INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600
RICE UNIVERSITY

A Modified Augmented Lagrangian Merit Function, and $Q$-Superlinear Characterization Results for Primal-Dual Quasi-Newton Interior-Point Method for Nonlinear Programming

by

Zeferino Parada García

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

Approved. Thesis Committee:

Richard A. Tapia. Chairman
Noah Harding Professor of Computational and Applied Mathematics

Thomas A. Badgwell
Assistant Professor of Chemical Engineering

William W. Symes
Professor of Computational and Applied Mathematics

Houston, Texas
April. 1997
Abstract

A Modified Augmented Lagrangian Merit Function, and $Q$-Superlinear Characterization Results for Primal-Dual Quasi-Newton Interior-Point Method for Nonlinear Programming

by

Zeferino Parada García

Two classes of primal-dual interior-point methods for nonlinear programming are studied. The first class corresponds to a path-following Newton method formulated in terms of the nonnegative variables rather than all primal and dual variables. The centrality condition is a relaxation of the perturbed Karush-Kuhn-Tucker condition and primarily forces feasibility in the constraints. In order to globalize the method using a linesearch strategy, a modified augmented Lagrangian merit function is defined in terms of the centrality condition. The second class is the Quasi-Newton interior-point methods. In this class the well known Boggs-Tolle-Wang characterization of $Q$-superlinear convergence for Quasi-Newton method for equality constrained optimization is extended. Critical issues in this extension are: the choice of the centering parameter, the choice of the steplength parameter, and the choice of the primary variables.
Acknowledgments

I would like to dedicate this dissertation to my mother Asunción who has always believed in education. Also I dedicate this dissertation to the memory of my father Zeferino and to my niece América Stephanie.

I would like to thank my sisters, brothers, sisters-in-law, and brothers-in-law for encouraging me during the years in Houston. To my nieces and nephews I would like to say that I missed many of their birthdays during my studies here at Rice, but I always keep them in my heart.

My profound gratitude goes to my advisor, Professor Richard Tapia who gave me the opportunity to enter the great world of Optimization Theory. He was generous to me in sharing his extensive teaching and research experiences. They enriched my professional career. Certainly, I could not finished this dissertation without his support, encouragement and advice. Professor Tapia was also a friend at Rice together with his charming wife Gina. The only words I have for them are: "Muchas Gracias".

My respect and profound thanks to the members of my committee, Professor Thomas Badgwell and Professor William Symes for their time, attention and comments to this dissertation.

I would like to thank Professor Héctor Martínez from La Universidad del Valle in Cali, Colombia for discussing in detail part of this research.

I appreciate the open attitude of Miguel Argáez for sharing with me his ideas in interior-point method, salsa music, and soccer.

Special thanks to Professor Amr El-Bakry for encouraging me to follow the ideas of this dissertation at the beginning of my research.

Also I want to thank Leticia Velázquez for exchanging ideas about computational
issues in interior-point methods.

I will be all my life in debt to my supportive friends in Mexico. They are Professor Virginia Abrín and Professor Pablo Barrera. This dissertation is a small repayment to their support, encouragement and friendship even since I was an undergraduate student at La Facultad de Ciencias, UNAM, in Mexico City.

I would like to thank the Department of Computational and Applied Mathematics CAAM and the Center for Research on Parallel Computation at Rice University for their support during my graduate career. This research was sponsored in part by the Department of Energy Grant DOE DE-FG03-93ER25178.
Contents

Abstract ii
Acknowledgments iii
List of Illustrations vii
List of Tables viii

1 Introduction 1

2 Preliminaries 4
  2.1 The Nonlinear Programming Problem ................. 4
  2.2 Definitions and Terminology .......................... 6
  2.3 Interpretation of the Perturbed KKT Conditions ......... 6
  2.4 The Logarithmic Barrier Function Method .............. 7
  2.5 The Philosophy of Primal-Dual Interior-Point Method ... 12
    2.5.1 The Perturbation Parameter ....................... 12
    2.5.2 The Steplength Parameter ......................... 13
    2.5.3 Path-Following Strategy .......................... 13
    2.5.4 Merit Function ................................... 14

3 A Modified Augmented Lagrangian Merit Function 15
  3.1 The Function ......................................... 15
  3.2 Descent Direction .................................... 17
  3.3 The Penalty Parameter ................................ 19

4 Path-Following Primal-Dual Interior-Point Method 21
4.1 Centrality Condition .................................................. 21
4.2 The Method ................................................................. 22
4.3 Updating the Penalty Parameter ................................. 24
4.4 Steplength Parameter .................................................. 24

5 Global Convergence Theory ........................................... 26
5.1 Assumptions .............................................................. 26
5.2 Inner Loop Exit .......................................................... 26
5.3 Global Convergence Theorems ...................................... 32

6 Numerical Results ........................................................ 35
6.1 Implementation .......................................................... 35
6.2 Numerical Experience .................................................. 36
6.3 Comments ................................................................. 37

7 Quasi-Newton Methods and a $Q$-Superlinear Result .......... 44
7.1 The Damped and Perturbed Newton Method .................. 44
7.2 Characterization for Damped and Perturbed Quasi-Newton Methods 45

8 Primal-Dual Quasi-Newton Interior-Point Methods .......... 49
8.1 The Method .............................................................. 49
8.2 An Equivalent Formulation ........................................ 51
8.3 $Q$-superlinear Convergence Characterization ................ 55

9 Concluding Remarks ..................................................... 57

Bibliography ................................................................. 58
Illustrations

6.1 The norm of the KKT conditions for the two strategies on Problem 81 \hspace{1cm} 41
6.2 The norm of the KKT conditions for the two strategies on Problem 104 \hspace{1cm} 41
6.3 The norm of the constraints for the two strategies on Problem 81 \hspace{1cm} 42
6.4 The norm of the constraints for the two strategies on Problem 104 \hspace{1cm} 42
Tables

6.1 Hock and Schittkowski test problems. The symbol '-' means no convergence. .................................................. 39
6.2 Hock and Schittkowski test problem (Continued). The symbol '-' means no convergence. ................................. 40
6.3 The role of the centrality condition on the penalty parameter. .... 43
Chapter 1

Introduction

Due to the computational success of the primal-dual interior-point method for Linear Programming (LP), recently there has been much activity proposing extensions for the more difficult case of Nonlinear Programming (NLP). In LP the primal-dual interior-point method, although not initially presented in this manner, is now recognized as a damped and perturbed Newton method applied to the Karush-Kuhn-Tucker (KKT) necessary conditions. This interpretation serves as the vehicle for its extension to NLP. There are two topics to be considered in formulating primal-dual interior-point methods for NLP that do not appear in LP. The first is the use of appropriate path-following strategies and merit functions for the primal-dual Newton method. The second is to replace the Hessian of the Lagrangian function by a matrix approximation when the second order derivatives are expensive to compute. This latter strategy has the potential of causing the fast convergence of Newton's method to deteriorate. Hence it is desirable to characterize those methods that generate $Q$-superlinear iterates in terms of their parametric choices. This dissertation investigates both topics separately.

In 1995 Argáez and Tapia [2] defined a centrality condition for primal-dual Newton methods consisting of the equality constraints from the NLP problem and the perturbed complementarity equation given in the KKT conditions. Hence their formulation includes only the nonnegative variables involved in the KKT conditions. Their centrality condition is a relaxation of the more restrictive centrality condition given by the perturbed Karush-Kuhn-Tucker (KKT) conditions. Implementations of the path-following primal-dual Newton method based on their centrality condition have
a better chance of meeting the centrality condition than do those methods whose path-following strategy is formulated by the perturbed KKT conditions. since we know that a perturbed KKT point may not exist for all choices of the parameter. In order to exploit the Argáez- Tapia centrality condition, an appropriate merit function must be used. This merit function must primarily enforce constraint satisfaction in the NLP problem. In this dissertation we propose a modified augmented Lagrangian merit function such that the augmentation term is the Argáez- Tapia centrality condition. The Newton step given for the perturbed KKT conditions becomes a descent direction of our modified augmented Lagrangian function. This simple fact permit us to develop a path-following primal-dual method for solving NLP using linesearch globalization.

The second part of this dissertation addresses the problem of replacing the Hessian matrix of the Lagrangian by a matrix approximation in the primal-dual interior-point method. Our interpretation of the primal-dual method is to view it as a damped and perturbed Quasi-Newton method applied to the KKT conditions. In 1993 Yamashita and Yabe [45] used the Dennis and Moré $Q$-superlinear result [13], to characterize primal-dual Quasi-Newton methods that gave $Q$-superlinear convergence in terms of all primal and dual variables involved in the KKT condition. However, we believe that this task is incomplete since we know that for the Equality Constrained Optimization Problem there exists a $Q$-superlinear characterization for the corresponding Quasi-Newton methods which is given in terms of the primal variable $x$ alone (see Bogg, Tolle, and Wang [5]). Then the primary variable for Quasi-Newton methods for Equality Constraints Optimization is the primal variable $x$. This understanding led us initially to try to obtain a characterization for primal-dual Quasi-Newton interior-point methods in terms of the primal variable $x$ alone. However, we could not do so without including an undesirable assumption on the interaction between the primal
variable and the dual nonnegative slack variable $z$. This in turn led us to search for a characterization in terms of both variables, the primal variable and the dual nonnegative variable under the standard Newton method assumptions. It is interesting then, that in the sense alluded to above the primary variables for primal-dual Quasi-Newton methods are the primal variable and the dual nonnegative variable.

This dissertation is organized as follows: In Chapter 2 we introduce the general Nonlinear Programming problem and the philosophy of primal-dual interior-point methods. In Chapter 3 we define our modified augmented Lagrangian merit function and explore its theoretical properties. In Chapter 4 we propose a path-following primal-dual interior-point method for solving NLP. Also, we discuss how the algorithmic parameters are chosen for the method. In Chapter 5 we consider our additional assumptions to prove global convergence for the method of the previous chapter. In Chapter 6 we detail our implementation of the method from the previous chapter and present numerical results on a subset of problems from Hock and Schittkowski[28] and Schittkowski [36]. In Chapter 7 we begin the second part of the dissertation. In this chapter we establish a $Q$-superlinear characterization result for damped and perturbed Quasi-Newton methods for solving nonlinear system of equations. In Chapter 8 we define the primal-dual Quasi-Newton interior-point method for NLP and establish our $Q$-superlinear characterizations results. In Chapter 9 we make some concluding remarks.
Chapter 2

Preliminaries

In this chapter we introduce general nonlinear programming (NLP) and the main ideas of primal-dual interior-point methods for NLP.

2.1 The Nonlinear Programming Problem

We consider the standard problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad h(x) = 0 \quad x \geq 0 \\
\end{align*}
\]

(2.1)

where \( f : \mathbb{R}^n \to \mathbb{R} \) and \( h : \mathbb{R}^n \to \mathbb{R}^m \) are twice continuously differentiable and \( m \leq n \).

The Lagrangian function associated with problem (2.1) is given by

\[
l(x, y, z) = f(x) + y^T h(x) - z^T x
\]

(2.2)

where \( y \in \mathbb{R}^m \) and \( z \in \mathbb{R}^n \) are the Lagrange multipliers associated with the constraints \( h(x) = 0 \) and \( x \geq 0 \) respectively.

As is common in constrained optimization, \( x \) is called the primal variable and \((y, z)\) are called the dual variables.

The Karush-Kuhn-Tucker (KKT) conditions for problem (2.1) are

\[
F(x, y, z) = \begin{pmatrix}
\nabla_x l(x, y, z) \\
h(x) \\
XZ\varepsilon
\end{pmatrix} = 0. \quad (x, z) \geq 0.
\]

(2.3)
where $X = \text{diag}(x)$, $Z = \text{diag}(z)$ and $e \in \mathbb{R}^n$ is the vector of all ones.

Observe that the inequality constraints in (2.1) can be written as $\epsilon_i^T x \geq 0$ for $i = 1, \ldots, n$.

where the vector $\epsilon_i$, $i = 1, \ldots, n$ corresponds to the $i$-th canonical vector whose $i$-th component is one and all others are zero. For a feasible point $x$ of (2.1) we set $B(x) = \{ i \in \{1, 2, \ldots, n\} \mid \epsilon_i^T x = 0 \}$. As usual in constrained optimization $B(x)$ is the set of indexes of binding or active inequality constraints at $x$. We will have need to consider the gradients of the active constraints. It should be clear that those gradients are $\{ \epsilon_i \mid i \in B(x) \}$.

