q(x)$ at the solution, the number of iterations the algorithm took to converge to that solution and the final barrier parameters. The table also provides statistics for the number of V-, O-, L-, M-,

Table 6.2.1: Control Parameters and Initial Values for Algorithm 5.5.1

Parameter	Value	Parameter	Value	Parameter	Value
ν	1.0	$k_{\max}$	1000	μ_0^E	1.0e-4
η	1.0e-2	$\tau_{\texttt{stop}}$	1.0e-6	μ_0^I	1.0e-5
β	1.0e-5	γ_α	0.5	τ_0	1.0e-2

and F-iterates used, where VOL% gives the percentage of all iterates that were V-L-, and O-iterates, *etc.* The $d \neq 0$ column gives the number of times a direction of negative curvature was computed. While the last column states the outcome: 1 if an optimal solution was found in under the maximum iterations allowed, 0 otherwise.

Table 6.2.2: Results for CUTEr test QPs

QP	n	m	OptQ	Itr	μ_E^F	μ_I^F	VOL%	M%	F%	$d \neq 0$	conv
AVGASA	8	10	-4.6319e+00	54	1.9e-12	1.0e-05	48.1	0	51.9	0	1
AVGASB	8	10	-4.4832e+00	60	6.3e-13	1.0e-05	51.7	0	48.3	0	1
BIGGSC4	4	7	-2.4500e+01	68	1.0e-05	1.0e-06	77.9	1.47	20.6	0	1
BQP1VAR	1	0	2.4567e-10	17	3.9e-15	1.0e-05	100	0	0	0	1
BQPGABIM	50	0	-3.7903e-05	21	1.0e-04	1.0e-05	100	0	0	0	1
BQPGASIM	50	0	-5.5198e-05	20	1.0e-04	1.0e-05	100	0	0	0	1
DUAL1	85	1	3.5013e-02	46	1.2e-10	1.0e-05	100	0	0	0	1
DUAL2	96	1	3.3734e-02	38	3.0e-10	1.0e-05	100	0	0	0	1
DUAL3	111	1	1.3576e-01	54	9.1e-10	1.0e-05	100	0	0	0	1
DUAL4	75	1	7.4609e-01	44	7.2e-10	1.0e-05	100	0	0	0	1
DUALC1	9	215	6.1553e+03	248	1.1e-12	1.0e-05	31.5	0	68.5	0	1
DUALC2	7	229	3.5513e+03	197	7.2e-16	1.0e-05	38.1	0	61.9	0	1
DUALC5	8	278	4.2723e+02	89	1.4e-18	1.0e-05	88.8	0	11.2	0	1
DUALC8	8	503	1.8309e+04	189	4.7e-10	1.0e-06	73	0	27	0	1
GENHS28	10	8	9.2717e-01	2	2.1e-15	1.0e-05	100	0	0	0	1
GMNCASE1	175	300	2.6697e-01	37	1.0e-04	1.0e-05	100	0	0	0	1
GMNCASE2	175	1050	-9.9444e-01	28	1.0e-04	1.0e-05	100	0	0	0	1
GMNCASE3	175	1050	1.5251e+00	25	1.0e-04	1.0e-05	100	0	0	0	1
GMNCASE4	175	350	5.9469e+03	6	1.0e-04	1.0e-05	100	0	0	0	1
GOFFIN	51	50	-1.1369e-13	2	1.0e-04	1.0e-05	100	0	0	0	1
GOULDQP1	32	17	-3.4853e+03	42	1.0e-04	1.0e-05	100	0	0	0	1
HATFLDH	4	7	-2.4500e+01	55	1.0e-05	1.0e-06	94.5	1.82	3.64	0	1
HS118	15	17	6.6482e+02	37	1.0e-04	1.0e-05	100	0	0	0	1
HS21	2	1	4.0000e-02	51	5.1e-16	1.0e-05	88.2	0	11.8	0	1
HS268	5	5	-1.4463e+04	94	8.9e-10	1.0e-05	100	0	0	0	1
HS3	2	0	-8.4703e-22	5	2.5e-32	1.0e-05	100	0	0	0	1
HS35	3	1	-8.8889e+00	22	1.3e-14	1.0e-05	100	0	0	0	1
HS35I	3	1	-8.8889e+00	22	1.4e-14	1.0e-05	100	0	0	0	1

