INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript 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. These are also available as one exposure on a standard 35mm slide or as a 17" x 23" black and white photographic print for an additional charge.

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
University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0500
Convergence rates for the variable, the multiplier, and the pair in SQP methods

Chiang, Jershan, Ph.D.

Rice University, 1989
RICE UNIVERSITY

CONVERGENCE RATES FOR THE VARIABLE, THE MULTIPLIER, AND THE PAIR IN SQP METHODS

by

JERSHAN CHIANG

A Thesis Submitted
In Partial Fulfillment Of The
Requirements For The Degree

DOCTOR OF PHILOSOPHY

Approved, Thesis Committee:

Richard A. Tapia,
Professor of Mathematical Sciences
Chairman

John E. Dennis, Jr.,
Professor of Mathematical Sciences

Shean-Tsong Chiu,
Assistant Professor of Statistics

Houston, Texas

May, 1989
CONVERGENCE RATES FOR THE VARIABLE, 
THE MULTIPLIER, AND THE PAIR IN SQP METHODS

JERSHAN CHIANG

ABSTRACT

In this work we consider relationships among the convergence rates for the variable $x$, for the multiplier $\lambda$ and for the pair $(x, \lambda)$ in SQP methods for equality constrained optimization. We show that if the convergence in $(x, \lambda)$ is $q$-superlinear, then the convergence in $x$ is at least two-step $q$-superlinear. Moreover, if the convergence in $(x, \lambda)$ and also in $x$ is $q$-superlinear, then the convergence in $\lambda$ is either $q$-superlinear or $q$-sublinear with unbounded $q_1$ factor. Finally we present a condition that guarantees $q$-superlinear convergence in $x$, $\lambda$ and $(x, \lambda)$. 
ACKNOWLEDGMENTS

I am deeply grateful to my thesis advisor, Professor Richard A. Tapia. It is a pleasure to thank him for his encouragement and inspiration throughout my graduate study at Rice. I would also like to thank professor John Dennis for helping me, through many discussions, during the research. Both have taught me a lot in numerical optimization and have shared with me their broad knowledge of mathematics. I would also like to thank the other member of my thesis committee, Professor Shean-Tsong Chiu, for his assistance.

I would like to thank Professor Richard Byrd for many helpful suggestions during his visit at Rice.

Special thanks are due to Vivian Choi for typing this thesis, and to Judy Forney for her help.

Finally, I would like to thank my family and friends for enduring encouragement and support.
# TABLE OF CONTENTS

Chapter 1  Introduction ...................................................... 1

Chapter 2  A Review of the Literature ..................................... 6
  2.1  Convergence Rate for the Pair \((x, \lambda)\) ....................... 6
  2.2  Convergence Rate for \(x\) ........................................ 8
  2.3  Convergence Rate for \(\lambda\) ..................................... 10

Chapter 3  The New Results ................................................ 12
  3.1  Preliminaries ....................................................... 12
  3.2  Convergence Rate for \(x\) ........................................ 18

Chapter 4  Summary and Concluding Remarks ............................ 37

Appendix ............................................................................... 39

Bibliography ......................................................................... 43
CHAPTER 1

Introduction

In this thesis we will be concerned with the equality constrained optimization problem

\[ \text{minimize } f(x) \]
\[ \text{subject to } h_i(x) = 0, \quad i = 1, 2, \ldots, m \]

where \( f, h_i \) are nonlinear functions defined from \( \mathbb{R}^n \) into \( \mathbb{R} \).

We denote by \( h(x) \) the vector whose components are \( h_i(x), \ i = 1, 2, \ldots, m \). The Lagrangian function associated with problem (1.1) is the function

\[ \ell(x, \lambda) = f(x) + \lambda^T h(x) \]  

(1.2)

where \( \lambda = (\lambda_1, \ldots, \lambda_m)^T \) is called the vector of Lagrange multipliers or simply the Lagrange multiplier. The augmented Lagrange function associated with problem (1.1) is the function

\[ L(x, \lambda; \rho) = f(x) + \lambda^T h(x) + \frac{1}{2} \rho h(x)^T h(x) \]  

(1.3)
The algorithm we are interested in is the SQP quasi-Newton method:

**ALGORITHM 1.1 (SQP quasi-Newton Method)**

Given $x_0, \lambda_0, B_0$

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

$$x_{k+1} = x_k + s_k \quad (1.4)$$

$$\lambda_{k+1} = \lambda_k + \Delta \lambda_k \quad (1.5)$$

$$B_{k+1} = \Pi B(x_k, s_k, \lambda_{k+1}, B_k) \quad (1.6)$$

where $s_k$ and $\Delta \lambda_k$ are the solution and the multiplier associated with the solution of the quadratic program

$$\begin{align*}
\text{minimize} & \quad \nabla_x L(x_k, \lambda_k; \rho)^T s + \frac{1}{2} s^T B_k s \\
\text{subject to} & \quad \nabla h(x_k)^T s + h(x_k) = 0.
\end{align*} \quad (1.7)$$

The matrix $B_k$ is interpreted as an approximation to $\nabla^2_x L(x_k, \lambda_k; \rho)$.

In order to facilitate the convergence analysis we will also consider the Diagonalized Multiplier Method (DMM) as defined by Tapia (1977).

**ALGORITHM 1.2 (The Diagonalized Multiplier quasi-Newton Method)**

Given $x_0, \lambda_0, B_0$

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

$$\lambda_{k+1} = U(x_k, \lambda_k, B_k) \quad (1.8)$$
\[ B_k s_k = -\nabla_x L(x_k, \lambda_{k+1}; \rho) \]  
(1.9)

\[ x_{k+1} = x_k + s_k \]  
(1.10)

\[ B_{k+1} = \mathcal{B}(x_k, s_k, \lambda_{k+1}, B_k) \]  
(1.11)

As before, \( B_k \) is interpreted as an approximation to \( \nabla^2_x L(x_k, \lambda_k; \rho) \).

We call \( U \) in (1.8) a multiplier update. The multiplier update

\[ U(x, \lambda, B) = \lambda + (\nabla h(x)^T B^{-1} \nabla h(x))^{-1}(h(x) - \nabla h(x)^T B^{-1} \nabla_x L(x, \lambda; \rho)) \]  
(1.12)

is called the Newton multiplier update. Under the mild assumptions that the quadratic program (1.7) has a unique solution and \( B_k \) is nonsingular, Tapia (1978) established the following result.

**Theorem 1.1** Algorithm 1.1 and Algorithm 1.2 using the Newton multiplier update generate identical iterates.

Actually the requirement of nonsingularity of \( B_k \) is a technicality which can be removed; however we will not concern ourselves with this issue in the present work.

The success of the convergence analysis in Fontecilla, Steihaug and Tapia (1987) has motivated us to use the framework of the DMM to study the relationship among the convergence rates for the pair \((x, \lambda)\), the variable \(x\) and the multiplier \(\lambda\).

We begin with some definitions concerning convergence rates. For the most part we follow Chapter 9 of Ortega and Rheinboldt (1970). However, our definition of
$r$-convergence is essentially that of Dennis and Schnabel (1983) which is known to be equivalent to the notion considered by Ortega and Rheinboldt (1970).

Let $\{x_k\} \subset \mathbb{R}^n$ be a convergent sequence with limit $x_*$ and assume that $x_k \neq x_*$ for all $k$. Consider a vector norm $\| \cdot \|$ on $\mathbb{R}^n$. For $p \in [1, \infty)$ the quantities

$$q_p = \lim_{k \to \infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|^p}$$

are called the $q_p$ factors of the sequence $\{x_k\}$ with respect to the norm $\| \cdot \|$.