In the study of Newton’s method, the standard assumptions for problem (2.1) are

A.1. (Existence) There exists $(x^*, y^*, z^*)$ a solution to problem (2.1) and its associated Lagrange multipliers satisfying the KKT conditions (2.3).

A.2. (Smoothness) The Hessian operators $\nabla^2 f$, $\nabla^2 h_i$, $i = 1, \ldots, m$ are locally Lipschitz continuous at $x^*$.

A.3. (Regularity) The set $\{ \nabla h_i(x^*) : i = 1, \ldots, m \} \cup \{ \epsilon_i : i \in B(x^*) \}$ is linearly independent.

A.4. (Second-Order Sufficiency) For all $\eta \neq 0$ satisfying $\nabla h_i(x^*)^T \eta = 0$, $i = 1, \ldots, m$: $\epsilon_i^T \eta = 0$, $i \in B(x^*)$ we have $\eta^T \nabla_x^2 l(x^*, y^*, z^*) \eta > 0$.

A.5. (Strict Complementarity) For all $i$, $z_i^* + x_i^* > 0$.

For a nonnegative parameter $\mu$, the perturbed KKT conditions associated with (2.3) are

$$F_\mu(x, y, z) = \begin{pmatrix} \nabla_x l(x, y, z) \\ h(x) \\ XZe - \mu e \end{pmatrix} = 0. \quad (x, z) \geq 0. \quad (2.4)$$
2.2 Definitions and Terminology

In this section we introduce some definitions and terminology that will be used throughout this work.

- We say that the point $x$ is a KKT point of problem (2.1) if there exist a pair $(y, z) \in \mathbb{R}^{n+m}$ such that the triple $(x, y, z)$ satisfies the KKT conditions (2.3).

- Given $\mu > 0$, we say that $x > 0$ is a perturbed KKT point (corresponding to $\mu$) if there exist $(y, z) \in \mathbb{R}^{m+n}$ such that the triple $(x, y, z)$ satisfies the perturbed KKT conditions (2.4) at $\mu$.

- We say that the triple $(x, y, z)$ is an interior point if $(x, z) > 0$.

- (From Argáez and Tapia [2]) We say that the interior point $(x, y, z)$ is a quasi-central point corresponding to $\mu$ if $h(x) = 0$ and $XZ\epsilon = \mu \epsilon$.

- (From Argáez and Tapia [2]) The collection of interior point that are a quasi-central point corresponding to some $\mu$ is called the quasi-central path.

2.3 Interpretation of the Perturbed KKT Conditions

In (2.4) the perturbation affects only the complementarity equation of (2.3). We briefly explain the role of this particular perturbation. Observe that (2.3) is not a square nonlinear system of equations due to the nonnegativity constraints. Hence Newton's method cannot be directly applied. Even if the inequalities $(x, z) \geq 0$ are ignored, we must deal with the following flaw. Consider the complementarity equation of (2.1)

$$XZ\epsilon = 0.$$  \hspace{1cm} (2.5)

Newton's method applied to the KKT conditions (2.3) will deal with the linearized form of (2.5). Let's consider the $i$-th component of this latter equation. We obtain

$$z_i \Delta x_i + x_i \Delta z_i = -x_i z_i.$$  \hspace{1cm} (2.6)
Assuming that \( x_i = 0 \) and \( z_i \neq 0 \), equation (2.6) tells us that \( \Delta x_i = 0 \). Therefore, the \( i \)-th component of the primal variable will remain zero in future Newton iterations. If the local solution \( x^* \) of (2.1) satisfies \( x_i^* > 0 \), we will never be able to reach this solution. The way to correct for this deficiency of Newton's method is to perturb the right-hand side of (2.6) by a quantity \( \mu > 0 \). Then, the equation (2.6) becomes

\[
\Delta z_i = -x_i z_i + \mu. \tag{2.7}
\]

and \( \Delta x_i \) is no longer equal to zero. Observe that (2.7) is the linearization of the \( i \)-th component of the equation \( XZ\varepsilon - \mu \varepsilon = 0 \).

### 2.4 The Logarithmic Barrier Function Method

In this section we describe the logarithmic barrier function method for NLP. Our purpose is to review the theoretical and Newton algorithmic equivalence between the perturbed KKT conditions (2.4) and the KKT conditions of the logarithmic barrier function method.

The logarithmic barrier function method for problem (2.1), consists in solving for each positive parameter \( \mu \), the equality constrained problem

\[
\begin{align*}
\text{minimize} & \quad f(x) - \mu \sum_{i=1}^{n} \log(x_i) \\
\text{subject to} & \quad h(x) = 0 \\
& \quad (x > 0). \tag{2.8}
\end{align*}
\]

Suppose that \( x(\mu) = x_\mu \) is a solution of (2.8). Under mild assumptions (see Fiacco and McCormick [17]) the collection of points \( \{x_\mu : \mu \geq 0\} \) defines a trajectory such that

\[
x_\mu \to x^* \quad \text{as} \quad \mu \to 0.
\]
where \( x^* \) is a local solution of problem (2.1).

The logarithmic barrier function method is the first known interior point method for solving the minimization problem (2.1). An interior-point method means that the variable \( x \) must remain in the interior of the set \( \{ x \geq 0 \} \). It is well known that the logarithmic barrier function method has impressive behavior far away from a local solution of (2.1), but it contains serious flaw near a binding solution of problem (2.1).

We briefly explain this flaw.

The KKT conditions of problem (2.8) are

\[
\hat{F}_\mu(x, y) = \begin{pmatrix}
    \nabla f(x) + \nabla h(x)^T y - \mu X^{-1} e \\
    h(x)
\end{pmatrix} = 0.
\]  

(2.9)

and the Jacobian of \( \hat{F}_\mu(x, y) \) is

\[
\hat{F}^\nu(x, y) = \begin{pmatrix}
    \nabla^2 f(x) + \nabla^2 h(x) y + \mu X^{-2} \nabla h(x) \\
    \nabla^T h(x)
\end{pmatrix}.
\]  

(2.10)

Let \( x^* \) be a local solution of (2.1). In order to explain the local behavior of the logarithmic barrier function method near a binding solution, we may assume that at least one component of \( x^* \) is zero. However, for the sake of simplicity we will assume that \( x^* = (0, 0, \ldots, 0)^T \) is a local solution of (2.1). Let \( y^*, z^* \) be the corresponding Lagrangian multiplier associated to \( x^* \) such that the standard assumptions A1-A5 hold at \( (x^*, y^*, z^*) \). Then for the points in the barrier trajectory we obtain

\[
\nabla f(x_{\mu}) + \nabla h(x_{\mu}) y_{\mu} - \mu X_{\mu}^{-1} \rightarrow \nabla x l(x^*, y^*, z^*), \quad \text{as} \quad \mu \rightarrow 0.
\]

and

\[
\nabla^2 f(x_{\mu}) + \nabla^2 h(x_{\mu}) y(x_{\mu}) + (\mu X_{\mu}^{-2}) \rightarrow \nabla^2 x l(x^*, y^*, z^*), \quad \text{as} \quad \mu \rightarrow 0.
\]

We necessarily have that

\[
\mu X_{\mu}^{-1} \rightarrow z^* \quad \text{as} \quad \mu \rightarrow 0.
\]
Hence
\[ \mu X_\mu^{-2} \to \infty \quad \text{as} \quad \mu \to 0. \]

So, the matrix \( F'_\mu(x_\mu, y_\mu) \) becomes ill-conditioned near \( x^* \). Notice that the bad-conditioning results from the gradient of \( \mu X^{-1}e \). If the latter expression is replacing by the auxiliary variable \( z = \mu X^{-1}e \), and we rewrite this relationship in the benign form \( XZe = \mu e \), then the KKT conditions (2.9) are transformed into the perturbed KKT conditions associated to problem (2.1) and the ill-conditioning problem is removed. The connection between the perturbed KKT conditions and the barrier KKT conditions are summarized in the following results.

**Proposition 2.1** The perturbed KKT conditions associated with problem (2.1) given by (2.4) and the KKT conditions for the logarithmic barrier function (2.8) given by (2.9) are equivalent in the sense that they have the same solution, that is, \( F'_\mu(x, y) = 0 \) if and only if \( F_\mu(x, y, \mu X^{-1}) = 0 \).

However, the equivalence in Proposition 2.1 is not extended to more theoretical properties. By a smooth optimization problem we mean a \( C^\infty \) problem.

**Proposition 2.2** The perturbed KKT conditions associated with problem (2.1) given by \( F_\mu(x, y, z) = 0 \) or any permutation of these equations, are not the KKT conditions for the logarithmic barrier function problem (2.8) or any other (smooth) unconstrained or equality constrained optimization problem.

The perturbed KKT conditions (2.4) are interpreted as the KKT conditions of (2.9) using the nonlinear transformation \( XZe = \mu e \). It is not the case that Newton’s method is invariant under this nonlinear transformation.
**Proposition 2.3** Consider $\mu > 0$ and an interior point $(x,y,z)$ such that $x_i z_i \neq \mu$ for $i = 1,...,n$. Assume that the matrices $F^\prime(x,y,z)$ and $F^\prime_\mu(x,y)$ are nonsingular. Let $(\Delta x, \Delta y, \Delta z)$ be the Newton step obtained from the nonlinear system $F(x,y,z) = 0$ given by (2.4). Let $(\Delta x^\prime, \Delta y^\prime)$ be the Newton step obtained from the nonlinear system $F^\prime(x,y) = 0$ given by (2.9). Then the following statements are equivalent

(i). $(\Delta x, \Delta y) = (\Delta x^\prime, \Delta y^\prime)$.

(ii). $\Delta x = 0$.

(iii). $\Delta x^\prime = 0$.

(iv). $x$ is a perturbed KKT point at $\mu$.

**Proof:** The Lagrangian function associated with the equality constrained optimization problem (2.8) is given by

$$\tilde{l}_\mu(x,y) = f(x) + h(x)^T y - \mu \sum_{i} \log(x).$$

The two linear systems that we are concerned with are

$$\nabla^2_x l(x,y,z) \Delta x + \nabla h(x) \Delta y - \Delta z = -\nabla_x l(x,y,z)$$  \hspace{1cm} (2.11)

$$\nabla h(x) \Delta x = -h(x)$$  \hspace{1cm} (2.12)

$$Z \Delta x + X \Delta z = -(XZ \epsilon - \mu \epsilon).$$  \hspace{1cm} (2.13)

and

$$\nabla^2_x l(x,y,z) \Delta x^\prime + \nabla h(x) \Delta y^\prime = \nabla_x \tilde{l}_\mu(x,y)$$  \hspace{1cm} (2.14)

$$\nabla h(x) \Delta x^\prime = -h(x).$$  \hspace{1cm} (2.15)

Solving for $\Delta z$ from equation (2.13), substituting in the equation (2.11) and observing that $\nabla_x l(x,y,z) + z - \mu X^{-1} \epsilon = \nabla_x \tilde{l}_\mu(x,y)$ we obtain

$$(\nabla^2_x l(x,y,z) + X^{-1}Z) \Delta x + \nabla h(x)^T \Delta y = -\nabla l(x,y).$$  \hspace{1cm} (2.16)
Proof of (i) ⇒ (ii). We observe that \( \nabla_x^2 l(x, y, z) + \mu X^{-2} = \nabla_x^2 \hat{I}_\mu(x, y) \). Hence, equation (2.14), equation (2.16) and the fact that \((\Delta x, \Delta y) = (\Delta x', \Delta y')\) imply

\[(X^{-1}Z - \mu X^{-2})\Delta x = 0.\]

Since \(x_i z_i \neq \mu\) for \(i = 1, \ldots, n\) we conclude that \(\Delta x = 0\).

Proof of (ii) ⇒ (iii). Since \(\Delta x = 0\), equation (2.16) can be written as

\[\nabla h(x)^T \Delta y = -\nabla \hat{I}_\mu(x, y).\]

Clearly, \(h(x) = 0\). Therefore \((0, \Delta y)\) solves the linear system (2.14) − (2.15) where \(\hat{F}_\mu(x, y)\) is nonsingular. In particular \(\Delta x' = 0\)

Proof of (iii) ⇒ (iv). Since \(\Delta x' = 0\), the equation (2.15) can be written as

\[\nabla f(x) + \nabla h(x)^T (y + \Delta y') - \mu X^{-1} \epsilon = 0.\]

Since \(h(x) = 0\), we conclude that \((x, y + \Delta y', \mu X^{-1} \epsilon)\) satisfies the perturbed KKT conditions corresponding to \(\mu\).