Table 6.2.2: Results for CUTEr test QPs, Continued

QP	n	m	OptQ	Itr	μ_E^F	μ_I^F	VOL%	M%	F%	$d \neq 0$	conv
HS35MOD	3	1	2.5000e-01	47	9.5e-10	1.0e-05	100	0	0	0	1
HS3MOD	2	0	-1.1102e-16	2	3.9e-24	1.0e-05	100	0	0	0	1
HS44	4	6	-1.3000e+01	50	4.5e-14	1.0e-05	70	0	30	9	1
HS44NEW	4	6	-1.5000e+01	21	2.6e-14	1.0e-05	100	0	0	2	1
HS51	5	3	0.0000e+00	1	8.3e-25	1.0e-05	100	0	0	0	1
HS52	5	3	5.3266e+00	2	7.4e-11	1.0e-05	100	0	0	0	1
HS53	5	3	4.0930e+00	20	1.6e-16	1.0e-05	100	0	0	0	1
HS76	4	3	-4.6818e+00	21	1.0e-04	1.0e-05	100	0	0	0	1
HS76I	4	3	-4.6818e+00	21	1.0e-04	1.0e-05	100	0	0	0	1
LOTSCHD	12	7	2.3984e+03	31	6.9e-15	1.0e-05	64.5	0	35.5	0	1
MAKELA4	21	40	3.1086e-15	8	1.0e-04	1.0e-05	100	0	0	0	1
OSLBQP	8	0	6.2500e+00	65	9.7e-10	1.0e-05	100	0	0	0	1
PALMER1C	8	0	-1.0741e+05	1	7.9e-12	1.0e-05	100	0	0	0	1
PALMER1D	7	0	-1.0741e+05	1	3.3e-12	1.0e-05	100	0	0	0	1
PALMER2C	8	0	-1.6103e+04	1	1.7e-12	1.0e-05	100	0	0	0	1
PALMER3C	8	0	-1.5824e+04	1	5.7e-13	1.0e-05	100	0	0	0	1
PALMER4C	8	0	-1.7275e+04	1	7.3e-13	1.0e-05	100	0	0	0	1
PRIMAL1	325	85	-3.5013e-02	39	1.0e-04	1.0e-05	100	0	0	0	1
PRIMAL2	649	96	-3.3734e-02	33	1.0e-04	1.0e-05	100	0	0	0	1
PRIMAL3	745	111	-1.3576e-01	48	1.0e-04	1.0e-05	100	0	0	0	1
PRIMAL4	1489	75	-7.4609e-01	50	1.0e-04	1.0e-05	96	0	4	0	1
PRIMALC1	230	9	-6.1553e+03	853	1.0e-05	1.0e-06	9.38	0.117	90.5	0	1
PRIMALC2	231	7	-3.5513e+03	77	1.0e-04	1.0e-05	50.6	0	49.4	0	1
PRIMALC5	287	8	-4.2723e+02	59	1.0e-04	1.0e-05	71.2	0	28.8	0	1
PRIMALC8	520	8	-1.8309e+04	291	1.0e-04	1.0e-05	15.5	0	84.5	0	1
QPCBLEND	83	74	-7.8413e-03	88	1.0e-05	1.0e-06	80.7	0	19.3	0	1
QPCBOEI1	384	351	1.8016e+05	1000	1.0e-04	1.0e-05	0.6	0	99.4	0	0
QPCBOEI2	143	166	1.5512e+07	1000	1.0e-14	1.0e-15	22.8	0	77.2	0	0
QPCSTAIR	467	356	4.2938e+05	1000	1.0e-04	1.0e-05	0.3	0	99.7	0	0
QPNBLEND	83	74	-8.7044e-03	102	1.0e-05	1.0e-06	71.6	0	28.4	0	1
QPNBOEI1	384	351	1.5600e+05	1000	1.0e-04	1.0e-05	0.6	0	99.4	0	0
QPNBOEI2	143	166	1.0776e+05	1000	1.0e-04	1.0e-05	0.4	0	99.6	217	0
QPNSTAIR	467	356	4.5354e+06	1000	1.0e-07	1.0e-08	6.1	0	93.9	0	0
S268	5	5	-1.4463e+04	94	8.9e-10	1.0e-05	100	0	0	0	1
SIM2BQP	2	0	5.6968e-10	22	1.4e-14	1.0e-05	100	0	0	0	1
SIMBQP	2	0	4.0354e-10	23	8.1e-15	1.0e-05	100	0	0	0	1
STEENBRA	432	108	1.6958e+04	307	5.9e-10	1.0e-06	26.4	0	73.6	5	1
TAME	2	1	0.0000e+00	33	1.1e-51	1.0e-05	100	0	0	0	1
TOINTQOR	50	0	-1.1598e+03	1	8.3e-21	1.0e-05	100	0	0	0	1
ZANGWIL2	2	0	-8.4267e+01	1	0.0e+00	1.0e-05	100	0	0	0	1
ZECEVIC2	2	2	-4.1250e+00	32	1.0e-04	1.0e-05	100	0	0	0	1
ALLINQP	100	50	-9.1593e+00	202	1.0e-05	1.0e-06	37.1	0.495	62.4	0	1
AUG2D	220	100	1.1080e+02	3	7.5e-22	1.0e-05	100	0	0	0	1
AUG2DC	220	100	1.8424e+02	3	2.3e-20	1.0e-05	100	0	0	0	1
AUG2DCQP	220	100	3.0399e+02	339	3.1e-13	1.0e-05	14.7	0	85.3	0	1