We define the $q$-order of convergence of $\{x_k\}$ to be $\inf \{p : q_p = \infty\}$. The $q_1$ factor will be of particular interest to us. If $q_1 < 1$ then the convergence is said to be $q$-linear, while if $q_1 \geq 1$, then the convergence is said to be $q$-sublinear. Clearly the ideal situation is when $q_1 = 0$ and in this case the convergence is said to be $q$-superlinear. The least ideal case is when $q_1 = +\infty$. We will refer to this convergence as $q$-sublinear with unbounded $q_1$ factor.

Suppose that we have $\{b_k\}$ converging to zero and such that $\|x_k - x_*\| \leq b_k$ for all $k$. If the sequence $\{b_k\}$ possesses a particular $q$-convergence property, then the sequence $\{x_k\}$ is said to possess the corresponding $r$-convergence property.

If for each $k$ the subsequence $x_k, x_{k+j}, x_{k+2j}, \ldots$ displays a particular convergence behavior, then we say that the original sequence $x_1, x_2, \ldots$ has this $j$-step convergence behavior.

It is of interest to observe that $r$-convergence properties are norm independent; so are the notions of $q$-order, $q$-superlinear and $q$-sublinear with unbounded $q_1$ factor.
However, the notions of $q$-linear and $q$-sublinear are norm dependent.

Convergence of order 2 is said to be quadratic and that of order 3 is said to be cubic. Unfortunately this standard terminology is such that a $q$-order of 1 does not imply $q$-linear convergence.

In Chapter 2 we review results that can be found in the literature. In Chapter 3 we first consider preliminaries including our notation and assumptions. We then give the new results concerning the convergence rates of the pair $(x, \lambda)$, $x$ and $\lambda$. In Chapter 4 we give some summary and concluding remarks. For completeness we catalog the various SQP secant methods in the Appendix.
CHAPTER 2

A Review of the Literature

2.1 Convergence Rate for the Pair \((x, \lambda)\)

Wilson (1963) formally presented the SQP Newton method, however it was undoubtedly known to workers in the calculus of variations in the first half of this century. He used the exact Hessian of the Lagrangian with respect to \(x\) as the Hessian of the quadratic objective function.

Garcia-Palomares and Mangasarian (1976) considered an SQP quasi-Newton method. The Hessian of the quadratic objective function was the approximation to the \(n \times n\) Hessian of the Lagrangian with respect to \(x\). In their algorithm an \((n + m) \times (n + m)\) matrix is updated in each iteration, but only the upper left \(n \times n\) submatrix is used in the quadratic programming subproblem. Under certain assumptions, they established various \(r\)-convergence results in \((x, \lambda)\).

Han (1976) improved the Garcia-Palomares and Mangasarian (1976) algorithm by presenting the SQP secant method as we know it today. He used a secant update directly on the \(n \times n\) matrix which approximated the Hessian of the (augmented) Lagrangian with respect to \(x\). He established local and \(q\)-superlinear convergence.
in \((z, \lambda)\) for several secant updates including the PSB and the DFP. The result for the latter update required the assumption that the Hessian of the Lagrangian with respect to \(x\) was positive definite at the solution.

Tapia (1977) defined the DMM algorithm. He showed that the diagonalized Newton multiplier method using the Newton multiplier update was equivalent to Newton's method on the extended problem \(\nabla L(z, \lambda) = 0\), here the differentiation is with respect to both \(z\) and \(\lambda\). Using the augmented Lagrangian and requiring the penalty constant \(\rho\) to be large enough so that positive definiteness of the Hessian of the augmented Lagrangian with respect to \(x\) near the solution was guaranteed, he established \(q\)-superlinear convergence in \((z, \lambda)\) for the diagonalized secant multiplier methods using the Newton multiplier update and Hessian updates given by the Broyden, PSB, DFP and BFGS secant updates. Therefore these results also implied that the SQP secant method was \(q\)-superlinearly convergent in \((z, \lambda)\) when the secant updates were Broyden, PSB, DFP and BFGS.

Glad (1979) proposed an algorithm which was essentially the diagonalized multiplier BFGS secant method. He used several multiplier updates including the Newton multiplier update. Using the augmented Lagrangian with an appropriate choice for the penalty constant \(\rho\), he established local and \(q\)-superlinear convergence in the pair \((z, \lambda)\) for the Newton multiplier update. Therefore Glad's result also gave local and \(q\)-superlinear convergence in \((z, \lambda)\) for the SQP and BFGS secant method.
through the equivalence established by Tapia (1978). Glad (1979) mentioned that the local and \( q \)-superlinear convergence in the pair \((x, \lambda)\) can also be obtained if the DFP secant update is used.

Han (1976), Tapia (1977) and Glad (1979) all used the Broyden-Dennis-Moré convergence theory for secant methods (1973) and applied it to a form of the extended system \( \nabla L(x, \lambda) = 0 \). Therefore it is no surprise that their convergence results are essentially the same for the various SQP secant methods.

## 2.2 Convergence Rate for \( x \)

In general a \( q \)-rate in \((x, \lambda)\) implies only the respective \( r \)-rate in \( x \). However, Tapia (1977) observed that if the multiplier update was independent of \( \lambda \), then the \( q \)-rate of \((x, \lambda)\) also applied to \( x \). Glad (1979) also made this observation.

Powell (1978) suggested a modification of the standard BFGS secant update for use in constrained optimization. Under the assumption that the sequence \( \{x_k\} \) generated by his modified BFGS secant method converged, he proved that the convergence of this sequence was \( r \)-superlinear.

Boggs, Tolle and Wang (1982) showed that the SQP method using the DFP or BFGS secant update gave \( q \)-superlinear convergence in \( x \) alone if the Hessian of the Lagrangian with respect to \( x \) at the solution was positive definite and \( \{x_k\} \) converged to \( x \), \( q \)-linearly. They also obtained a characterization of \( q \)-superlinear
convergence in \( x \) under the assumption that \( \{x_k\} \) converged to \( x^* \) \( q \)-linearly.

Fontecilla, Steihaug and Tapia (1987) proved that the diagonalized multiplier secant method with the Newton multiplier update, equivalently, the SQP secant method gave \( q \)-superlinear convergence in \( x \) for the PSB, DFP and BFGS updates. This result for DFP and BFGS requires the Hessian of the augmented Lagrangian with respect to \( x \) to be positive definite at the solution. They also obtained the Boggs-Tolle-Wang characterization of \( q \)-superlinear convergence in \( x \) without the assumption that \( \{x_k\} \) converged to \( x^* \) \( q \)-linearly.

Tapia (1988) proposed two new classes of SQP secant methods. One class of methods uses the SQP augmented Lagrangian formulation, while the other class uses the SQP Lagrangian formulation. He proved that in both cases the BFGS and DFP versions of the algorithm were locally convergent and gave \( q \)-superlinear convergence in \( x \) under the standard assumptions.

Coleman and Conn (1984) proposed a reduced Hessian algorithm and showed that their algorithm gave two-step \( q \)-superlinear convergence in \( x \) if the quasi-Newton updates were DFP or BFGS. Reduced Hessian or projected Hessian methods are a special case of SQP methods.