Proof of (iv) ⇒ (i). If \(x\) is a perturbed KKT point at \(\mu\) then there exists \((\tilde{y}, \tilde{z}) \in \mathbb{R}^{m+n}\) such that \((x, \tilde{y}, \tilde{z})\) satisfies the perturbed KKT conditions at \(\mu\). Therefore

\[\nabla f(x) + \nabla h(x)^T \tilde{y} - \tilde{z} = 0.\]

It follows that \((0, \tilde{y} - y, \tilde{z} - z)\) solves (2.11) − (2.13), and \((0, \tilde{y} - y)\) solves (2.14) − (2.15). Since the two linear systems have nonsingular matrices, we conclude that \((\Delta x, \Delta y) = (\Delta x', \Delta y')\).

\[\square\]

The proposition 2.3 must be interpreted correctly. It is wrong to interpret it as saying only that both Newton steps agree at a perturbed KKT point. It says more.
It says that iterates agree if and only if there is no movement in \( x \). In particular, it takes out the redundant case of already having a perturbed KKT point at \( \mu \) and we are looking for another perturbed KKT point at \( \mu \). For a perturbed KKT point \( x \) at \( \mu \) if we look for a perturbed KKT point at \( \mu' \neq \mu \), we no longer have that \( (\Delta x, \Delta y) = (\Delta x', \Delta y') \).

### 2.5 The Philosophy of Primal-Dual Interior-Point Method

The primal-dual interior-point method for NLP solves the KKT conditions (2.3) associated to the optimization problem (2.1). The vehicle is Newton’s method applied to the perturbed KKT conditions (2.4). The nonnegative condition given in (2.1) is obtained by damping the Newton step in order to generate interior point iterates. Fundamental issues for Newton’s method applied to the perturbed KKT conditions (2.4) are: the choice of the perturbation parameter, the steplength for damping the Newton step, the option of using a path-following strategy, and the choice of merit function.

#### 2.5.1 The Perturbation Parameter

In the primal-dual interior-point method, the perturbation parameter can be used to guide interior point iterates towards the solution of the KKT conditions (2.3). The choice of the perturbation parameter can depend on whether we are concerned with local or global convergence. Given a particular perturbation parameter, the first question is about existence of corresponding perturbed KKT points. Locally the answer is the affirmative. Under standard assumption A1-A5 we can invoke the Implicit Theorem of Calculus to ensure existence and uniqueness of perturbed KKT points for sufficiently small perturbation parameters. However, perturbed KKT points may not exist for large perturbation parameters.
2.5.2 The Steplength Parameter

The Newton step from the nonlinear equation $F_\mu(x, y, z) = 0$, can be damped in order to maintain interior point iterates. Certainly, damping the Newton step by a small positive scalar ensures that we obtain an interior point iterate, but convergence may deteriorate as a consequence of staying in the interior of $\{(x, z) \geq 0\}$. The steplength parameter for forcing interior point iterates is not a choice in the primal-dual interior-point method. It merely is information given for the current interior point and its Newton step. But, we may choose how far we want to move to the boundary. This choice affects the behavior of interior-point methods.

2.5.3 Path-Following Strategy

The primal-dual interior-point methods must prevent the property of sticking to the boundaries, $\epsilon_i^T x = 0$, $i = 1,..., n$. One idea is to impose a condition that forces interior-point iterates to be 'more in the interior'. Hence, path-following strategies avoid sticking to the boundaries described above by producing interior point iterates that follow a centrality condition. In general, centrality conditions are defined by information in the perturbed KKT conditions. Then a path-following strategy is obtained by fixing a perturbation parameter and applying several Newton iterations to the corresponding perturbed KKT condition until an interior point satisfies the centrality condition. We see that centering interior point iterates by a path-following strategy can deteriorate the global behavior of the method if we accurately satisfy the centrality conditions. Then, a path-following strategy can be seen as a trade-off between avoiding sticking to the boundary and fast global convergence.
2.5.4 Merit Function

The Newton method is not globally convergent. Hence a merit function must be chosen to measure progress to a solution between two iterates generated by the Newton method. No rules exist to prefer a particular merit function, but some issues can be considered in selecting it. For example, we would expect that a merit function reflects as long as possible all the information in the problem. In particular if a minimization problem is been solved, it is desired (but not required) that the merit function is also the objective function of another minimization problem. Properties such as smoothness and cheap evaluation of merit functions are important in order to save computational work. For interior-point methods a few merit functions exist for globalizing Newton's method (See El-Bakry et al [16], and Yamashita [44]). These merit functions depend of the option of path following strategy and do not satisfy all properties mentioned above. A merit function to be useful in a path-following strategy has the task of reaching the corresponding centrality condition rather than the KKT conditions. In this dissertation we will propose a novel merit function for the centrality condition given by the quasi-central path.
Chapter 3

A Modified Augmented Lagrangian Merit Function

In this chapter we define a modified augmented Lagrangian function associated with XLP problem (2.1) which will be used as a merit function in our primal-dual Newton interior-point method of Chapter 4.

3.1 The Function

We define the modified augmented Lagrangian function associated with the nonnegative perturbation parameter $\mu$ as

$$\sigma_\mu(x, y, z; C) = l(x, y, z) + \frac{C}{2} \nu_\mu(x, z).$$

(3.1)

where

$$\nu_\mu(x, z) = h(x)^T h(x) + (XZ\epsilon - \mu\epsilon)^T (XZ\epsilon - \mu\epsilon).$$

(3.2)

the function $l(x, y, z)$ is the Lagrangian function given in (2.2). and $C \geq 0$ is our penalty parameter.

Observe that (3.2) is well defined and nonnegative for all pairs $(x, z)$. Notice that our modified augmented Lagrangian function (3.1) is a generalization of the augmented Lagrangian function for equality optimization problem (see Hestenes [27]). Also, our augmented Lagrangian function function (3.1) satisfies a similar minimization property in the primal variable $x$ as the corresponding augmented Lagrangian function does for the equality constrained optimization problem.
Proposition 3.1  Let $(x_\mu, y_\mu, z_\mu)$ be a perturbed KKT point at $\mu > 0$.

Then

(i). The triple $(x_\mu, y_\mu, z_\mu)$ is a stationary point in the primal variable $x$.

of $\phi_\mu(x, y, z; C)$ for any parameter $C \geq 0$.

(ii). Moreover, there exists $C_* \geq 0$ such that the Hessian matrix

$$ \nabla_x^2 \phi_\mu(x_\mu, y_\mu, z_\mu; C) $$

is positive definite for all $C \geq C_*$. 

Proof of (i). Taking the derivative of (3.1) with respect to $x$, we obtain

$$ \nabla_x \phi_\mu(x, y, z; C) = \nabla_x l(x, y, z) + C[\nabla h(x)h(x) + Z(XX\varepsilon - \mu)] $$

therefore

$$ \nabla_x \phi_\mu(x_\mu, y_\mu, z_\mu) = 0. $$

Proof of (ii). Notice that

$$ \nabla_x^2 \phi_\mu(x_\mu, y_\mu, z_\mu; C) = \nabla_x^2 l(x_\mu, y_\mu, z_\mu) + C[\nabla h(x_\mu)^T \nabla h(x_\mu) + Z_\mu^2]. $$

since $z_\mu > 0$, there exists $C_* \geq 0$ such that $\nabla_x^2 \phi(x_\mu, y_\mu, z_\mu; C)$ is positive definite for all $C \geq C_*$. 

\[\Box\]

Corollary 3.1  There exists $C_* \geq 0$ such that

$$ x_\mu = \arg \min_x \phi_\mu(x, y_\mu, z_\mu; C) \text{ for all } C \geq C. $$  \hspace{1cm} (3.3)

Proof. The proof follows from Proposition (3.1).

\[\Box\]
3.2 Descent Direction

In the folklore of optimization the major part of using an augmented Lagrangian is relegated to the augmentation term and the penalty parameter. Our current application is no exception. Our task is to demonstrate that the modified augmented Lagrangian function (3.1) is a merit function for Newton’s method applied to the perturbed KKT conditions (2.4). Basically, we will exploit a straightforward connection between the Newton step obtained from the perturbed KKT conditions (2.4) and the augmented function (3.2). Hence our primal-dual interior-point method of Chapter 4 will be formulated in the reduced variable \((x, z)\) instead of the triple \((x, y, z)\). Recall that the nonlinear equation \(F_\mu(x, y, z) = 0\) was defined in (2.4). For now, we assume that the Jacobian matrix \(F_\mu'(x, y, z)\) is nonsingular. The Newton step \((\Delta x, \Delta y, \Delta z)^T\) for the nonlinear equation \(F_\mu(x, y, z) = 0\) is the solution of the linear system

\[
F_\mu'(x, y, z) \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} = -F_\mu'(x, y, z) \tag{3.4}
\]

Writing out the linear system (3.4) we obtain

\[
\begin{pmatrix}
\nabla_z^2 I(w) & \nabla h(x) & -I \\
\nabla h(x)^T & 0 & 0 \\
Z & 0 & X
\end{pmatrix}
\begin{pmatrix}
\Delta x \\
\Delta y \\
\Delta z
\end{pmatrix} =
\begin{pmatrix}
\nabla_z I(w) \\
h(x) \\
XZ \epsilon
\end{pmatrix} +
\mu
\begin{pmatrix}
0 \\
0 \\
\epsilon
\end{pmatrix}. \tag{3.5}
\]

Now, we establish our basic result.

**Proposition 3.2** Let \(\mu > 0\) be a perturbation parameter. Consider an interior point \((x, y, z)\) such that \(F_\mu'(x, y, z)\) is nonsingular. Let \((\Delta x, \Delta y, \Delta z)^T\) be the Newton step obtained from the linear system (3.5).

Set \(\Delta v = (\Delta x, \Delta z)^T\). Then
(i).
\[ \nabla v_\mu(x,y)^T \Delta v = -2v_\mu(x,z) \leq 0. \] (3.6)

with equality if and only if \( h(x) = 0 \) and \( XZ \epsilon = \mu \epsilon \).

(ii). Moreover, suppose that \( v_\mu(x,z) > 0 \), then there exists a threshold real number \( \hat{C} \) such that for any \( C > \hat{C} \), the reduced Newton step \( \Delta v \) is a descent direction for the modified augmented Lagrangian function (3.1) in the sense that
\[ \nabla_{(x,z)} \phi_\mu(x,y,z;C)^T \Delta v < 0. \] (3.7)

**Proof:** (i). A straightforward calculation gives us that
\[ \nabla v_\mu(x,z)^T \Delta v = 2[h(x)^T \nabla h(x)^T \Delta x + (XZ \epsilon - \mu \epsilon)^T (Z \Delta x + X \Delta z)]. \] (3.8)

Since \((\Delta x, \Delta y, \Delta z)\) is the Newton step, in particular we have that
\[ \nabla h(x)^T \Delta x = -h(x) \]
\[ Z \Delta x + X \Delta z = -(XZ \epsilon - \mu \epsilon). \] (3.9)

Therefore our result (3.6) follows from (3.8) and (3.9).

**Proof of (ii).** Notice that (3.1), and (3.6) give us
\[ \nabla_{(x,z)} \phi_\mu(x,y,z;C)^T \Delta v = \nabla_{(x,z)} l(x,y,z)^T \Delta v - C v_\mu(x,z). \] (3.10)

Since \( \phi_\mu(x,z) > 0 \), we consider the threshold parameter
\[ \hat{C} = \frac{\nabla_{(x,z)} l(x,y,z)^T \Delta v}{v_\mu(x,z)}. \] (3.11)

If we choose \( C \) according to the formula
\[ C = \hat{C} + \rho, \text{ where } \rho > 0. \] (3.12)

we obtain from (3.10) that
\[ \nabla_{(x,z)} \phi_\mu(x,y,z;C)^T \Delta v = -\rho v_\mu(x,z) < 0. \] (3.13)
We observe that the penalty parameter in (3.12) could be a negative real number. Since we will have need for considering nonnegative penalty parameters for our modified augmented Lagrangian function (3.1), we will select our penalty parameter in a different way than (3.12).