Table 6.2.2: Results for CUTEr test QPs, Continued

QP	n	m	OptQ	Itr	μ_E^F	μ_I^F	VOL%	M%	F%	$d \neq 0$	conv
AUG2DQP	220	100	1.7797e+02	319	1.0e-09	1.0e-05	33.5	0	66.5	1	1
AUG3D	156	27	8.3333e-02	2	8.2e-20	1.0e-05	100	0	0	0	1
AUG3DC	156	27	3.5843e+01	2	2.4e-14	1.0e-05	100	0	0	0	1
AUG3DCQP	156	27	3.9288e+01	144	5.3e-14	1.0e-05	36.1	0	63.9	0	1
AUG3DQP	156	27	4.1833e+00	216	1.6e-10	1.0e-06	62.5	0.463	37	17	1
BIGGSB1	100	0	-1.9850e+00	58	1.0e-05	1.0e-06	96.6	1.72	1.72	0	1
BLOCKQP1	210	101	-9.4000e+01	33	1.0e-04	1.0e-05	87.9	0	12.1	13	1
BLOCKQP2	210	101	-9.3806e+01	22	1.0e-04	1.0e-05	100	0	0	1	1
BLOCKQP3	210	101	-4.5000e+01	132	1.0e-04	1.0e-05	27.3	0	72.7	119	1
BLOCKQP4	210	101	-4.5776e+01	25	1.0e-04	1.0e-05	100	0	0	2	1
BLOCKQP5	210	101	-4.5000e+01	157	1.0e-04	1.0e-05	22.3	0	77.7	141	1
BLOWEYA	202	102	-4.4543e-01	1000	1.0e-10	1.0e-11	31.9	0.6	67.5	0	0
BLOWEYB	202	102	-2.9699e-01	1000	4.8e-12	1.0e-12	35.2	0.7	64.1	0	0
BLOWEYC	202	102	-3.0275e-01	1000	3.1e-12	1.0e-12	36.4	0.7	62.9	0	0
CHENHARK	100	0	-2.0000e+00	39	9.5e-10	1.0e-05	100	0	0	0	1
CVXBQP1	100	0	2.2725e+02	20	2.7e-12	1.0e-05	100	0	0	0	1
CVXQP1	100	50	1.1591e+04	27	9.7e-11	1.0e-05	100	0	0	0	1
CVXQP2	100	25	8.1209e+03	29	2.8e-14	1.0e-05	100	0	0	0	1
CVXQP3	100	75	1.1943e+04	36	9.1e-13	1.0e-05	100	0	0	0	1
DEGENQP	10	1005	1.5528e-10	31	3.7e-14	1.0e-05	100	0	0	0	1
DEGTRID	101	0	-9.9500e+01	47	9.2e-10	1.0e-05	100	0	0	0	1
DEGTRID2	101	0	-9.9500e+01	115	1.6e-10	1.0e-06	99.1	0.87	0	0	1
DEGTRIDL	101	1	5.0000e-01	51	8.7e-10	1.0e-05	100	0	0	0	1
DIXON3DQ	100	0	-2.0000e+00	1	0.0e+00	1.0e-05	100	0	0	0	1
DQDRTIC	100	0	0.0000e+00	1	0.0e+00	1.0e-05	100	0	0	0	1
FERRISDC	400	103	-5.8247e-05	267	1.0e-06	1.0e-07	80.5	0	19.5	77	1
GOULDQP2	199	99	9.3976e-07	32	1.5e-08	1.0e-05	100	0	0	0	1
GOULDQP3	199	99	2.2133e-03	23	1.7e-10	1.0e-05	100	0	0	0	1
HARKERP2	100	0	-4.9999e-01	54	5.6e-10	1.0e-05	100	0	0	0	1
HILBERTA	2	0	2.9582e-31	1	3.3e-24	1.0e-05	100	0	0	0	1
HILBERTB	10	0	4.0388e-29	1	5.0e-21	1.0e-05	100	0	0	0	1
HUES-MOD	100	2	3.4830e+07	51	2.9e-19	1.0e-05	90.2	0	9.8	0	1
HUESTIS	100	2	3.4830e+09	173	4.1e-16	1.0e-06	53.8	0	46.2	0	1
JNLBRNG1	100	0	-1.7896e-01	78	9.2e-14	1.0e-05	53.8	0	46.2	0	1
JNLBRNG2	100	0	-3.9528e+00	34	7.4e-12	1.0e-05	100	0	0	0	1
JNLBRNGA	100	0	-3.6116e-01	76	1.7e-13	1.0e-05	72.4	0	27.6	0	1
JNLBRNGB	100	0	-7.2552e+00	76	2.4e-15	1.0e-05	77.6	0	22.4	0	1
LISWET1	103	100	-2.5861e+01	76	8.3e-10	1.0e-05	86.8	0	13.2	0	1
LISWET2	103	100	-1.7360e+01	92	8.5e-10	1.0e-05	91.3	0	8.7	0	1
LISWET3	103	100	-1.0562e+01	84	9.1e-10	1.0e-05	100	0	0	0	1
LISWET4	103	100	-7.6508e+00	59	5.4e-10	1.0e-05	96.6	0	3.39	0	1
LISWET5	103	100	-1.6540e+02	74	9.9e-10	1.0e-05	100	0	0	0	1
LISWET6	103	100	-2.2469e+01	79	8.8e-10	1.0e-05	97.5	0	2.53	0	1
LISWET7	103	100	-2.5485e+01	182	3.6e-10	1.0e-05	36.8	0	63.2	0	1
LISWET8	103	100	-2.5489e+01	73	1.3e-10	1.0e-05	82.2	0	17.8	0	1