Nocedal and Overton (1985) presented two algorithms. They called them one-sided and two-sided projected Hessian updating algorithms. The two-sided projected Hessian updating algorithm is a reduced Hessian method approximating the
projected Hessian. They proved that the two-sided projected Hessian updating algorithm gave two-step $q$-superlinear convergence in $x$ if either the PSB, DFP or BFGS update was used. This is the standard convergence rate result for a reduced Hessian method.

2.3 Convergence Rate for $\lambda$

To our knowledge there are no results in the literature that concern the $q$-convergence rate of the multiplier in an SQP method or one of its equivalent formulations. Several recent articles in the literature make various assumptions concerning this rate. For example, Boggs and Tolle (1985) considered the SQP Lagrangian BFGS and DFP secant methods. They showed that the convergence in $x$ was $q$-superlinear assuming that the convergence in $x$ and in $\lambda$ was $q$-linear and $\{x_k\}$ satisfied a condition called tangential convergence. This result did not require the Hessian of the Lagrangian with respect to $x$ to be positive definite at the solution.

Gill, Murray, Saunders and Wright (1986) proposed an SQP Lagrangian secant method for generating a search direction and determined the step length from a line-search strategy with an augmented Lagrangian as the merit function. They established under the assumptions that the convergence in $x$ and $\lambda$ was $q$-superlinear and $\|\Delta x_k\|/\|\Delta \lambda_k\| > M > 0$. for $k$ sufficiently large, where $\Delta x_k = x_{k+1} - x_k$ and $\Delta \lambda_k = \lambda_{k+1} - \lambda_k$, that eventually their steplength choice would be 1.
Recently Tapia and Whitley (1987) investigated the convergence rate of the projected Newton method applied to the symmetric eigenvalue problem. This algorithm can be viewed as Newton SQP followed by a 2-norm normalization. They established a $q$-rate of convergence of $1 + \sqrt{2}$ for both the variable $x$ and the multiplier $\lambda$.

This result of Tapia and Whitley (1987) strengthened our already strong desire to determine relationships among the convergence rates for $(x, \lambda)$, for $x$ and for $\lambda$ in SQP methods.
CHAPTER 3

The New Results

3.1 Preliminaries

In this section we will state the standard assumptions that will be assumed throughout the remainder of the thesis and collect some preliminary results which will be used in later sections where we establish our new convergence rate results. In an effort to simplify our notation and give a cleaner presentation we will work exclusively with the SQP Lagrangian formulation in the remainder of this thesis, i.e., we will effectively choose $\rho = 0$ in the augmented Lagrangian. No loss of generality will result from this simplification as long as we remember that the requirement that the Hessian of the Lagrangian with respect to $x$ be positive definite at the solution, can be dealt with by working with the augmented Lagrangian and choosing $\rho$ sufficiently large.

Let $x_*$ be a local solution of problem (1.1) with associated multiplier $\lambda_*$. We will use the notation $\nabla h_k = \nabla h(x_k)$, $\nabla f_k = \nabla f(x_k)$, $\nabla h_* = \nabla h(x_*)$ and $A_* = \nabla^2_2 \ell(x_*, \lambda_*)$. Both the $\ell_2$ vector-norm and the corresponding induced matrix norm will be denoted by $| \cdot |$. We will use $\| \cdot \|$ to denote an arbitrary but fixed matrix
norm.

Throughout this work we make the following assumptions:

A1. The functions $f$ and $h$ have continuous second derivatives in an open neighborhood $D$ of a local solution $x_*$ of problem (1.1) and these second derivatives are Lipschitz continuous at $x_*$. 

A2. $\nabla h_*$ has full rank.

A3. $z^T A_* z > 0$ for all $z \neq 0$ satisfying $\nabla h_*^T z = 0$.

A4. The sequence $\{(x_k, \lambda_k)\}$ has been generated by a particular SQP quasi-Newton method and converges to $(x_*, \lambda_*)$.

A1, A2, and A3 are standard assumptions in the study of quasi-Newton methods for constrained optimization. Assumption A3 is the second-order sufficiency condition. The following lemmas are technical results and will be useful tools in the proof of several of our results.

**Lemma 3.1.1** There exists a sequence of matrices $\{\Gamma_k\}$ such that $\{\Gamma_k\}$ converges to $A_*$ and

$$x_{k+1} - x_* + B_k^{-1} \nabla h_k (\lambda_{k+1} - \lambda_*) = (I - B_k^{-1} \Gamma_k) (x_k - x_*).$$

*Proof.* By Theorem 1.1

$$x_{k+1} = x_k - B_k^{-1} \nabla_x \ell(x_k, \lambda_{k+1}).$$
Therefore

\[
x_{k+1} - x_* = x_k - x_* - B_k^{-1}[\nabla x_\ell(x_k, \lambda_{k+1}) - \nabla x_\ell(x_*, \lambda_*)] \\
= x_k - x_* - B_k^{-1}[\nabla f_k + \nabla h_k \lambda_{k+1} - \nabla f_* - \nabla h_* \lambda_*] \\
= B_k^{-1}[B_k(x_k - x_*) - (\nabla f_k - \nabla f_*) - \nabla h_k(\lambda_{k+1} - \lambda_*) - (\nabla h_k - \nabla h_*) \lambda_*] \\
= B_k^{-1} \left\{ B_k(x_k - x_*) - \left[ \int_0^1 \nabla^2 f(x_* + t(x_k - x_*)) dt \right] (x_k - x_*) \right\} - B_k^{-1} \nabla h_k(\lambda_{k+1} - \lambda_*)
\]

where the integral of the matrix-valued function is interpreted componentwise. (For more details see Dennis and Schnabel (1983), Chapter 4.)

Let \( \Gamma_k = \int_0^1 \nabla^2 f(x_* + t(x_k - x_*)) dt + \int_0^1 \nabla^2 h(x_* + t(x_k - x_*)) \lambda_* dt \). Then we have

\[
x_{k+1} - x_* = B_k^{-1}[B_k(x_k - x_*) - \Gamma_k(x_k - x_*)] - B_k^{-1} \nabla h_k(\lambda_{k+1} - \lambda_*),
\]
or

\[
x_{k+1} - x_* + B_k^{-1} \nabla h_k(\lambda_{k+1} - \lambda_*) = (I - B_k^{-1} \Gamma_k)(x_k - x_*).
\]

By the definition of \( \Gamma_k \) and the fact that \( \{(x_k, \lambda_k)\} \) converges to \( (x_*, \lambda_*) \) we have \( \{\Gamma_k\} \) converges to \( A_* \).

\( \square \)

**Lemma 3.1.2** We have

\[
\lambda_{k+1} - \lambda_* = (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1}[\nabla h_k^T(I - B_k^{-1} \Gamma_k)(x_k - x_*) + O(|x_k - x_*|^2)],
\]

where \( \{\Gamma_k\} \) is as in Lemma 3.1.1.
Proof. By Theorem 1.1 we have

\[ \lambda_{k+1} = \lambda_k + (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1} \left[ h_k - \nabla h_k^T B_k^{-1} \nabla x(\ell(x_k, \lambda_k)) \right] \]

\[ = \lambda_k + (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1} \left[ h_k - h_* - \nabla h_k^T B_k^{-1} \left[ \nabla x(\ell(x_k, \lambda_k)) - \nabla x(\ell(x_*, \lambda_*)) \right] \right] \]

Define \( \Gamma_k \) by the same formula used in Lemma 3.1.1 and perform the same algebra to obtain

\[ \lambda_{k+1} = \lambda_k + (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1} \left[ \left( \int_0^1 \nabla h(x_* + t(x_k - x_*))^T dt \right) (x_k - x_*), \right. \]

\[ - \nabla h_k^T B_k^{-1} \Gamma_k (x_k - x_*) - (\lambda_k - \lambda_*), \]

or

\[ \lambda_{k+1} - \lambda_* = (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1} \left[ \nabla h_k^T (I - B_k^{-1} \Gamma_k) (x_k - x_*) + O(|x_k - x_*|^2) \right]. \]

\[ \square \]

Considering \( a_k \geq 0 \) and \( b_k > 0 \) we write \( a_k = O(b_k) \) if there exists \( m \) such that \( \frac{a_k}{b_k} \leq m \) for all \( k \). We now use Lemma 3.1.2 to establish that \( |\lambda_{k+1} - \lambda_*| = O(|x_k - x_*|) \) whenever \( \{B_k\} \) and \( \{B_k^{-1}\} \) are bounded.