3.3 The Penalty Parameter

Clearly, a sufficiently large penalty parameter ensures a descent direction for our modified augmented Lagrangian function. However, we have need to control the behavior of the penalty parameter from the computational and theoretical point of view. Hence we will impose a condition on the penalty parameter that reflects the structure of our modified augmented Lagrangian merit function. We point out that the penalty parameter depends on the current point \((x, y, z)\) and the reduced Newton step \(\Delta v = (\Delta x, \Delta z)\). Then according to Proposition 3.2, we select the penalty parameter as the solution of the linear program

\[
\begin{align*}
\text{minimize} & \quad C \\
\text{s. t.} & \quad \nabla_{(x, z)}\phi_{\mu}(x, y, z; C)^T \Delta v \leq -[|\nabla_{(x, z)} I(x, y, z)^T \Delta v| + 2\epsilon_{\mu}(x, z)].
\end{align*}
\]

(3.14)

The linear constraint in (3.14) is the condition we impose on the penalty parameter. This condition states that at least the rate of decrease along the reduced Newton step is bounded above for the rate of decrease of each component on our modified augmented Lagrangian merit function.

The minimization problem (3.14) has a positive solution given by

\[
C^* = 2 \left\{ \frac{|\nabla_{(x, z)} I(x, y, z)^T \Delta v| + 1}{\epsilon_{\mu}(x, z)} \right\}.
\]

(3.15)
where $|.|_+$ is the real function defined by

$$|r|_+ = \begin{cases} r & \text{if } r \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.16)$$

It is worth noticing that the linear constraint in (3.14) is binding at $C_*$. 
Chapter 4

Path-Following Primal-Dual Interior-Point Method

In this chapter we present our interior-point method for solving the optimization problem (2.1).

4.1 Centrality Condition

We will adopt the notion of centrality defined as the quasi-central path by Argáez and Tapia [2]. The quasi-central path is the set of the interior points \((x, y, z)\) such that \(h(x) = 0\) and \(XZe = \mu e\) for some \(\mu > 0\). This quasi-central path is a relaxation of the more restrictive condition of a perturbed KKT point. For a fixed perturbation parameter \(\mu\) we do not intend to find a quasi-central point, because the process leads to an impractical or a costly method. In fact, if the fixed perturbation parameter \(\mu\) is relative large we are not interested in one of its corresponding quasi-central path points. Since the perturbed KKT conditions will be a guide towards obtaining a KKT point, we will follow the accepted scheme of shrinking neighborhoods around the centrality condition (See Anstreicher and Vial [1], Yamashita [44], González-Lima [25], and Argáez and Tapia [2]). So, we will attempt to find for a fixed \(\mu\) an interior point in the set

\[ N(\mu; \gamma) = \{(x, y, z) \mid \|h(x)\|_2^2 + \|XZe - \mu e\|_2^2 \leq \gamma \mu \}. \tag{4.1} \]

where \(\gamma\) is a constant in \((0, 1)\).
4.2 The Method

In this section we present our path-following primal-dual Newton interior-point method. Basically, the method is a damped and perturbed Newton method applied to the perturbed KKT conditions. We will use a path-following strategy based on the quasi-central path. As a globalization strategy we will utilize a linesearch on our modified augmented Lagrangian merit function (3.1). The method will consist of the following general steps: choose a perturbation parameter $\mu$ and then find an interior point in (4.1) using Newton's method on the perturbed KKT conditions. Then, decrease the value of $\mu$ and continue the process until a stopping criteria based on the KKT conditions is achieved. For sake of clarity, the parametric choices are specified in subsequent sections. Recall that $F(x, y, z)$ is the residual function given by (2.3).

Algorithm 1 (Path-Following Primal-Dual Newton Interior-Point Method)

Let $w_0 = (x_0, y_0, z_0)$ be an initial interior point. Let $p, J, \gamma \in (0, 1)$ be fixed parameters. Set $k = 0$, $v_k = (x_k, z_k)$, and $\mu_{k-1} = 0$.

Step 1. Test for convergence using $F(w_k)$.

Step 2. Set $\mu_k = \sigma_k v_{\mu_{k-1}}(v_k)$, where $\sigma_k \in (0, 1)$.

Step 3. Set $l = 0$ and $w^l = w_k$.

Step 4. (Inner loop) If $w^l \in N(\mu_k, \gamma)$ go to Step 5.

4.1. Find $\Delta w^l = (\Delta x^l, \Delta y^l, \Delta z^l)^T$ as a solution of the linear system

$$F'_{\mu_k}(w^l)\Delta w^l = -F_{\mu_k}(w^l).$$

(4.2)

4.2. Compute the penalty parameter $C^l$ such that (3.7) holds.

4.3. Choose $\alpha^l$ such that $w_k + \alpha^l \Delta w^l$ is an interior point.

4.4. (Backtracking) Find the first natural number $s$ for which the steplength
\[ \alpha^l = p^l \tilde{\alpha}^l \text{ satisfies} \]

\[ o_{\mu_k}(v^l + \alpha^l \Delta v^l, y^l; C^l) \leq o_{\mu_k}(w^l; C^l) + 3\alpha^l \nabla_v o_{\mu_k}(w^l; C^l)^T \Delta v^l \quad (4.3) \]

where \( v^l = (x^l, z^l) \) and \( \Delta v^l = (\Delta x^l, \Delta z^l) \).

**4.5.** Set \( w^{l+1} = w^l + \alpha^l \Delta w^l \)

\( l \leftarrow l + 1 \). go to **Step 4**.

**Step 5.** Set \( w_{k+1} = w^l \). where \( l \) is the last index in Step 4.

\( k \leftarrow k + 1 \). go to **Step 1**.

The Algorithm 1 generates two different classes of iterates. One class corresponds to the path following strategy defined by Step 4.1 - Step 4.5. and its goal is to approximate our centrality condition. The second class is the outer loop iterates indexed by \( k \). The parametric choices of Algorithm 1 are \( \sigma_k, C^l, \) and \( \alpha^l \). The parameter \( \sigma_k \) tells us how much centering we expect in the next outer iterate. The penalty parameter \( C^l \) indicates the modified augmented Lagrangian merit function \((3.1)\) to be used in the backtracking process of Step 4.4. The steplength parameter \( \alpha^l \) points out how close we want to be to the boundaries.

Indeed, Argáez and Tapia [2] established Algorithm 1 for NLP problem \((2.1)\) using a different modified augmented Lagrangian function and a slightly modified neighborhood of the quasi-central path. Similar interior-point methods to Algorithm 1 have been used before in constrained optimization. Yamashita [41] proposed a global path-following interior-point method for problem \((2.1)\). His method is entirely formulated in the primal variable \( x \). Anstreicher and Vial [1] established a path following primal-dual interior-point method for convex programming. They also exploited a straightforward relation between the Newton step and a potential merit function as we did with our modified augmented Lagrangian merit function. However, their
method can not be directly generalized to NLP, because it requires the existence of perturbed KKT points for relative large $\mu$.

4.3 Updating the Penalty Parameter

We propose a positive monotone nondecreasing penalty parameter update for the Step 4.2 in Algorithm 1. Basically, our penalty parameter choice will serve to prove the global convergence theory of Algorithm 1. Recall that the perturbation parameter is $\mu_k$. we update the penalty parameter at the inner iteration $l$ with the following scheme:

**Algorithm 2** (Penalty Parameter Update)

Let $C^{l-1}$ be the previous penalty parameter. Let $(x^l, y^l, z^l)$ be the current interior point

(1) Compute $C_{trial}$ according to the formula (3.15).

(2) Set

$$C^l = \begin{cases} 
C_{trial} & \text{if } C^{l-1} \leq C_{trial} \\
C^{l-1} & \text{otherwise.}
\end{cases}$$

Hence the penalty parameter $C^l$ satisfies

$$\nabla_{(x,z)} \hat{O}_{\mu_k}(x^l, y^l, z^l; C^l) \Delta v \leq \nabla_{(x,z)} \hat{O}_{\mu_k}(x^l, y^l, z^l; C_{trial}) \Delta v.$$

So, our penalty parameter $C^l$ is a feasible point of the linear program (3.14) defined at $(x^l, y^l, z^l)$ and $\mu_k$.

4.4 Steplength Parameter

We imitate the steplength parameter update given by El-Bakry et al [16]. This update will enforce that limit points produced by the inner loop (steps 4.1 – 4.3) are interior
points. We will have need to consider the nonlinear function

\[ G_\mu(x, z) = \begin{pmatrix} h(x) \\ XZ\epsilon - \mu e \end{pmatrix}. \]  

(4.5)

For the sake of clarity, we suppress the subindex \( k \) in the perturbation parameter and the superindex \( l \) at the current point.

For any steplength parameter \( \alpha \in (0, 1) \), we consider the update

\[ (x(\alpha), y(\alpha), z(\alpha)) = (x, y, z) + \alpha(\Delta x, \Delta y, \Delta z). \]

For a given initial interior point \((x_0, y_0, z_0)\) in the inner loop, we set

\[ \tau_1 = \min \frac{(X_0Z_0)e}{\|x_0^Tz_0\|}, \quad \tau_2 = x_0^Tz_0/\|G_\mu(x_0, z_0)\|_2. \]

We define the following functions

\[ g_1(\alpha) = \min (X(\alpha)Z(\alpha)e - \delta \tau_1 x(\alpha)^Tz(\alpha))/n. \]

and

\[ g_2(\alpha) = x(\alpha)^Tz(\alpha) - \delta \tau_2 \|G_\mu(x(\alpha), z(\alpha))\|_2. \]

where \( \delta \in (0, 1) \) is a constant.

**Algorithm 3 (Steplength Parameter)**

1. Compute for \( j = 1, 2 \),

\[ \alpha_j = \max \{ \alpha \in [0, 1] : g_j(\alpha') \geq 0 \text{ for all } \alpha' \leq \alpha \}. \text{ for } j = 1, 2. \]

2. Set \( \bar{\alpha} = \min (\alpha_1, \alpha_2). \)

More details about the functions \( g_j(\alpha) \) and a proof that \( \alpha_j > 0 \) for \( j = 1, 2 \) can be found in El-Bakry et al [16].
Chapter 5

Global Convergence Theory

In this Chapter we establish our global convergence theory for the primal-dual Newton interior-point method of Chapter 4.

5.1 Assumptions

In addition to the standard Newton method assumptions. A1-A5 in Chapter 3, we consider the following assumptions for our global convergence theory.

B1.- (Smoothness) The functions \( f(x) \) and \( h(x) \) are twice continuously differentiable. Moreover, the function \( h(x) \) is Lipschitz continuous for \( x \geq 0 \).

B2.- (Regularity) \( \nabla h(x) \) has full column rank for all \( x \geq 0 \).

B3.- The matrix \( \nabla^2 l(x, y, z) + X^{-1}Z \) is nonsingular and positive definite on the subspace \( \{ u : \nabla h(x)^Tu = 0 \} \) for \( x > 0 \).

B4.- (Boundedness) For fixed \( \mu \), the inner loop defined by Steps 4.1 - 4.5 without the stopping criteria given in Step 4, generates an inner iteration sequence \( \{(x^i, y^i, z^i)\} \) such that the sequence \( \{(x^i, z^i)\} \) is bounded.

In our assumption B4, the boundedness of \( \{x^i\} \) can be enforced by box constraints, \(-M \leq x \leq M\), for sufficiently large \( M > 0 \).

5.2 Inner Loop Exit

In this section we will demonstrate that the inner loop (Steps 4.1-4.5) generate at least one interior point in our neighborhood around the quasi-central path. We will follow a standard technique for proving global convergence for similar interior-point methods
(see El-Bakry et al [16]). Toward this end let us define for a fixed perturbation parameter $\mu$ and for $\varepsilon \geq 0$, the set

$$\Omega(\varepsilon) = \{(x, z) : \varepsilon \leq \varepsilon_{\mu}(x, z), \min (XZ\varepsilon) \geq \tau_1(x^T z) / 2n \cdot x^T z \geq \tau_2 \|G_{\mu}(x, z)\|_2 / 2\}.$$ 

Certainly our set $\Omega(\varepsilon)$ depends on $\mu$, but we will not write this dependency and assume that it is clear from the context. The set $\Omega(\varepsilon)$ will be the tool to demonstrate that we will obtain and interior point inside the neighborhood (4.1) of the quasi-central path.

The following observations are in order.

**O1.** $\Omega(\varepsilon)$ is a closed set.

**O2.** $\{x^l, z^l\} \subseteq \Omega(0)$, where $\{x^l, y^l, z^l\}$ is the inner iteration sequence.

**O3.** For $\varepsilon > 0$, and $(x, z) \in \Omega(\varepsilon)$. $x^T z$ is uniformly bounded away from zero.

**O4.** For $\varepsilon > 0$, and $(x, z) \in \Omega(\varepsilon)$. $XZ\varepsilon$ is uniformly bounded away from zero.