Table 6.2.2: Results for CUTEr test QPs, Continued

QP	n	m	OptQ	Itr	μ_E^F	μ_I^F	VOL%	M%	F%	$d \neq 0$	conv
LISWET9	102	100	-5.5606e+00	154	5.6e-10	1.0e-05	26.6	0	73.4	0	1
LISWET10	103	100	-2.5997e+01	60	4.0e-10	1.0e-05	91.7	0	8.33	0	1
LISWET11	103	100	-2.4788e+01	364	2.2e-15	1.0e-06	26.1	0	73.9	0	1
LISWET12	103	100	-5.6573e+00	461	1.4e-17	1.0e-09	40.3	0	59.7	0	1
MOSARQP1	100	10	-1.5420e+02	28	1.6e-13	1.0e-05	100	0	0	0	1
MOSARQP2	100	10	-2.0652e+02	26	3.0e-16	1.0e-05	100	0	0	0	1
NCVXBQP1	100	0	-1.9956e+06	273	1.0e-07	1.0e-08	38.1	0	61.9	194	1
NCVXBQP2	100	0	-5.5390e+04	1000	1.0e-04	1.0e-05	1.5	0	98.5	1000	0
NCVXBQP3	100	0	-5.8559e+04	1000	1.0e-04	1.0e-05	1.3	0	98.7	1000	0
NCVXQP1	100	50	-7.2975e+05	410	1.0e-05	1.0e-06	20.2	0	79.8	232	1
NCVXQP2	100	50	-5.4468e+05	287	1.0e-06	1.0e-07	40.1	0	59.9	197	1
NCVXQP3	100	50	-2.7833e+05	532	1.0e-09	1.0e-10	36.7	0	63.3	404	1
NCVXQP4	100	25	-9.1739e+05	298	1.0e-06	1.0e-07	31.5	0	68.5	150	1
NCVXQP5	100	25	-6.3385e+05	313	1.0e-06	1.0e-07	38.7	0	61.3	194	1
NCVXQP6	100	25	-3.3018e+05	340	1.0e-07	1.0e-08	36.8	0	63.2	244	1
NCVXQP7	100	75	-4.9111e+05	433	1.0e-06	1.0e-07	26.1	0	73.9	72	1
NCVXQP8	100	75	-3.4298e+05	656	1.0e-06	1.0e-07	20.9	0.152	79	260	1
NCVXQP9	100	75	-2.1121e+05	178	1.0e-06	1.0e-07	48.9	0	51.1	102	1
NOBNDTOR	100	0	-5.5211e-01	24	1.0e-04	1.0e-05	100	0	0	0	1
OBSTCLAE	100	0	1.3979e+00	28	3.2e-11	1.0e-05	100	0	0	0	1
OBSTCLAL	100	0	1.3979e+00	53	2.6e-11	1.0e-05	100	0	0	0	1
OBSTCLBL	100	0	2.8750e+00	21	1.0e-04	1.0e-05	100	0	0	0	1
OBSTCLBM	100	0	2.8750e+00	17	1.0e-04	1.0e-05	100	0	0	0	1
OBSTCLBU	100	0	2.8750e+00	21	1.0e-04	1.0e-05	100	0	0	0	1
PENTDI	100	0	-7.5000e-01	1000	1.0e-09	1.0e-05	10.4	0	89.6	0	0
PORTSNQP	100	2	3.1801e+01	35	2.1e-12	1.0e-05	100	0	0	3	1