**Theorem 3.1.1** If \( \{B_k\} \) and \( \{B_k^{-1}\} \) are bounded, then \( |\lambda_{k+1} - \lambda_*| = O(|x_k - x_*|) \).

**Proof.** From Lemma 3.1.2 we have

\[ \lambda_{k+1} - \lambda_* = (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1} \left[ \nabla h_k^T (I - B_k^{-1} \Gamma_k) (x_k - x_*) + O(|x_k - x_*|^2) \right] \quad (3.1.1) \]
From Lemma 3.7 in Fontecilla (1988) we have

\[ |(\nabla h_k^T B_k^{-1} \nabla h_k)^{-1}| \leq |(\nabla h_k^T \nabla h_k)^{-1} \nabla h_k^T|^2 |B_k|. \]

Since \( \{B_k\} \) is bounded and \( \nabla h_* \) has full rank, it follows that there exists \( C_1 \) such that

\[ |(\nabla h_k^T B_k^{-1} \nabla h_k)^{-1}| \leq C_1 \quad (3.1.2) \]

Furthermore

\[
\begin{align*}
|\nabla h_k^T (I - B_k^{-1} \Gamma_k)| \\
= |\nabla h_k^T B_k^{-1}(B_k - \Gamma_k)| \\
\leq |\nabla h_k^T| |B_k^{-1}| |B_k - \Gamma_k|.
\end{align*}
\]

The facts that \( \{B_k^{-1}\} \) is bounded and \( \{\Gamma_k\} \) converges to \( A_* \) lead to

\[ |\nabla h_k^T (I - B_k^{-1} \Gamma_k)| \leq C_2 \quad (3.1.3) \]

for some constant \( C_2 \).

Combining (3.1.1), (3.1.2) and (3.1.3) we have \( \frac{|\lambda_{k+1} - \lambda_*|}{|x_k - x_*|} \leq C_3 \) for some constant \( C_3 \).

Therefore \( |\lambda_{k+1} - \lambda_*| = O(|x_k - x_*|) \). \( \square \)

Using Theorem 3.1.1 we can establish the finiteness of the \( q_1 \) factor of the sequence \( \{x_k\} \).

**Corollary 3.1.1** If \( \{B_k\} \) and \( \{B_k^{-1}\} \) are bounded, then \( |x_{k+1} - x_*| = O(|x_k - x_*|) \).
Proof. By Lemma 3.1.1 we have

\[ x_{k+1} - x_* - (I - B_k^{-1} \Gamma_k)(x_k - x_*) = -B_k^{-1} \nabla h_k(\lambda_{k+1} - \lambda_*). \quad (3.1.4) \]

From Theorem 3.1.1 and the fact that \( \{B_k^{-1}\} \) is bounded we have

\[
\frac{|B_k^{-1} \nabla h_k(\lambda_{k+1} - \lambda_*)|}{|x_k - x_*|} \leq |B_k^{-1}| \frac{|\nabla h_k| |\lambda_{k+1} - \lambda_*|}{|x_k - x_*|} \leq C_1
\]

for some constant \( C_1 \). Furthermore

\[
\frac{|(I - B_k^{-1} \Gamma_k)(x_k - x_*)|}{|x_k - x_*|} = \frac{|B_k^{-1}(B_k - \Gamma_k)(x_k - x_*)|}{|x_k - x_*|} \leq |B_k^{-1}| |B_k - \Gamma_k|.
\]

By assumption \( \{|B_k^{-1}\}| \) is bounded, therefore

\[
\frac{|(I - B_k^{-1} \Gamma_k)(x_k - x_*)|}{|x_k - x_*|} \leq C_2
\]

for some constant \( C_2 \). Combining (3.1.4), (3.1.5), and (3.1.6) we have

\[ |x_{k+1} - x_*| = O \left(|x_k - x_*|\right). \]

The requirement that \( \{B_k\} \) and \( \{B_k^{-1}\} \) be bounded is quite mild. In general if the quasi-Newton update satisfies bounded deterioration, then we have that \( \{B_k\} \) and \( \{B_k^{-1}\} \) are bounded if the initial \((x_0, B_0)\) is close to \((x_*, A_*)\). Moreover, the well-known secant updates Broyden, PSB, DFP and BFGS all satisfy bounded deterioration. For more details see the argument used in Broyden, Dennis and Moré (1973) or Theorems 3.1 and 3.2 in Fontecilla, Steihaug and Tapia (1987).
3.2 Convergence Rate for $x$

Han (1976), (1977), Tapia (1977) and Glad (1978) independently established local and $q$-superlinear convergence for the pair $(x, \lambda)$ for various SQP secant methods as mentioned in Section 2.1. In general, $q$-superlinear convergence for the pair $(x, \lambda)$ only implies $r$-superlinear convergence for $x$. However, we will now show that for the SQP quasi-Newton method the $q$-superlinear convergence of the pair $(x, \lambda)$ implies at least two-step $q$-superlinear convergence for $x$ provided that $\{B_k\}$ and $\{B^{-1}_k\}$ are bounded. The result will follow from the following lemma.

Lemma 3.2.1 Let $a_k > 0$, $b_k \geq 0$, $a_{k+1} = O(a_k)$ and $b_{k+1} = O(a_k)$. If $\{(a_k, b_k)\}$ converges to $(0, 0)$ $q$-superlinearly, then the convergence of $\{a_k\}$ to 0 is at least two-step $q$-superlinear. Moreover

$$\lim_{k \to \infty} \frac{b_{k+1}}{a_{k-1}} = 0.$$  

Proof. Let $b_{k+1} = M_k a_k$. By assumption

$$\lim_{k \to \infty} \sqrt{\frac{a_{k+1}^2 + b_{k+1}^2}{a_k^2 + b_k^2}} = 0.$$  

So

$$\lim_{k \to \infty} \sqrt{\frac{a_{k+1}^2 + b_{k+1}^2}{a_k^2 + M_{k-1}^2 a_{k-1}^2}} = 0.$$  

Hence

$$\lim_{k \to \infty} \sqrt{\frac{a_k^2 + b_k^2}{a_{k-1}^2 + M_{k-1}^2}} = 0. \quad (3.2.1)$$
Since $a_k = O(a_{k-1})$ and $\{M_k\}$ is bounded, it follows from (3.2.1) that $\lim_{k \to \infty} \frac{a_{k+1}}{a_{k-1}} = 0$ and $\lim_{k \to \infty} \frac{b_{k+1}}{a_{k-1}} = 0$. 