We will focus our attention on proving that whenever the inner iteration sequence $\{x^l, y^l, z^l\}$ satisfies

$$\{x^l, z^l\} \subseteq \Omega(\varepsilon) \quad \text{for some } \varepsilon > 0,$$

then the Newton step sequence $\{\Delta x^l, \Delta y^l, \Delta z^l\}$ is bounded and the steplength parameter sequence $\{\alpha^l\}$ is bounded away from zero. We begin by stating some useful results.

**Lemma 5.1** The iteration sequence $\{x^l, y^l, z^l\}$ is bounded.

**Proof:** By the smoothness of $f$ and $h$ (assumption B1), and regularity on $h$ (assumption B2) we obtain

$$y^l = [\nabla h(x^l)^T \nabla h(x^l)]^{-1} \nabla h(x^l) \nabla_x l(x^l, y^l, z^l) - \nabla f(x^l) + z^l.$$
Now, appealing to the boundedness of \(\{(x^l, z^l)\}\) (assumption B4), we conclude that there exists a constant \(M\) such that
\[
\|y^l\| \leq M.
\]

\[\square\]

**Lemma 5.2** In \(\Omega(\epsilon)\) the sequence \(\{x^l, z^l\}\) is bounded component-wise away from zero.

**Proof:** By definition of \(\Omega(\epsilon)\), for component index \(i\) we have
\[
[x^l]_i[z^l]_i \geq \psi_\mu(x^l, z^l)\tau_1 / 2 \geq \epsilon \tau_1 / 2 > 0.
\]
Hence \(\{[x^l]_i\}\) bounded implies \(\{[z^l]_i\}\) bounded away from zero. Now invoking assumption B4 the result follows.

\[\square\]

**Lemma 5.3** If \(\{x^l, z^l\} \subset \Omega(\epsilon)\), then \(\{[F_\mu(x^l, y^l, z^l)]^{-1}\}\) is bounded.

**Proof:** For the sake of clarity we suppress the superscript \(l\) and the arguments of functions in the proof. Recall that \(F'_{\mu} = F'\). We know that the Jacobian matrix
\[
F' = \begin{pmatrix}
\nabla^2 l & \nabla h & -I \\
\n\nabla h^T & 0 & 0 \\
Z & 0 & X
\end{pmatrix},
\]
is nonsingular if and only if the matrix
\[
\begin{pmatrix}
\nabla^2 l + X^{-1}Z & \nabla h \\
\n\nabla h^T & 0
\end{pmatrix}
\]
is nonsingular. The latter matrix is well known to be nonsingular under assumptions B2 and B3. This equivalence also states that the Newton step given in Step 4.1 is
well defined for interior points. Now, we compute \([F']^{-1}\). Rearranging the order of the rows and columns of \(F'\), we obtain the matrix.

\[
F' = \begin{pmatrix}
A & B \\
B^T & 0
\end{pmatrix}
\]

where

\[
A = \begin{pmatrix}
\nabla^2_x l & -I \\
Z & X
\end{pmatrix}, \quad B^T = (\nabla h^T \ 0).
\]

Under assumptions B2, B4, and Lemma 6.2 we have that \(A^{-1}\) exists. Moreover

\[
A^{-1} = \begin{pmatrix}
XH^{-1}X^{-1} & H^{-1}X^{-1} \\
\nabla^2_x l H^{-1} - I & \nabla^2_x l H^{-1}X^{-1}
\end{pmatrix}
\]

where \(H = \nabla^2_x l + X^{-1}Z\).

Finally, a straightforward calculation give us

\[
[F']^{-1} = \begin{pmatrix}
A^{-1} - A^{-1}B(B^TA^{-1}B)^{-1}A^{-1}B(B^TA^{-1}B)^{-1} & A^{-1}B(B^TA^{-1}B)^{-1} \\
(B^TA^{-1}B)^{-1}B^TA^{-1} & -(B^TA^{-1}B)^{-1}
\end{pmatrix}.
\] (5.1)

Since the sequence of inner iterates is bounded (Lemma 6.1), assumptions B1 and B2 imply boundedness for each matrix in (5.1). Hence we obtain our result.

\(\square\)

**Corollary 5.1** If \((x^i, z^i) \subseteq \Omega(\epsilon)\), then the Newton direction sequence \(\{\Delta x^i, \Delta y^i, \Delta z^i\}\) is bounded.

**Proof:** From the linear system in Step 4.1 we have that

\[
\|\Delta x^i, \Delta y^i, \Delta z^i\|_2 \leq \|F^{-1}_\mu(x^i, y^i, z^i)\|_2 \|F_\mu(x^i, y^i, z^i)\|_2.
\]

The result follows from Lemma 6.2.

\(\square\)
Lemma 5.4  Assume that \( \{(x^l, z^l)\} \subset \Omega(\varepsilon) \). Then \( \{\alpha^l\} \) is bounded away from zero.

Proof: See Lemma 6.3 in El-Bakry et al [16].

\( \square \)

Now we establish our main result of this section.

Theorem 5.1  Consider \( \mu > 0, \gamma \in (0, 1) \), and \((x^0, y^0, z^0)\) an interior point. Let \( \{(x^l, y^l, z^l)\} \) be the sequence generated by the inner loop (steps 4.1 – 4.5) in Algorithm 1. Then there exists an index \( l^* \) such that \((x^{l^*}, y^{l^*}, z^{l^*}) \in \mathcal{N}(\mu, \gamma)\).

Proof: We will prove our result by contradiction. Suppose that the result is false. i.e.

\[
0 < \mu \gamma < \psi_\mu(x^l, z^l) \quad \text{for all index } l.
\]

(5.2)

The following observations are in order.

\((D1)\). From construction of \( \{(x^l, y^l, z^l)\} \) we have that \( \{(x^l, z^l)\} \subset \Omega(\mu \gamma) \).

\((D2)\). The penalty parameter sequence \( \{C^l\} \) converges, say to \( C^* \). To see this, recall that we have a monotone nondecreasing penalty parameter update (see Algorithm 2). therefore \( \{C^l\} \) is either convergent or unbounded. Suppose that \( \{C^l\} \) is unbounded.

Then there exists an unbounded subsequence \( \{C^{l''}\} \) given by

\[
C^{l''} = 2 \left\{ \frac{\|\nabla(x, z)l(x^{l''}, y^{l''}, z^{l''})^T \delta v^{l''}\|}{\psi_\mu(x^{l''}, z^{l''})} + 1 \right\},
\]

where \( \delta v^{l''} = (\Delta x^{l''}, \Delta z^{l''}) \).

Now boundedness of \( \{(x^l, z^l)\} \) and Corollary 5.1 imply that

\[
\psi_\mu(x^{l''}, z^{l''}) \to 0 \quad \text{as } l'' \to \infty.
\]

This leads to a contradiction of \(D1\). Therefore \( \{C^l\} \) converges.

In place of our assumption \( B4 \), the observations \( D1, D2, \) and boundedness away from
zero of \( \{(x^l, z^l)\} \) we may assume that there exists a subsequence \( \{(x^{l''}, y^{l''}, z^{l''})\} \) such that:

(i) This subsequence converges to an interior point \((x^*, y^*, z^*)\).

(ii) The penalty parameter subsequence \( \{C^{l''}\} \) is either constant and equal to \( C^* \) for sufficiently large index \( l'' \) or strictly increasing.

Observe that

\[ o_{\mu}(w^{l''}; C^*) = o_{\mu}(w^{l''}; C^{l''}) - o_{\mu} \psi_{\mu}(v^{l''}). \]

and

\[ \nabla o_{\mu}(w^{l''}; C^*)^T \Delta v^l = \nabla o_{\mu}(w^{l''}; C^{l''})^T \Delta v^l - 2 o_{\mu} \psi_{\mu}(w^{l''}). \]

where \( o_{\mu} \to 0 \) as \( l'' \to \infty \). Since the residual sequence \( \{\psi_{\mu}(w^{l''})\} \) is bounded we can choose for sufficiently large index \( l'' \) the penalty parameter value \( C^* \) instead of \( C^{l''} \) without affecting the backtracking power \( s \) in Step 4.4. Hence we obtain the same iterate \((x^{l''+1}, y^{l''+1}, z^{l''+1})\) in Step 4.5 using either penalty parameter \( C^{l''} \) or \( C^* \).

Now, we collect all our observations on the sequence \( \{w^{l''}\} \). For this sequence we are performing a backtracking scheme on the fixed modified augmented Lagrangian merit function

\[ o_{\mu}(x, y, z) \equiv o_{\mu}(x, y, z; C^*). \]

It is worth noticing that

\[ \nabla_{(x, z)} o_{\mu}^*(x^{l''}, y^{l''}, z^{l''})^T \Delta v^{l''} \leq -2 \psi_{\mu}(x^{l''}, z^{l''}). \]  

(5.3)

Since the steplength sequence \( \{\alpha^{l''}\} \) is bounded away from zero we have from standard linesearch theory (see Ortega and Rheinboldt [34], and Byrd and Nocedal [4]) that.

\[ \frac{\nabla_{(x, z)} o_{\mu}^*(x^{l''}, y^{l''}, z^{l''})^T \Delta v^{l''}}{\Delta v^{l''}} \to 0. \]

In particular, \( \{(x^{l''}, z^{l''})\} \subset \Omega(\mu \gamma) \), hence \( \{\Delta v^{l''}\} \) is bounded. We conclude from (5.3) that

\[ -2 \psi_{\mu}(x^{l''}, z^{l''}) \to 0. \]
This is a contradiction of the original assumption (5.2).

\[ \square \]

5.3 Global Convergence Theorems

In this section we establish our convergence theory for Algorithm 1. The first result states that any limit point of the outer iteration sequence is a quasi-central point corresponding to \( \mu = 0 \). This result is not surprising since Algorithm 1 was designed around the quasi-central path. Our second results guarantees a basic and fundamental property for any method for solving (2.1). This property merely states that if the method generates a convergent sequence, the limit of that sequence is a KKT point.

For now on we consider Algorithm 1 without the stopping criteria given in Step 1. We begin by stating our first global convergence result. Recall that our perturbation parameter update in Step 2 of Algorithm 1 is given by

\[ \mu_k = \sigma_k \nu_{\mu_{k-1}}(x_k, z_k). \quad \sigma_k \in (0, 1). \quad (5.4) \]

**Theorem 5.2** Assume that B1-B4 hold. Let \( \{ (x_k, y_k, z_k) \} \) be the outer sequence generated by Algorithm 2 with the choice of \( \mu_k \) given by (5.4) such that \( \{ \sigma_k \} \) is bounded away from zero. Then \( \mu_k \to 0 \). Q-linearly. Moreover, any limit point of \( \{ (x_k, y_k, z_k) \} \) satisfies the equations \( h(x) = 0 \) and \( XYZ = 0 \).

**Proof.** Theorem (5.1) implies that the outer sequence \( \{ (x_k, y_k, z_k) \} \) is well defined. From Step 4, our perturbation parameter update (5.4), and the boundedness of \( \{ \sigma_k \} \), we have

\[ \mu_k = \sigma_k \nu_{\mu_{k-1}}(x_k, z_k) < \gamma \mu_{k-1} \]

Since \( \gamma < 1 \), \( \{ \mu_k \} \) converges to zero Q-linearly.
Let \((x^*, y^*, z^*)\) be a limit point of \(\{ (x_k, y_k, z_k) \}\). Let \(\{ (x_{k'}, y_{k'}, z_{k'}) \}\) be a subsequence that converges to \((x^*, y^*, z^*)\). Then \(\mu_{k'-1} \rightarrow 0\) and \(\mu_{k'} = \sigma_{k'} \nu_{k'-1} (x_{k'}, z_{k'}) \rightarrow 0\) as \(k' \rightarrow \infty\). In particular \(\{ \sigma_{k'} \}\) is bounded away from zero. We appeal to continuity of the function \(h\) to conclude that \(h(x^*) = 0\) and \(X^*Z^*e = 0\).

\[\square\]

Let us consider the notation \(w = (x, y, z)\). Theorem 5.1 ensures that for each index \(k\), our Algorithm 1 will construct only a finite number of inner iterations of the inner loop iteration

\[w_k^0, w_k^1, \ldots, w_k^{l^*_k},\]

where \(w_k = w_k^0\) and \(l^*_k\) corresponds to the first index \(l\) such that \(w_k^l \in \mathcal{N}(\mu_k, \gamma)\). We define the sequence generated by Algorithm 1 without the stopping criteria in Step 1. as

\[\{w_k^{l_j}\} = \{w_k^0, \ldots, w_k^0, w_k^0, \ldots, w_k^{l_j}, \ldots\}\]

**Theorem 5.3** Assume that B1-B4 holds. Let \(\{w_k^{l_j}\}\) be the sequence generated by Algorithm 2. If \(\{w_k^{l_j}\}\) converges to \(w^* = (x^*, y^*, z^*)\) and \(F'(w^*)\) is nonsingular then \(w^*\) is a KKT point.