PORTSQP	100	1	3.1587e+01	25	3.5e-11	1.0e-05	100	0	0	0	1
POWELL20	100	100	5.2703e+04	629	5.9e-11	1.0e-05	5.41	0	94.6	0	1
QUDLIN	120	0	-7.2000e+05	30	1.0e-04	1.0e-05	100	0	0	7	1
SOSQP1	200	101	4.1605e-16	19	3.2e-05	1.0e-05	100	0	0	0	1
SOSQP2	200	101	-4.8738e+01	202	1.0e-04	1.0e-05	24.3	0	75.7	0	1
STCQP1	257	128	4.0405e+03	32	1.0e-04	1.0e-05	100	0	0	0	1
STCQP2	257	128	1.4294e+03	28	4.3e-13	1.0e-05	100	0	0	0	1
STNQP1	257	128	-4.4730e+03	22	1.0e-04	1.0e-05	100	0	0	1	1
STNQP2	257	128	-7.2320e+03	24	1.0e-04	1.0e-05	100	0	0	2	1
TESTQUAD	1000	0	0.0000e+00	1	0.0e+00	1.0e-05	100	0	0	0	1
TORSION1	100	0	-4.9234e-01	26	1.0e-04	1.0e-05	100	0	0	0	1
TORSION2	100	0	-4.9234e-01	24	1.0e-04	1.0e-05	100	0	0	0	1
TORSION3	100	0	-1.2705e+00	23	1.0e-04	1.0e-05	100	0	0	0	1
TORSION4	100	0	-1.2705e+00	22	1.0e-04	1.0e-05	100	0	0	0	1
TORSION5	100	0	-2.8971e+00	25	1.0e-04	1.0e-05	100	0	0	0	1
TORSION6	100	0	-2.8971e+00	24	1.0e-04	1.0e-05	100	0	0	0	1
TORSIONA	100	0	-4.0570e-01	27	1.0e-04	1.0e-05	100	0	0	0	1
TORSIONB	100	0	-4.0570e-01	25	1.0e-04	1.0e-05	100	0	0	0	1
TORSIONC	100	0	-1.1766e+00	23	1.0e-04	1.0e-05	100	0	0	0	1

Table 6.2.2: Results for CUTEr test QPs, Continued

QP	n	m	OptQ	Itr	μ_E^F	μ_I^F	VOL%	M%	F%	$d \neq 0$	conv
TORSIOND	100	0	-1.1766e+00	22	1.0e-04	1.0e-05	100	0	0	0	1
TORSIONE	100	0	-2.7984e+00	25	1.0e-04	1.0e-05	100	0	0	0	1
TORSIONF	100	0	-2.7984e+00	24	1.0e-04	1.0e-05	100	0	0	0	1
TRIDIA	100	0	-1.0000e+00	1	3.1e-19	1.0e-05	100	0	0	0	1
UBH1	99	60	1.1474e+00	160	9.5e-10	1.0e-05	74.4	0	25.6	0	1
YAO	202	200	8.1384e+00	1000	9.3e-06	1.0e-06	7.1	0.1	92.8	0	0

Bibliography

[1] J. M. Borwein. Necessary and sufficient conditions for quadratic minimality. *Numer. Funct. Anal. and Optimiz.*, 5:127–140, 1982.