**Theorem 3.2.1** If $\{(x_k, \lambda_k)\}$ converges to $(x_*, \lambda_*)$ $q$-superlinearly and $\{B_k\}$ and $\{B_k^{-1}\}$ are bounded, then $\{x_k\}$ converges to $x_*$ at least two-step $q$-superlinearly and

\[
\lim_{k \to \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|x_{k+1} - x_*|} = 0.
\]

**Proof.** Let $a_k = |x_k - x_*|$ and $b_k = |\lambda_k - \lambda_*|$. From Theorem 3.1.1 and Corollary 3.1.1 we have

\[
b_{k+1} = O(a_k) \quad \text{and} \quad a_{k+1} = O(a_k).
\]

Therefore from Lemma 3.2.1 we have two-step $q$-superlinear convergence of $\{x_k\}$ and $\lim_{k \to \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|x_{k-1} - x_*|} = 0$. 

We now give a condition that ensures $q$-superlinear convergence of $\{x_k\}$ provided that the pair $\{(x_k, \lambda_k)\}$ converges $q$-superlinearly. Furthermore the well-known SQP secant methods satisfy this condition. Let $\{n_k\}$ be a sequence of positive integers. We write $\Delta x_{n_k}$ for $x_{n_k+1} - x_{n_k}$.

**Theorem 3.2.2** Let $P_k = I - \nabla h_k(\nabla h_k^T \nabla h_k)^{-1} \nabla h_k$. Assume that

\[
\lim_{k \to \infty} \frac{|P_{n_k}(B_{n_k} - A_*)\Delta x_{n_k}|}{|\Delta x_{n_k}|} = 0
\]

for any sequence of positive integers $\{n_k\}$ such that $\lim_{k \to \infty} \frac{|\lambda_{n_k} - \lambda_*|}{|x_{n_k} - x_*|} = \infty$. If $\{(x_k, \lambda_k)\}$ converges to $(x_*, \lambda_*)$ $q$-superlinearly, then we also have that $\{x_k\}$ converges to $x_*$ $q$-superlinearly.
Proof. By assumption we have

\[ |x_{k+1} - x_*| \leq |(x_{k+1} - x_*, \lambda_{k+1} - \lambda_*)| \]
\[ \leq c_k |(x_k - x_*, \lambda_k - \lambda_*)| \]
\[ \leq c_k (|x_k - x_*| + |\lambda_k - \lambda_*|), \]

where \( \{c_k\} \) converges to 0. It follows that

\[ \frac{|x_{k+1} - x_*|}{|x_k - x_*|} \leq c_k \left( 1 + \frac{|\lambda_k - \lambda_*|}{|x_k - x_*|} \right). \]  \hspace{1cm} (3.2.2)

Suppose

\[ \lim_{k \to \infty} \frac{|x_{k+1} - x_*|}{|x_k - x_*|} = \delta > 0. \]

Then there exists a sequence of positive integers \( \{n_k\} \) such that

\[ \lim_{k \to \infty} \frac{|x_{n_k+1} - x_*|}{|x_{n_k} - x_*|} = \delta. \]  \hspace{1cm} (3.2.3)

Case (i) If

\[ \lim_{k \to \infty} \frac{|\lambda_{n_k} - \lambda_*|}{|x_{n_k} - x_*|} < \infty, \]

then from (3.2.2) we have

\[ \lim_{k \to \infty} \frac{|x_{n_k+1} - x_*|}{|x_{n_k} - x_*|} = 0. \]

This contradicts (3.2.3).
Case(ii) If 
\[ \lim_{k \to \infty} \frac{|\lambda_{n_k} - \lambda_*|}{|x_{n_k} - x_*|} = \infty, \]
then there exists a subsequence of \( \{n_k\} \), say \( \{m_k\} \), such that 
\[ \lim_{k \to \infty} \frac{|\lambda_{m_k} - \lambda_*|}{|x_{m_k} - x_*|} = \infty. \]

By hypothesis we have 
\[ \lim_{k \to \infty} \frac{|P_{m_k} (B_{m_k} - A_*) \Delta x_{m_k}|}{|\Delta x_{m_k}|} = 0. \tag{3.2.4} \]

If we look closely at the proof of the characterization of \( q \)-superlinear convergence in Dennis and Moré (1974), we see that it can also be used for subsequences. Hence by the characterization of \( q \)-superlinear convergence for \( x \) (Boggs, Tolle and Wang (1982)) we have 
\[ \lim_{k \to \infty} \frac{|x_{m_k+1} - x_*|}{|x_{m_k} - x_*|} = 0. \]
This also contradicts (3.2.3). Both Case (i) and Case (ii) lead to a contradiction. Hence we must have \( \delta = 0 \) and it follows that \( \lim_{k \to \infty} \frac{|x_{k+1} - x_*|}{|x_k - x_*|} = 0. \)

The known result that the SQP Broyden, PSB, DFP and BFGS secent methods give \( q \)-superlinear convergence in \( x \) follows directly from Theorem 3.2.2. When we say that an SQP quasi-Newton method is locally convergent we mean that there exist \( \epsilon \) and \( \delta \) such that if \( |x_0 - x_*| < \epsilon \) and \( |B_0 - A_*| < \delta \) then the method is well-defined and \( \{x_k\} \) converges to \( x_* \).
Corollary 3.2.1 The SQP Broyden, PSB, DFP and BFGS secant methods, in the case of DFP and BFGS we assume that the matrix $A_*$ is positive definite, are locally convergent and both \{$(x_k, \lambda_k)$\} and \{$x_k$\} converge $q$-superlinearly.

Proof. The local convergence and the fact that \{$(x_k, \lambda_k)$\} converges $q$-superlinearly has been established independently by Glad (1979), Han (1976), (1977) and Tapia (1977). Moreover, the Broyden, PSB, DFP and BFGS secant methods satisfy

$$\lim_{k \to \infty} \frac{|(B_k - A_*) \Delta x_k|}{|\Delta x_k|} = 0.$$ \hfill (3.2.5)

For the details see the argument used in Broyden, Dennis and Moré (1973) or Corollary 5.5 in Fontecilla, Steihaug, and Tapia (1987). Formula (3.2.5) will imply

$$\lim_{k \to \infty} \frac{|P_k(B_k - A_*) \Delta x_k|}{|\Delta x_k|} = 0.$$ 

It follows from Theorem 3.2.2 that \{$x_k$\} converges to $x_*$ $q$-superlinearly.

The $q$-superlinear convergence in $x$ of these secant methods was originally established by Boggs, Tolle and Wang (1982). See also Fontecilla, Steihaug and Tapia (1987).

Corollary 3.2.2 If \{$(x_k, \lambda_k)$\} converges to $(x_*, \lambda_*)$ $q$-superlinearly, then the following two statements are equivalent:

1. \(\lim_{k \to \infty} \frac{|P_{n_k}(B_{n_k} - A_*) \Delta x_{n_k}|}{|\Delta x_{n_k}|} = 0\) for any subsequence \{\$n_k$\} of \{\$k$\} such that \(\lim_{k \to \infty} \frac{\lambda_{n_k} - \lambda_*}{x_{n_k} - x_*} = \infty\).
(ii) \( \{x_k\} \) converges to \( x_* \) \( q \)-superlinearly.

**Proof.** The fact that (ii) implies (i) follows from the Boggs-Tolle-Wang characterization of \( q \)-superlinear convergence, see Boggs, Tolle and Wang (1982) or Fontecilla, Steihaug and Tapia (1987). The fact that (i) implies (ii) follows from Theorem 3.2.2.

\[ \square \]

At this point we give another condition that will ensure \( q \)-superlinear convergence in \( x \) provided that we have \( q \)-superlinear convergence in the pair \( (x, \lambda) \).