**Proof:** Observe that the subsequence \(\{w_k^0\}\) is merely the outer iteration sequence \(\{w_k\}\), that also converges to \(w^*\). Hence by Theorem (5.2) we conclude that

\[h(x^*) = 0\quad \text{and} \quad X^*Z^*e = 0.\]

For this subsequence, we obtain the next iterate as

\[w_k^1 = w_k^0 + \alpha_k^0 \Delta w_k^0\]

where the associated step length parameter \(\alpha_k^0\) are bounded away from zero. Since \(\{w_k\}\) converges to \(w^*\) and \(F'(w^*)\) is nonsingular, we conclude that

\[\Delta w_k^0 \rightarrow 0\quad \text{as} \quad k \rightarrow \infty.\]
Writing out the first equation in (3.4) we obtain

\[ \nabla f(x_k^0) + \nabla(y_k^0 + \Delta y_k^0 - (z_k^0 + \Delta z_k^0)) = -\nabla_x^2 l(u_k^0) \Delta x_k^0. \]  

(5.5)

Now if we take the limit in both sides of (5.5) when \( k \to \infty \), we obtain

\[ \nabla_x l(u^*) = 0 \]

Therefore \( u^* \) is a KKT point.
Chapter 6

Numerical Results

In this chapter we present numerical results for the Newton path-following primal-dual Newton interior-point method of Chapter 4 (Algorithm 1).

6.1 Implementation

We coded our program in Matlab 4.2 using a Sun workstation with 64 bit arithmetic. The stopping criteria in Step 1 was

\[
\frac{\|F(w_k)\|_2}{1 + \|w_k\|_2} \leq \varepsilon_{exit} = 10^{-8}.
\]

The centering parameter in Step 2 was given by

\[
\sigma_k = 0.5.
\]

The neighborhood around the quasi-central path in our inner stopping criteria (Step 4) was chosen as \(N(\mu_k; 0.8)\). The second order derivatives were computed by finite differences. The steplength parameter \(\tilde{\alpha}^l\) given in Algorithm 3 was used to prove our convergence results of Chapter 5. In order to compute this steplength parameter we must obtain the first positive solution of the nonlinear equation given by \(g_2(\alpha) = 0\). Hence we chose in our implementation an easier computable steplength parameter. Our steplength parameter was given by \(\alpha^l = \min (1, 0.995\tilde{\alpha}^l)\), where

\[
\tilde{\alpha}^l = \min \left(\frac{-1}{\min (((X^l)^{-1}) \Delta x^l, -1)}, \frac{-1}{\min (((Z^l)^{-1}) \Delta z^l, -1)}\right) \tag{6.1}
\]

This steplength parameter is the smaller steplength to the boundary. Just notice that \((x^l, z^l) + \tilde{\alpha}^l(\Delta x^l, \Delta z^l)\) has at least one component (in \(x\) or \(z\)) equal to zero. In
the backtracking scheme (Step 4.4). we set $\beta = 10^{-4}$. and $p = 0.5$. The maximum number of linear solver that are allowed was 100.

6.2 Numerical Experience

The test problems are from Hock and Schittkowski [28], and Schittkowski [36]. We labeled them with the same number than they have in [28], and [36] Firstly, we compare the role of the centrality condition on our modified augmented Lagrangian merit function (3.1). We summarize our numerical results in tables (6.3) and (6.3). Both tables are formed by six columns as follows: The first column contains the problem number. The second column is the dimension of the primal variable $x$, referred by $n$. The third and fourth columns are the number of equality constraints ($m$) and inequality constraints ($p$) respectively. The fifth and sixth columns are the number of linear system solves (Step 4.1) for each problem depending on the path following strategy (Centrality) or not (No Centrality). The option of ‘No Centrality’, means that the inner loop (Step 4.1-4.5) is performed only once. This gives a linesearch damped and perturbed Newton method applied to the KKT conditions (2.3) using as merit function our modified augmented Lagrangian function (3.1). The starting points in the primal variable are the same as those in [28] and [36]. We solved 60 problems. In 40 of them we found the solution reported in [28] and [36]. In most of the test problems the number of linear system solves using centrality or not are similar. But the use of ’Centrality’ or ’No Centrality’ produced different iterates as it is shown in problems 81 and 104. These two problems are not solved by pure Newton’s method i.e by the ’No centrality’ option without linesearch. Then we plot for both problems, each inner loop (counting the linear systems solved) versus the $l_2$ norm of the KKT condition in the interior point given by Step 4.5 in Algorithm 1. See Figure 6.3 and Figure 6.3. We observe that the path-following strategy decreases the norm of the
KKT condition faster that the option of 'No Centrality' far away of the solution. The path-following strategy enforces the centrality condition faster. This is shown in problems 81 and 104 in Figure 6.3 and Figure 6.3 respectively. Also, the behavior of the penalty parameter should be different between the options of 'Centrality' and 'No Centrality'. In table (6.3) for each test problem the next two columns correspond to the last penalty parameter using path-following strategy or not, respectively. The option of 'No Centrality' gives in general a smaller penalty parameter than does the 'Centrality' option. This emphasizes the role of the centrality condition which may force larger penalty parameters. We solved Problem 13 in which the constrained qualifications does not hold. Problem 13 has been difficult to solve for interior-point codes (see El-Bakry et al [16], and Yamashita [44]).

6.3 Comments

We summarize our numerical results in the following comments. Smaller choices of $\sigma_k$ in Step 2 deteriorates the global behavior of Algorithm 1. Since we require too much accuracy in the centrality conditions. Also, values of $\sigma_k$ close to 1 produce short steps in the satisfaction of the centrality conditions, hurting the global behavior of Algorithm 1. For $\sigma_k \in [0.4, 0.6]$, our numerical results are much the same. Therefore we chose $\sigma_k = 0.5$. We did not consider a dynamic choice of $\sigma_k$. For the updating penalty parameter scheme (Algorithm 3), we used only Step 1. This scheme produces in practice a monotone nondecreasing update and similar numerical results. If the factor 2 in Step 1 of Algorithm 2 is replaced by a larger number the numerical results are not altered. However, replacing the factor 2 by a smaller positive number causes the convergence of Algorithm 1 to deteriorate. This emphasizes the role of $C_{tral}$ in the rate of decrease for our modified augmented Lagrangian merit function. Our theory did not guarantee boundedness or unboundedness of the penalty parameter.
(See Table 6.3). This property depends on the problem and the initial interior point. However, unboundedness may not lead to bad behavior. For instance, we solved problem 13 which has been one of the most difficult problems to solve for interior-point methods. Algorithm 1 was designed around the centrality conditions $h(x) = 0$ and $XZe = \mu e$. The numerical results clearly indicate this feature and validate our convergence theory.
<table>
<thead>
<tr>
<th>Problem</th>
<th>$n$</th>
<th>$m$</th>
<th>$p$</th>
<th>Linear Systems</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Centrality</td>
<td>No Centrality</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>20</td>
<td>33</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>13</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>29</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>26</td>
<td>39</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>24</td>
<td>2</td>
<td>0</td>
<td>5</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>25</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>29</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>31</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>32</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>34</td>
<td>3</td>
<td>0</td>
<td>8</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td>35</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td>36</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>37</td>
<td>3</td>
<td>0</td>
<td>8</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>38</td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>23</td>
<td>19</td>
</tr>
<tr>
<td>41</td>
<td>4</td>
<td>1</td>
<td>8</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>43</td>
<td>4</td>
<td>0</td>
<td>10</td>
<td>14</td>
<td>-</td>
</tr>
<tr>
<td>44</td>
<td>4</td>
<td>0</td>
<td>10</td>
<td>11</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 6.1  Hock and Schittkowski test problems.  
The symbol ' - ' means no convergence.
<table>
<thead>
<tr>
<th>Problem</th>
<th>n</th>
<th>m</th>
<th>p</th>
<th>Linear Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Centrality</td>
</tr>
<tr>
<td>45</td>
<td>5</td>
<td>0</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>53</td>
<td>5</td>
<td>3</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>60</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>62</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>64</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>65</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>66</td>
<td>3</td>
<td>0</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>71</td>
<td>4</td>
<td>1</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>72</td>
<td>4</td>
<td>0</td>
<td>10</td>
<td>23</td>
</tr>
<tr>
<td>73</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>74</td>
<td>4</td>
<td>3</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>75</td>
<td>4</td>
<td>3</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td>76</td>
<td>4</td>
<td>0</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>80</td>
<td>5</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>81</td>
<td>9</td>
<td>13</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>83</td>
<td>5</td>
<td>0</td>
<td>16</td>
<td>38</td>
</tr>
<tr>
<td>86</td>
<td>5</td>
<td>0</td>
<td>15</td>
<td>44</td>
</tr>
<tr>
<td>93</td>
<td>6</td>
<td>0</td>
<td>8</td>
<td>27</td>
</tr>
<tr>
<td>104</td>
<td>8</td>
<td>0</td>
<td>22</td>
<td>15</td>
</tr>
<tr>
<td>227</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>233</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>250</td>
<td>3</td>
<td>0</td>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>251</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>262</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>325</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>339</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>341</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>342</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>353</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>354</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 6.2  Hock and Schittkowski test problem (Continued). The symbol '-' means no convergence.
Figure 6.1  The norm of the KKT conditions for the two strategies on Problem 81

Figure 6.2  The norm of the KKT conditions for the two strategies on Problem 104
Figure 6.3  The norm of the constraints for the two strategies on Problem 81

Figure 6.4  The norm of the constraints for the two strategies on Problem 104
<table>
<thead>
<tr>
<th>Problem</th>
<th>Penalty Parameter</th>
<th>Problem</th>
<th>Penalty Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Centrality</td>
<td>No centrality</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>53</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>50</td>
<td>62</td>
</tr>
<tr>
<td>5</td>
<td>10^3</td>
<td>70</td>
<td>64</td>
</tr>
<tr>
<td>10</td>
<td>2.1</td>
<td>-</td>
<td>65</td>
</tr>
<tr>
<td>11</td>
<td>3.2</td>
<td>14.5</td>
<td>66</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>2</td>
<td>71</td>
</tr>
<tr>
<td>13</td>
<td>10^8</td>
<td>10^8</td>
<td>72</td>
</tr>
<tr>
<td>14</td>
<td>15.5</td>
<td>11</td>
<td>73</td>
</tr>
<tr>
<td>16</td>
<td>10^4</td>
<td>-</td>
<td>74</td>
</tr>
<tr>
<td>18</td>
<td>20</td>
<td>19</td>
<td>75</td>
</tr>
<tr>
<td>20</td>
<td>700</td>
<td>10^4</td>
<td>76</td>
</tr>
<tr>
<td>21</td>
<td>90</td>
<td>250</td>
<td>80</td>
</tr>
<tr>
<td>22</td>
<td>6.5</td>
<td>14</td>
<td>81</td>
</tr>
<tr>
<td>23</td>
<td>60</td>
<td>17</td>
<td>83</td>
</tr>
<tr>
<td>24</td>
<td>10^4</td>
<td>150</td>
<td>86</td>
</tr>
<tr>
<td>25</td>
<td>250</td>
<td>40</td>
<td>93</td>
</tr>
<tr>
<td>29</td>
<td>2</td>
<td>180</td>
<td>104</td>
</tr>
<tr>
<td>30</td>
<td>400</td>
<td>50</td>
<td>227</td>
</tr>
<tr>
<td>31</td>
<td>800</td>
<td>45</td>
<td>233</td>
</tr>
<tr>
<td>32</td>
<td>10^3</td>
<td>10^3</td>
<td>250</td>
</tr>
<tr>
<td>34</td>
<td>30</td>
<td>11</td>
<td>251</td>
</tr>
<tr>
<td>35</td>
<td>10^3</td>
<td>10^3</td>
<td>262</td>
</tr>
<tr>
<td>36</td>
<td>10^4</td>
<td>10^4</td>
<td>325</td>
</tr>
<tr>
<td>37</td>
<td>10^3</td>
<td>10^4</td>
<td>339</td>
</tr>
<tr>
<td>38</td>
<td>10^6</td>
<td>10^7</td>
<td>341</td>
</tr>
<tr>
<td>41</td>
<td>10^3</td>
<td>500</td>
<td>342</td>
</tr>
<tr>
<td>43</td>
<td>250</td>
<td>-</td>
<td>353</td>
</tr>
<tr>
<td>44</td>
<td>10^3</td>
<td>200</td>
<td>354</td>
</tr>
</tbody>
</table>

Table 6.3  The role of the centrality condition on the penalty parameter.
Chapter 7

Quasi-Newton Methods and a $Q$-Superlinear Result

In this chapter we establish a $Q$-superlinear characterization for Quasi-Newton methods for solving systems of nonlinear equations.