[2] J. R. Bunch and L. Kaufman. Some stable methods for calculating inertia and solving symmetric linear systems. *Math. Comput.*, 31:163–179, 1977.

[3] J. R. Bunch, L. Kaufman, and B. N. Parlett. Decomposition of a symmetric matrix. *Numer. Math.*, 27:95–109, 1976.

[4] J. R. Bunch and B. N. Parlett. Direct methods for solving symmetric indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 8:639–655, 1971.

[5] L. B. Contesse. Une caractérisation complète des minima locaux en programmation quadratique. *Numer. Math.*, 34:315–332, 1980.

[6] Anthony V. Fiacco and Garth P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley and Sons, Inc., New York-London-Sydney, 1968.

[7] Anders Forsgren and Philip E. Gill. Primal-dual interior methods for nonconvex nonlinear programming. *SIAM J. Optim.*, 8:1132–1152, 1998.

[8] Anders Forsgren, Philip E. Gill, and Joseph R. Shinnerl. Stability of symmetric ill-conditioned systems arising in interior methods for constrained optimization. *SIAM J. Matrix Anal. Appl.*, 17:187–211, 1996.

[9] Anders Forsgren, Philip E. Gill, and Margaret H. Wright. Interior methods for nonlinear optimization. *SIAM Rev.*, 44:525–597, 2002.

[10] Anders Forsgren and Walter Murray. Newton methods for large-scale linear equality-constrained minimization. *SIAM J. Matrix Anal. Appl.*, 14:560–587, 1993.

[11] Philip E. Gill and Walter Murray. Newton-type methods for unconstrained and linearly constrained optimization. *Math. Program.*, 7:311–350, 1974.

[12] Philip E. Gill and Daniel P. Robinson. A primal-dual augmented Lagrangian. *Computational Optimization and Applications*, 51:1–25, 2012.

[13] Philip E. Gill and Daniel P. Robinson. A globally convergent stabilized SQP method. *SIAM J. Optim.*, 23(4):1983–2010, 2013.

[14] John Greenstadt. On the relative efficiencies of gradient methods. *Math. Comp.*, 21:360–367, 1967.

[15] M. R. Hestenes. Multiplier and gradient methods. *J. Optim. Theory Appl.*, 4:303–320, 1969.

[16] Antal Majthay. Optimality conditions for quadratic programming. *Math. Programming*, 1:359–365, 1971.

[17] O. L. Mangasarian and S. Fromovitz. The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. *J. Math. Anal. Appl.*, 17:37–47, 1967.

[18] Garth P. McCormick. Anti-zig-zagging by bending. *Management Science*, 15(5):315–320, 1969.

[19] Jorge J. Moré and Danny C. Sorensen. On the use of directions of negative curvature in a modified Newton method. *Math. Program.*, 16:1–20, 1979.

[20] P. M. Pardalos and G. Schnitger. Checking local optimality in constrained quadratic programming is NP-hard. *Oper. Res. Lett.*, 7(1):33–35, 1988.

[21] Panos M. Pardalos and Stephen A. Vavasis. Quadratic programming with one negative eigenvalue is NP-hard. *J. Global Optim.*, 1(1):15–22, 1991.

[22] R. Polyak. Modified barrier functions (theory and methods). *Math. Program.*, 54(2, Ser. A):177–222, 1992.

[23] Michael J. D. Powell. A method for nonlinear constraints in minimization problems. In Roger Fletcher, editor, *Optimization*, pages 283–298, London and New York, 1969. Academic Press.

[24] Daniel P. Robinson. *Primal-Dual Methods for Nonlinear Optimization*. PhD thesis, Department of Mathematics, University of California San Diego, La Jolla, CA, 2007.

[25] R. B. Schnabel and E. Eskow. A new modified Cholesky factorization. *SIAM J. Sci. and Statist. Comput.*, 11:1136–1158, 1990.

[26] Margaret H. Wright. Why a pure primal Newton barrier step may be infeasible? *SIAM J. Optim.*, 5(1):1–12, 1995.

[27] Margaret H. Wright. Ill-conditioning and computational error in interior methods for nonlinear programming. *SIAM J. Optim.*, 9(1):84–111, 1998.