**Theorem 3.2.3** Assume \( \frac{|\Delta x_k|}{|\Delta \lambda_k|} > M > 0 \) for some constant \( M \) and \( k \) sufficiently large. If \( \{(x_k, \lambda_k)\} \) converges to \( (x_*, \lambda_*) \) \( q \)-superlinearly, then \( \{x_k\} \) converges to \( x_* \) \( q \)-superlinearly.

**Proof.** According to the Dennis-Moré superlinear convergence criterion (see Dennis-Moré (1974)) we have \( q \)-superlinear convergence of the sequence \( \{(x_k, \lambda_k)\} \) (see Tapia (1977) for details) if and only if

\[
\lim_{k \to \infty} \frac{|(B_k - A_*)\Delta x_k + (\nabla h_k - \nabla h_*)(\Delta \lambda_k)|}{\sqrt{|\Delta x_k|^2 + |\Delta \lambda_k|^2}} = 0 \tag{3.2.6}
\]

and

\[
\lim_{k \to \infty} \frac{|(\nabla h_k - \nabla h_*)^T \Delta x_k|}{\sqrt{|\Delta x_k|^2 + |\Delta \lambda_k|^2}} = 0 \tag{3.2.7}
\]

From (3.2.6) we have

\[
\lim_{k \to \infty} \frac{|(B_k - A_*)\Delta x_k + (\nabla h_k - \nabla h_*)(\Delta \lambda_k)|}{\sqrt{1 + \frac{|\Delta \lambda_k|^2}{|\Delta x_k|^2}}} = 0 \tag{3.2.8}
\]
By assumption \( \frac{|\Delta \lambda_k|}{|\Delta x_k|} < \frac{1}{M} \) for \( k \) sufficiently large. It follows that (3.2.8) implies
\[
\lim_{k \to \infty} \frac{|(B_k - A_\star) \Delta x_k|}{|\Delta x_k|} = 0.
\]
Immediately we have
\[
\lim_{k \to \infty} \frac{|P_k(B_k - A_\star) \Delta x_k|}{|\Delta x_k|} = 0.
\]
Therefore from Theorem 3.2.2 we have \( \{x_k\} \) converges to \( x_\star \) \( q \)-superlinearly. \( \square \)

In their work Gill, Murray, Saunders and Wright (1986) made three assumptions in their local convergence analysis. These three assumptions were (i) \( \{x_k\} \) converged to \( x_\star \) \( q \)-superlinearly, (ii) \( \{\lambda_k\} \) converged to \( \lambda_\star \) \( q \)-superlinearly and (iii) \( \frac{|\Delta x_k|}{|\Delta \lambda_k|} > M > 0 \) for \( k \) sufficiently large. We know from Theorem 3.2.3 that if the pair \( (x, \lambda) \) converges \( q \) superlinearly and (iii) holds then (i) will hold. Hence Gill, Murray, Saunders and Wright could have assumed superlinear convergence in the pair instead of (i).

### 3.3 Convergence Rate for \( \lambda \)

In this section we will show that if the pair \( (x, \lambda) \) and the variable \( x \) converge \( q \)-superlinearly, then the multiplier \( \lambda \) converges \( q \)-superlinearly or \( q \)-sublinearly with unbounded \( q_1 \) factor. The following lemma is a technical result which will be used later.
Lemma 3.3.1 Let \( \{x_{n_k}\} \) and \( \{\lambda_{n_k}\} \) be subsequences of \( \{x_k\} \) and \( \{\lambda_k\} \). Assume \( \{B_k^{-1}\} \) is bounded. If
\[
\lim_{k \to \infty} \frac{|\lambda_{n_k} - \lambda_*|}{|x_{n_k} - x_*|} = 0,
\]
then
\[
\lim_{k \to \infty} \frac{|(I - B_{n_k}^{-1} \Gamma_{n_k-1})(x_{n_k-1} - x_*)|}{|x_{n_k} - x_*|} = 1
\]
where \( \Gamma_k \) is as in Lemma 3.1.1.

Proof. By Lemma 3.1.1 we have
\[
\frac{x_{n_k} - x_* - (I - B_{n_k}^{-1} \Gamma_{n_k-1})(x_{n_k-1} - x_*)}{|x_{n_k} - x_*|} = -B_{n_k}^{-1} \nabla h_{n_k-1} \frac{(\lambda_{n_k} - \lambda_*)}{|x_{n_k} - x_*|}.
\]
By assumption \( \lim_{k \to \infty} \frac{|\lambda_{n_k} - \lambda_*|}{|x_{n_k} - x_*|} = 0 \) and \( \{B_k^{-1}\} \) is bounded. Therefore we have
\[
\lim_{k \to \infty} \frac{|B_{n_k}^{-1} \nabla h_{n_k-1}(\lambda_{n_k} - \lambda_*)|}{|x_{n_k} - x_*|} = 0.
\]
It follows from (3.3.1) that
\[
\lim_{k \to \infty} \frac{|x_{n_k} - x_* - (I - B_{n_k}^{-1} \Gamma_{n_k-1})(x_{n_k-1} - x_*)|}{|x_{n_k} - x_*|} = 0.
\]
Therefore
\[
\lim_{k \to \infty} \frac{|(I - B_{n_k}^{-1} \Gamma_{n_k-1})(x_{n_k-1} - x_*)|}{|x_{n_k} - x_*|} = 1. \quad \Box
\]

Lemma 3.3.2 Let \( Q_k V_k = W_k \) where \( \{Q_k\} \) is a sequence of \( n \times m \) matrices with full column rank and \( \{V_k\} \) is a sequence of vectors. Assume \( V_k \neq 0 \) for \( k \) sufficiently large, \( \{Q_k\} \) is bounded, and the limit points of \( \{Q_k\} \) have full column rank. If \( \{W_k\} \) converges to 0 \( q \)-superlinearly then \( \{V_k\} \) converges to 0 \( q \)-superlinearly.
Proof. By the assumption that \( \{Q_k\} \) is bounded and each \( \{Q_k\} \) is of full column rank and the limit points of \( \{Q_k\} \) have full column rank we have

\[
m|V_k| \leq |W_k| \leq M|V_k|
\]

for some positive constants \( m \) and \( M \) independent of \( k \). Therefore the fact that \( \{W_k\} \) converges to 0 \( q \)-superlinearly will imply \( \{V_k\} \) converges to 0 \( q \)-superlinearly. \( \square \)

**Theorem 3.3.1** If \( \{(x_k, \lambda_k)\} \) and \( \{x_k\} \) converge to \((x_\ast, \lambda_\ast)\) and \( x_\ast \) \( q \)-superlinearly and \( \{B_k\} \) and \( \{B_k^{-1}\} \) are bounded, then either

(i) \( \lim_{k \to \infty} \frac{|\lambda_{k+1} - \lambda_\ast|}{|\lambda_k - \lambda_\ast|} = 0 \) (i.e. \( \lambda_k \) converges \( q \)-superlinearly)

or

(ii) \( \lim_{k \to \infty} \frac{|\lambda_{k+1} - \lambda_\ast|}{|\lambda_k - \lambda_\ast|} = \infty \) (i.e. \( \lambda_k \) converges \( q \)-sublinearly with unbounded \( q_1 \) factor).