7.1 The Damped and Perturbed Newton Method

Given an initial $x_0$, by a damped and perturbed Newton method for solving the nonlinear equation (2.1). we mean the iterative process

$$
x_{k+1} = x_k - \alpha A_k^{-1}(F(x_k) - r_k), \quad k = 1, 2, \ldots
$$

(7.1)

In (7.1), $0 < \alpha \leq 1$, is the steplength parameter. $r_k \in \mathbb{R}^n$ is the perturbation vector. and $A_k$ is a matrix approximation to $F'_\mu(x_k)$. We do not intend to study in detail the iterative process (7.1). therefore we will not be overly concerned with corresponding parametric choices. The damped and perturbed Quasi-Newton methods will be used as a tool to gain understanding of our primal-dual Quasi-Newton interior-point method in Chapter 8. In particular we are interested in a $Q$-superlinear characterization of (7.1) in terms of its parametric choices applied to our interior point methods. However, we were not able to find such characterization in the optimization literature. For now, we concentrate our efforts on filling this theoretical gap.
7.2 Characterization for Damped and Perturbed Quasi-Newton Methods

We begin by collecting some known useful facts. Toward this end let \( \epsilon_k = x_k - x^* \)
and \( s_k = x_{k+1} - x_k \); assume \( S1 - S3 \), and that \( \{x_k\} \) converges to \( x^* \).

There exists a constant \( \rho > 0 \) such that for \( k \) sufficiently large

\[
\frac{1}{\rho} \| \epsilon_k \| \leq \| F(x_k) \| \leq \rho \| \epsilon_k \|. \tag{7.2}
\]

A proof of (7.2) can be found, for example, in Dembo, Eisenstat, and Steihaug [10]. It follows that

\[
\frac{\| \epsilon_{k+1} \|}{\| \epsilon_k \|} \rightarrow 0 \quad \Rightarrow \quad \frac{\| s_k \|}{\| \epsilon_k \|} \rightarrow 1. \tag{7.3}
\]

and

\[
\frac{\| \epsilon_{k+1} \|}{\| \epsilon_k \|} \rightarrow 0 \Leftrightarrow \frac{\| F_{k+1} \|}{\| s_k \|} \rightarrow 0. \tag{7.4}
\]

To establish (7.3) we merely need to observe that \( \epsilon_{k+1} = s_k + \epsilon_k \). Moreover, (7.4) follows directly once we write

\[
\frac{\| F_{k+1} \|}{\| \epsilon_k \|} = \frac{\| F_{k+1} \|}{\| s_k \|} \frac{\| s_k \|}{\| \epsilon_k \|}. \]

The next two theorems will motivate choices for the steplength \( \alpha_k \) and the perturbation vector \( r_k \).

Theorem 7.1 Let \( \{x_k\} \) be generated by (7.1). Assume that S1, S2, and S3 hold and that \( x_k \rightarrow x^* \). Then any two of the following statements imply the third:

(i) \( x_k \rightarrow x^* \) Q-superlinearly.

(ii) \( \lim_{k \to \infty} \frac{\| s_k (1 - \alpha_k) F(x_k) \|}{\| s_k \|} = 0. \)

(iii) \( \lim_{k \to \infty} \frac{\| (A_k - F(x^*) s_k) \|}{\| s_k \|} = 0. \)
Proof: Adding and subtracting the appropriate quantities, we have

\[ F(x_{k+1}) = [F(x_{k+1}) - F(x_k) - F'(x^*)s_k] - [A_k - F'(x^*)]s_k + [\alpha_k r_k + (1 - \alpha_k)F(x_k)]. \quad (7.5) \]

From (7.4), (i) is equivalent to

\[ \lim_{k \to \infty} \frac{\|F(x_{k+1})\|}{\|s_k\|} = 0. \]

Using Lemma 4.1.15 in [14] we have

\[ \|F(x_{k+1}) - F(x_k) - F'(x^*)s_k\| \leq \frac{\gamma\|s_k\|}{2}(\|\epsilon_{k+1}\| + \|\epsilon_k\|). \quad (7.6) \]

The remainder of the proof is fairly straightforward.

\[ \square \]

Observe that if for all \( k \), \( \alpha_k = 1 \) and \( r_k = 0 \), then (7.1) becomes the standard quasi-Newton method; moreover, in this case condition (ii) is trivially satisfied and Theorem 2.1 reduces to the standard Dennis-Moré characterization.

Condition (ii) tells us that essentially for Q-superlinear convergence we must have \( \alpha_k \to 1 \) and \( r_k = o(\|s_k\|) \). We are somewhat concerned with this latter requirement for the following reason. Our expectation is to be able to control the size of the perturbation vector \( r_k \); however, at the beginning of the iteration when we must choose \( r_k \), the step \( s_k \) is unknown to us. For this reason we look for a similar condition involving \( \|F_k\| \), a quantity which is readily available. However, we must add an assumption concerning the rate of convergence of \( \{x_k\} \).

**Theorem 7.2** Let \( \{x_k\} \) be generated by (7.1). Assume that S1, S2, and S3 hold and that \( x_k \to x^* \). Then any two of the following statements imply the third.
(i)' \( x_k \to x^* \) Q-superlinearly.

(ii)' \( \lim_{k \to -\infty} \frac{\|A_k x_k + (1 - \alpha_k) F(x_k)\|}{\|F(x_k)\|} = 0 \) and the convergence of \( \{x_k\} \) to \( x^* \) is Q-linear.

(iii)' \( \lim_{k \to -\infty} \frac{\|A_k - F'(x^*) x_k\|}{\|s_k\|} = 0. \)

**Proof:** We must show that any two conditions in Theorem 2.1 are equivalent to the corresponding two conditions in Theorem 2.2. Observe that from (7.2), the fact that \( s_k = \epsilon_{k+1} - \epsilon_k \), and the Q-linear convergence of \( \{x_k\} \) to \( x^* \), there exist positive constants \( \beta_1 \) and \( \beta_2 \) such that for \( k \) sufficiently large

\[ \frac{\beta_1}{\rho \|F(x_k)\|} \leq \frac{1}{\|s_k\|} \leq \frac{\rho \beta_2}{\|F(x_k)\|}. \]  

(7.7)

The proof of the theorem now follows from Theorem 2.1. and (7.7).

\( \square \)

The assumption in (ii)' concerning the rate of convergence of \( \{x_k\} \) can be replaced by the following weaker statement:

The set

\[ Q_1^*(\{x_k\}) = \{ \text{limit points of} \left\{ \frac{\|\epsilon_{k+1}\|}{\|\epsilon_k\|} \right\} \} . \]

does not contain one and \( \infty \), for at least one norm.

Clearly the set \( Q_1^*(\{x_k\}) \) depends on the norm selected. The largest element of \( Q_1^*(\{x_k\}) \) is the well-known \( Q_1 \)-factor. For more detail on this issue, see Chapter 9 of Ortega and Rheinboldt [34].

In terms of secant methods the assumption that \( \{x_k\} \) converges to \( x^* \) Q-linearly seems not to be restrictive. In fact if the matrices \( \{A_k\} \) satisfy a standard bounded deterioration property, as do the well-known secant methods, then in an appropriate norm \( x_k \to x^* \) Q-linear. (see Chapter 8 of Dennis and Schnabel [14] for more detail
Theorem 2.2 tells us that in order to obtain Q-superlinear convergence we should have \( r_k = o(\|F_k\|) \) and \( \alpha_k \to 1 \). We find it interesting that this is exactly the condition given by Dembo, Eisenstat, and Steihaug [10] for Q-superlinear convergence of their inexact Newton method. Actually, they chose \( \alpha_k = 1 \) for all \( k \). An obvious choice for the perturbation vector is \( r_k = \sigma_k \|F_k\| \) where \( \sigma_k \in (0, 1] \) and \( \sigma_k \to 0 \) as \( k \to \infty \).
Chapter 8

Primal-Dual Quasi-Newton Interior-Point Methods

In this chapter we describe the primal-dual Quasi-Newton interior-point method. The main characteristic of these methods is to substitute for the Jacobian of the perturbed KKT conditions a matrix approximation. In fact due to the structure of the Jacobian we only consider matrix approximations to the Hessian of the Lagrangian. Appealing to our $Q$-superlinear characterization of Chapter 7, we will impose a condition on the matrix approximation in order to obtain $Q$-superlinear convergence.

8.1 The Method

We now describe a primal-dual Quasi-Newton interior-point method for solving (2.1). For the sake of clarity, at iteration $x_k$ we denote $F(x_k)$ by $F_k$, and $\nabla h(x_k)$ by $\nabla h_k$; similar notation will be used in other quantities.

Algorithm 4 Let $w_0 = (x_0, y_0, z_0)$ be an initial interior point.

For $k = 0, 1, \ldots$, until convergence do

Step1. Choose $\sigma_k \in (0, 1]$ and set $\mu_k = \sigma_k R_k$ for some $R_k \in \mathbb{R}$.

Step2. Obtain $\Delta w_k = (\Delta x_k, \Delta y_k, \Delta z_k)^T$ as the solution of the linear system

$$M_k \Delta w_k = -F_{\mu_k}(w_k)$$  \hspace{1cm} (8.1)

where

$$M_k = \begin{pmatrix} G_k & \nabla h_k & -I_n \\ \nabla h_k^T & 0 & 0 \\ Z_k & 0 & X_k \end{pmatrix}.$$
Step 3. Choose $\tau_k \in (0, 1)$ and set

\[ \alpha_k = \min(1, \tau_k \tilde{\alpha}_k) \]

\[ \alpha_{k,y} = 1 \text{ or } \alpha_{k,y} = \alpha_k \]

where

\[ \tilde{\alpha}_k = \min \left\{ \frac{-1}{\min(X_k^{-1} \Delta x_k, -1)}, \frac{-1}{\min(Z_k^{-1} \Delta z_k, -1)} \right\} \]

Step 4. Update

\[ w_{k+1} = w_k + \lambda_k \Delta w_k \]

where \( \lambda_k = \text{diag}(\alpha_k, \ldots, \alpha_k, \alpha_{k,y}, \ldots, \alpha_{k,y}, \alpha_k, \ldots, \alpha_k) \)

in the above the three groups of scalars have \( n \), \( m \), and \( n \) members respectively.

The choice for \( R_k \) will be in general \( \|F(w_k)\| \): however we leave it open to obtain a certain amount of needed flexibility in the statement of our theorems in Section 3.

The choice \( G_k = \nabla^2 x l(w_k) \) corresponds to Newton's method. For this choice El-Bakry, Tapia, Tsuchiya, and Zhang [16] established local convergence, superlinear convergence, and quadratic convergence for Algorithm 1 for the appropriate choices of \( \tau_k \) and \( R_k \). Yamashita [44] considered a somewhat different steplength than that described in Step 3, this choice was based on a particular merit function. He then established a global convergence result for his line-search algorithm. El-Bakry et al [16] also gave a global convergence result for a line-search globalization of their form of Algorithm 1. Observe that the choice of steplength in Step 3, \( \alpha_k = \tau_k \tilde{\alpha}_k \) and \( \tau_k \in (0, 1) \) keep \( x_{k+1} \) and \( z_{k+1} \) positive. If \( \tau_k \) was chosen to be equal to one, then at least one component of \( x_{k+1} \) or \( z_{k+1} \) would be zero. We could use different steplength also for the \( x \) and \( z \) variables. The obvious choice would be to let \( \alpha_{k,x} = \)
min(1, \tau_k \tilde{\alpha}_{kz}) \text{ and } \alpha_{kz} = \min(1, \tau_k \tilde{\alpha}_{kz}). \text{ where }
\tilde{\alpha}_{kx} = \frac{-1}{\min(X_k^{-1} \Delta x_k, -1)}.

and
\tilde{\alpha}_{kz} = \frac{-1}{\min(Z_k^{-1} \Delta z_k, -1)}.