Proof. If \( |\lambda_{k+1} - \lambda_\ast| = O(|\lambda_k - \lambda_\ast|) \) is not true, then

\[
\lim_{k \to \infty} \frac{|\lambda_{k+1} - \lambda_\ast|}{|\lambda_k - \lambda_\ast|} = \infty .
\]

Now suppose that \( |\lambda_{k+1} - \lambda_\ast| = O(|\lambda_k - \lambda_\ast|) \). By hypothesis we have

\[
|\lambda_{k+1} - \lambda_\ast| \leq |(x_{k+1} - x_\ast, \lambda_{k+1} - \lambda_\ast)|
\]
\[
\leq c_k|(x_k - x_\ast, \lambda_k - \lambda_\ast)|
\]
\[
\leq c_k(|x_k - x_\ast| + |\lambda_k - \lambda_\ast|),
\]
where \( \{c_k\} \) converges to 0. It follows that

\[
\frac{|\lambda_{k+1} - \lambda_*|}{|\lambda_k - \lambda_*|} \leq c_k \left( 1 + \frac{|x_k - x_*|}{|\lambda_k - \lambda_*|} \right).
\]  

(3.3.2)

Suppose

\[
\lim_{k \to \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|\lambda_k - \lambda_*|} = \delta > 0.
\]

Then there exists a sequence of positive integers \( \{n_k\} \) such that

\[
\lim_{k \to \infty} \frac{|\lambda_{n_k+1} - \lambda_*|}{|\lambda_{n_k} - \lambda_*|} = \delta.
\]  

(3.3.3)

Case (i) If

\[
\lim_{k \to \infty} \frac{|x_{n_k} - x_*|}{|\lambda_{n_k} - \lambda_*|} < \infty,
\]

then from (3.3.2) we have

\[
\lim_{k \to \infty} \frac{|\lambda_{n_k+1} - \lambda_*|}{|\lambda_{n_k} - \lambda_*|} = 0.
\]

However, this contradicts (3.3.3).

Case (ii) If

\[
\lim_{k \to \infty} \frac{|x_{n_k} - x_*|}{|\lambda_{n_k} - \lambda_*|} = \infty,
\]

then there exists a subsequence of \( \{n_k\} \), say \( \{m_k\} \), such that

\[
\lim_{k \to \infty} \frac{|\lambda_{m_k} - \lambda_*|}{|x_{m_k} - x_*|} = 0.
\]  

(3.3.4)
Now
\[ \frac{|\lambda_{k+1} - \lambda_k|}{|x_k - x_*|} \leq \frac{|\lambda_{k+1} - \lambda_*|}{|x_k - x_*|} + \frac{|\lambda_k - \lambda_*|}{|x_k - x_*|}. \]

From (3.3.4) and the assumption that $|\lambda_{k+1} - \lambda_*| = O(|\lambda_k - \lambda_*|)$ we have
\[ \lim_{k \to \infty} \frac{|\lambda_{m_k+1} - \lambda_{m_k}|}{|x_{m_k} - x_*|} = 0. \]

From (3.3.4) and Lemma 3.3.1 we have
\[ \lim_{k \to \infty} \frac{|(I - B_{m_k-1}^{-1}) (x_{m_k-1} - x_*)|}{|x_{m_k} - x_*|} = 1. \tag{3.3.6} \]

By the characterization of $q$-superlinear convergence of $\{(x_k, \lambda_k)\}$ (see Dennis and Moré (1974) or Tapia (1977) for details) we have
\[ \lim_{k \to \infty} \frac{|(B_k - A_*) \Delta x_k + (\nabla h_k - \nabla h_*) \Delta \lambda_k|}{\sqrt{||\Delta x_k|^2 + |\Delta \lambda_k|^2}} = 0. \]

It follows that
\[ \lim_{k \to \infty} \frac{|(B_k - A_*) \frac{\Delta x_k}{|\Delta x_k|} + (\nabla h_k - \nabla h_*) \frac{\Delta \lambda_k}{|\Delta \lambda_k|}|}{\sqrt{1 + \frac{|\Delta \lambda_k|^2}{|\Delta x_k|^2}}} = 0. \tag{3.3.7} \]

By (3.3.5) and the fact that $\{x_k\}$ converges to $x_*$ $q$-superlinearly we have from (3.3.7)
\[ \lim_{k \to \infty} \frac{|(B_{m_k} - A_*) \Delta x_{m_k}|}{|\Delta x_{m_k}|} = 0. \tag{3.3.8} \]

Since $\{\Gamma_k\}$ converges to $A_*$,
\[ \lim_{k \to \infty} \frac{|(B_{m_k} - \Gamma_{m_k}) (x_{m_k} - x_*)|}{|x_{m_k} - x_*|} = 0. \tag{3.3.9} \]
Let $W_{k+1} = x_{k+1} - x_* - B_k^{-1}(B_k - \Gamma_k)(x_k - x_*)$ and $Q_k = -B_k^{-1}\nabla h_k$. Then from Lemma 3.1.1 we have

$$W_{k+1} = Q_k(\lambda_{k+1} - \lambda_*).$$

$$\frac{|W_{k+1}|}{|W_k|} = \frac{|x_{k+1} - x_* - B_k^{-1}(B_k - \Gamma_k)(x_k - x_*)|}{|x_k - x_* - B_{k-1}^{-1}(B_{k-1} - \Gamma_{k-1})(x_{k-1} - x_*)|}$$

$$= \frac{|\Delta x_k + B_k^{-1}\Gamma_k(x_k - x_*)|}{|\Delta x_{k-1} + B_{k-1}^{-1}\Gamma_{k-1}(x_{k-1} - x_*)|}. \quad (3.3.10)$$

Since $\{x_k\}$ converges to $x_*$ $q$-superlinearly we have

$$\lim_{k \to \infty} \frac{|W_{m_k+1}|}{|W_{m_k}|} = \lim_{k \to \infty} \frac{|(B_{m_k}^{-1}\Gamma_{m_k} - I)(x_{m_k} - x_*)|}{|(B_{m_k}^{-1}\Gamma_{m_k-1} - I)(x_{m_k-1} - x_*)|}$$

$$= \lim_{k \to \infty} \frac{|(B_{m_k}^{-1}\Gamma_{m_k} - I)(x_{m_k} - x_*)|}{|(B_{m_k}^{-1}\Gamma_{m_k-1} - I)(x_{m_k-1} - x_*)|} \cdot \frac{|x_{m_k} - x_*|}{|x_m_x - x_*|}.$$

From (3.3.6), (3.3.9), and the fact that $\{B_k^{-1}\}$ is bounded we have

$$\lim_{k \to \infty} \frac{|W_{m_k+1}|}{|W_{m_k}|} = 0.$$

This contradicts (3.3.3).

Both case (i) and case (ii) lead to a contradiction. Hence we must have $\delta = 0$ and it follows that

$$\lim_{k \to \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|\lambda_k - \lambda_*|} = 0. \quad \square$$

A highly desirable feature of an iterative procedure is the property that should an iterate happen to coincide with a solution, then the subsequent iterate is also equal to the solution. Clearly, an iterative procedure which lacks this fundamental
property cannot have good theoretical $q$-convergence behavior. The error could be zero at one iteration and nonzero in the subsequent iteration. This implies that in any analysis which considers the worst case, the $q_1$-factor would be unbounded. Even if the error were not zero at any iteration it could be arbitrarily small and one would expect similar statements to hold.