Since the asymptotic properties of these choices are essentially the same, we will not concern ourselves with other choices of steplength parameters. It should be clear that the algorithmic choices are the choices of \( \tau_k \), \( \sigma_k \), and \( G_k \) the approximation to \( \nabla^2 \phi(w_k) \). Our objective is to characterize Q-superlinear convergence in terms of the algorithmic choices. A straightforward application of Theorem 2.2 would lead to a characterization in terms of all the variables \((x, y, z)\). Such activity would be incomplete since for equality constrained optimization, where the \( z \)-variable is not present, the Boggs-Tolle-Wang characterization is in terms of the \( x \)-variable alone. Effectively, the \( y \)-variable can be removed from the problem as demonstrated by Stoer and Tapia [38]. Our first initial efforts in the current research attempted to obtain such a characterization for Algorithm 1; however we could not do so without making assumptions which we considered undesirable. Therefore, we turned to attempting a characterization in terms of the \((x, z)\)-variables and were successful. It follows then that in this application the primary variables are \( x \) and \( z \), each carries independent information and can not be removed from the problem. In retrospective we find this occurrence fitting and not surprising.

8.2 An Equivalent Formulation.

In this section we imitate the approach taken by Stoer and Tapia [38] in deriving the Boggs-Tolle-Wang characterization for equality constrained optimization. Our task is to construct a quasi-Newton method that involves only the \((x, z)\)-variables.
equivalent to Algorithm 1 of Section 3, and has the form of a damped and perturbed quasi-Newton method as described by (7.1). This equivalence will allow us, in Section 5, to apply our characterization Theorem 2.2.

Assumption A3 allows us to locally, i.e., in a neighborhood of $x^*$, consider the projection operator

$$P(x) = I - \nabla h(x)[\nabla h(x)^T \nabla h(x)]^{-1} \nabla h(x)^T.$$  \hfill (8.1)

In turn this allows us to consider the nonlinear equation

$$F_0(x, z) = \left( P(x)(\nabla f(x) - z) + \nabla h(x)h(x) \right) + XZ\epsilon = 0.$$  \hfill (8.2)

Observe that $F_0 : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{2n}$. We now demonstrate that Algorithm 1 is equivalent to a damped and perturbed quasi-Newton method applied to equation (8.2). Toward this end let $(x_k, y_k, z_k), G_k$, and $\mu_k$ be as in the $k$-th iteration of Algorithm 1 and consider the linear system

$$\begin{pmatrix} P_k G_k + \nabla h_k \nabla h_k^T & -P_k \\ Z_k & X_k \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta z_k \end{pmatrix} = -(F_0(x_k, z_k) - \mu_k \epsilon).$$  \hfill (8.3)

In (8.3), $\epsilon$ is the $2n$-vector whose first $n$ components are zero and whose last $n$ components are one. We will also need to consider the formula

$$y_k^+ = - (\nabla h_k^T \nabla h_k)^{-1} \nabla h_k^T (G_k \Delta x_k + \nabla f_k - (z_k + \Delta z_k)),$$  \hfill (8.4)

where $(\Delta x_k, \Delta z_k)$ is the solution of (8.3).

**Proposition 8.1** Let $(x^*, y^*, z^*)$ be a solution of the KKT conditions (2.3) at which the standard assumptions A1-A5 hold. Then $(x^*, z^*)$ is a solution of the nonlinear equation (8.2) and the standard Newton's method assumptions S1-S3 hold for $F_0$ at this solution. Moreover, if $(\Delta x_k, \Delta y_k, \Delta z_k)$
is a solution of the linear system (8.1). Then \((\Delta x_k, \Delta z_k)\) is a solution of the linear system (8.3). Conversely, if \((\Delta x_k, \Delta z_k)\) is a solution of the linear system (8.3) and we let \(\Delta y_k = y_k^+ - y_k\), where \(y_k^+\) is given by (8.4), then \((\Delta x_k, \Delta y_k, \Delta z_k)\) is a solution of the linear system (8.1).

**Proof.** We begin by establishing the equivalence between the linear systems (8.1) and (8.3).

Writing out (8.1) in detail gives

\[
G_k \Delta x_k + \nabla h_k \Delta y_k - \Delta z_k = - (\nabla f_k + \nabla h_k y_k - z_k)
\]

\[
\nabla h_k^T \Delta x_k = - h_k
\]  

\[
Z_k \Delta x_k + X_k \Delta z_k = - X_k Z_k + \mu_k \varepsilon .
\]  

Writing out (8.3) in detail gives

\[
(P_k G_k + \nabla h_k \nabla h_k^T) \Delta x_k - P_k \Delta z_k = -(P_k(\nabla f_k - z_k) + \nabla h_k h_k)
\]

\[
Z_k \Delta x_k + X_k \Delta z_k = - X_k Z_k + \mu_k \varepsilon .
\]  

We observe that we can write

\[
P_k [G_k \Delta x_k + \nabla f_k - (z_k + \Delta z_k)] = G_k \Delta x_k + \nabla f_k - (z_k + \Delta z_k) + \nabla h_k y_k^+
\]  

where \(y_k^+\) is given by (8.4).

Now, suppose \((\Delta x_k, \Delta y_k, \Delta z_k)\) solves (8.5). Multiplying the first equation by \(P_k\), the second equation by \(\nabla h_k\), adding the two resulting equations, and recalling that \(P_k \nabla h_k = 0\) leads us to the first equation in (8.6). Hence \((\Delta x_k, \Delta z_k)\) solves (8.6). Conversely, suppose \((\Delta x_k, \Delta z_k)\) solves (8.6). Multiplying the first equation by \(\nabla h_k^T\) gives the second equation in (8.5). This in turn tells us that the first equation in (8.6) now implies that the left-hand side of (8.7) is zero. Hence the right-hand side is zero and the first equation in (8.5) holds with \(y_k + \Delta y_k = y_k^+\). This establishes the equivalence of the two linear systems (8.5) and (8.6).
If \((x^*, y^*, z^*)\) solves (2.3), then clearly \((x^*, z^*)\) solves (8.2). Observing that \(P(x)(\nabla f(x) - z) = P(x)(\nabla f(x) + \nabla h(x)y^+(x^*, z^*) - z)\) and \(y^+(x^*, z^*) = y^*\) we see that
\[
F_0'(x^*, z^*) = \begin{pmatrix}
P_x \nabla_x^2 l(x^*, y^*, z^*) + \nabla h_x \nabla h_x^T & -P_x \\
Z_x & X_x
\end{pmatrix}.
\] (8.8)

An argument along the lines of the one given above can be used to show that the linear system
\[
F_0'(x^*, z^*) \begin{pmatrix}
\eta_x \\
\eta_z
\end{pmatrix} = 0
\] (8.9)
is equivalent to the linear system
\[
F'(x^*, y^*, z^*) \begin{pmatrix}
\eta_x \\
\eta_y \\
\eta_z
\end{pmatrix} = 0
\] (8.10)

where \(F\) is given by (2.3). Under the standard assumptions \textbf{A1-A5}. for \(F\) given by (2.3), we know that \(F'(x^*, y^*, z^*)\) is nonsingular. Hence \(F_0'(x^*, z^*)\) must also be nonsingular. It should be clear that \(F_0\) and \(F\) have the same smoothness properties. This says that assumptions \textbf{S1-S3}. appropriately stated, hold for \(F_0\) at \((x^*, z^*)\). We have now established our equivalence proposition.

\(\square\)

We have shown that obtaining \((x_k, z_k)\) from Algorithm 1 can be viewed as obtaining \((x_k, z_k)\) from a damped and perturbed quasi-Newton method applied to the nonlinear equation \(F_0(x, z) = 0\) given by (8.2). Moreover, the approximate Jacobian has the form
\[
\begin{pmatrix}
P_k G_k + \nabla h_k \nabla h_k^T & -P_k \\
Z_k & X_k
\end{pmatrix}
\] (8.11)

and the Jacobian at the solution is given by (8.8).

We are now ready to state our Q-superlinear convergence results.
8.3 Q-superlinear Convergence Characterization.

In this section we apply the theory developed in Chapter 2 to the primal-dual quasi-
Newton interior-point method described by Algorithm 1 of Section 1. Recall that $G_k$
is our approximation to $G_\ast = \nabla^2 f(x^\ast) + \nabla^2 h(x^\ast)y^\ast$. Also $R_k$ appears in Step 1 of
Algorithm 1.

Theorem 8.1. Let $\{(x_k, y_k, z_k)\}$ be generated by Algorithm 1. Assume that $\{(x_k, y_k, z_k)\}$ converges to $(x^\ast, y^\ast, z^\ast)$ and assumptions A1-A5 hold at $(x^\ast, y^\ast, z^\ast)$. Furthermore, assume that $\tau_k$ and $\sigma_k$ have been chosen so that

(i) $\tau_k \to 1$.

(ii) $\sigma_k \to 0$.

Assume that either $R_k = O(\|s_k\|)$, where $s_k = (x_{k+1}, y_{k+1}, z_{k+1}) - (x_k, y_k, z_k)$,
or $R_k = O(\|F(x_k, y_k, z_k)\|)$ and $\{(x_k, y_k, z_k)\}$ converges to $(x^\ast, y^\ast, z^\ast)$ Q-
linearly.

Then $\{(x_k, y_k, z_k)\}$ converges Q-superlinearly to $(x^\ast, y^\ast, z^\ast)$ if and only if

$$\frac{\|(G_k - G_\ast)(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\| + \|y_{k+1} - y_k\| + \|z_{k+1} - z_k\|} \to 0. \quad (8.12)$$

Assume that either $R_k = O(\|s_k\|)$ where $s_k = (x_{k+1}, z_{k+1}) - (x_k, z_k)$ or
$R_k = O(\|F_0(x_k, z_k)\|)$, where $F_0$ is given by (8.2). and $\{(x_k, z_k)\}$ converges
to $(x^\ast, z^\ast)$ Q-linearly. Then $\{(x_k, z_k)\}$ converges Q-superlinearly to $(x^\ast, z^\ast)$
if and only if

$$\frac{\|P_k(G_k - G_\ast)(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\| + \|z_{k+1} - z_k\|} \to 0. \quad (8.13)$$

Proof. The proof of the theorem follows by applying Theorem 2.1. Theorem 2.2. and
Proposition 4.1. and using (8.1), (8.8). and (8.11). We have used the following fact
concerning norms in finite dimensional spaces. Let $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$. Also let $\| \cdot \|_n$ be a norm on $\mathbb{R}^n$, $\| \cdot \|_m$ a norm on $\mathbb{R}^m$, and $\| \cdot \|_{n+m}$ a norm on $\mathbb{R}^{n+m}$. Then there exist positive constants $\theta_1$ and $\theta_2$ such that

$$\theta_1(\|u\|_n + \|v\|_m) \leq \|(u, v)\|_{n+m} \leq \theta_2(\|u\|_n + \|v\|_m).$$  \hspace{1cm} (8.14)

A proof of (8.14) can be obtained by working with the $l_1$ norm and the equivalence of norms property. We also used the fact that $\tau_k \to 1$ implies $\alpha_k \to 1$ (see Step 3 of Algorithm 1) under our assumptions. This fact can be found in Yamashita and Yabe [45]. Finally, we have removed all quantities that converged to zero and were redundant in the characterization result.

□

Yamashita and Yabe [45] gave a characterization which has the flavor of (8.12). However, their assumptions were somewhat more restrictive.
Chapter 9

Concluding Remarks

We have presented two primal-dual interior-point methods approaches for solving general NLP problems. The first approach is a global path-following primal-dual Newton interior-point method. For this method we used a novel modified augmented Lagrangian merit function together with a relaxed centrality condition of the perturbed KKT conditions. We have demonstrated the numerical behavior of our primal-dual Newton interior-point method on a subset of standard test problem for NLP. In the future we would like to apply our method to larger NLP problems. In order to accomplish this task we will require iterative linear solvers for the Newton linear system and we also incorporate subroutines that compute first order derivatives. The second point was purely theoretical. Basically, we studied the case where the Hessian of the Lagrangian in the primal variable is replaced by a matrix approximation, giving the so-called primal-dual Quasi-Newton interior-point methods. We gave a characterization of $Q$-superlinear convergence in terms of the parametric choices in the methods that only contains the nonnegative variables $x$ and $z$. In the near future we would like to establish an effective Quasi-Newton method using some of the well known matrix symmetric approximation (PSB or BFGS) in constrained optimization.
Bibliography