Let us now look at the SQP iterative procedure in terms of $(x, \lambda)$, $x$ and $\lambda$ from this point of view. It follows that if $x_k = x_*$, then $\nabla f_k = -\nabla h_k \lambda_*$ and $h_k = 0$; so from (1.12) $\lambda_{k+1} = \lambda_*$ and from (1.9) and (1.10) $x_{k+1} = x_*$. Therefore the establishment of good $q$-convergence behavior in $(x, \lambda)$ and in $x$ for the SQP method should not be viewed as a complete surprise.

From (3.1.1) we see that $\lambda_{k+1}$ does not depend explicitly on $\lambda_k$. We should not expect to have $\lambda_{k+1} = \lambda_*$ whenever $\lambda_k = \lambda_*$. Moreover, in most cases there will exist a manifold $\Omega \in \mathbb{R}^n$ of dimension $n - m$ such that $\lambda_{k+1} = \lambda_*$ whenever $x_k \in \Omega$. It follows that in the worst-case analysis given by Theorem 3.3.1 the unbounded $q_1$-factor situation is to be expected and cannot be removed from the theorem. The surprise is that Theorem 3.3.1 says that if the $q$-convergence behavior in $\lambda$ is not arbitrarily bad (unbounded $q_1$-factor), then it is essentially optimal ($q_1$-factor of zero). It is interesting that both notions are norm independent. We believe that while our numerical experience dictates that in most cases we should expect $q$-superlinear convergence in $\lambda$, Theorem 3.3.1 is actually sharp.
It is interesting to point out that the r-convergence in \( \lambda \) is always superlinear and the unbounded \( q_1 \)-factor occurs because the estimate of the multiplier is exceptionally good an infinite number of times.

It is also interesting to point out that in the modified SQP method studied by Tapia and Whitley (1988) if it happens that \( \lambda_k = \lambda_\ast \), then the algorithm will converge in the subsequent iteration, i.e., \((x_{k+1}, \lambda_{k+1}) = (x_\ast, \lambda_\ast)\). This is due to the very special structure of the eigenvalue problem. Hence it is not unreasonable that they were able to establish the surprising \( q \)-convergence rate of \( 1 + \sqrt{2} \) for the pair \((x, \lambda)\), the variable \( x \) and the multiplier \( \lambda \).

### 3.4 Convergence Rate for \((x, \lambda)\), \( x \) and \( \lambda \)

In this section we will discuss relationships among the convergence rates of \((x, \lambda)\), \( x \) and \( \lambda \). Of course Theorem 3.3.1 already gives one such relationship. The following result holds for any sequences \( \{x_k\} \) and \( \{\lambda_k\} \) no matter whether they were generated by an SQP quasi-Newton method or not.

**Theorem 3.4.1** If \( \{x_k\} \) converges to \( x_\ast \) \( q \)-superlinearly and \( \{\lambda_k\} \) converges to \( \lambda_\ast \) \( q \)-superlinearly, then \( \{(x_k, \lambda_k)\} \) converges to \((x_\ast, \lambda_\ast)\) \( q \)-superlinearly.

**Proof.** Since \( q \)-superlinear convergence is independent of norm, we can work with the max norm.
Let $\| \cdot \|$ denote the max norm. By assumption there exist \( \{ c_k \} \) and \( \{ \tilde{c}_k \} \) such that

\[
\| x_{k+1} - x_* \| \leq c_k \| x_k - x_* \| 
\]

and

\[
\| \lambda_{k+1} - \lambda_* \| \leq \tilde{c}_k \| \lambda_k - \lambda_* \| 
\]  \hspace{1cm} (3.4.1)

where \( \{ c_k \} \) and \( \{ \tilde{c}_k \} \) converge to 0. Then

\[
\| (x_{k+1}, \lambda_{k+1}) - (x_*, \lambda_*) \| = \| (x_{k+1} - x_*, \lambda_{k+1} - \lambda_*) \|
\]

\[
= \max \{ \| x_{k+1} - x_* \|, \| \lambda_{k+1} - \lambda_* \| \}
\]

\[
\leq \max \{ c_k \| x_k - x_* \|, \tilde{c}_k \| \lambda_k - \lambda_* \| \}
\]

\[
\leq (\max \{ c_k, \tilde{c}_k \}) \| (x_k - x_*, \lambda_k - \lambda_*) \| .
\]

So

\[
\| (x_{k+1}, \lambda_{k+1}) - (x_*, \lambda_*) \| \leq (\max \{ c_k, \tilde{c}_k \}) \| (x_k - x_*, \lambda_k - \lambda_*) \| . 
\]  \hspace{1cm} (3.4.2)

Since \( \{ c_k \} \) and \( \{ \tilde{c}_k \} \) converge to zero, then

\[
\lim_{k \to \infty} \max \{ c_k, \tilde{c}_k \} = 0 .
\]  \hspace{1cm} (3.4.3)

It follows that \( \{ (x_k, \lambda_k) \} \) converges to \((x_*, \lambda_*)\) q-superlinearly.

The following result gives a condition that guarantees that \((x, \lambda), x \) and \( \lambda \) all converge q-superlinearly.

**Theorem 3.4.2** Assume \( \{ B_k \} \) and \( \{ B_k^{-1} \} \) are bounded. If \( \lim_{k \to \infty} \frac{|\lambda_k - \lambda_*|}{|x_k - x_*|} = \infty, \)

then
(i) \( \{x_k\} \) converges to \( x_* \) \( q \)-superlinearly,

(ii) \( \{\lambda_k\} \) converges to \( \lambda_* \) \( q \)-superlinearly,

and

(iii) \( \{(x_k, \lambda_k)\} \) converges to \( (x_*, \lambda_*) \) \( q \)-superlinearly.

Proof. By assumption \( \lim_{k \to \infty} \frac{|\lambda_k - \lambda_*|}{|x_k - x_*|} = \infty \) and by Theorem 3.1.1 we have 
\( \frac{|\lambda_{k+1} - \lambda_*|}{|x_k - x_*|} \) is bounded. It follows that \( \lim_{k \to \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|\lambda_k - \lambda_*|} = 0 \). This proves (ii).

Furthermore the fact that \( |\lambda_{k+1} - \lambda_*| = O(|x_k - x_*|) \) implies that
\[
O\left(\frac{|x_k - x_*|}{|x_k - x_*|}\right) = \frac{|\lambda_{k+1} - \lambda_*|}{|x_k - x_*|}.
\]
So we have
\[
\lim_{k \to \infty} \frac{|x_k - x_*|}{|x_{k+1} - x_*|} = \infty.
\]

Therefore
\[
\lim_{k \to \infty} \frac{|x_{k+1} - x_*|}{|x_k - x_*|} = 0.
\]

This proves (i).

By Theorem 3.4.1, (i) and (ii) give (iii).

Observe that the condition given in Theorem 3.4.2 precludes \( \lambda_k = \lambda_* \) an infinite number of times. The example (1) studied by Byrd (1985) satisfies this condition if we choose the quasi-Newton update for the full Hessian appropriately. For more details see Byrd (1985). However, this is just a sufficient condition. The projected Newton method for the symmetric eigenvalue problem studied by Tapia and Whitley
(1988) does not satisfy this condition. But it still gives the same convergence rate for the pair \((x, \lambda)\), the variable \(x\) and the multiplier \(\lambda\).

Gill, Murray, Saunders and Wright (1986) made the assumption \(\frac{|\Delta x_k|}{|\Delta \lambda_k|} > M > 0\) in the local analysis of their SQP algorithm. This assumption will lead to the following result.

**Theorem 3.4.3** Assume that \(\{(x_k, \lambda_k)\}\) converges to \((x_*, \lambda_*)\) \(q\)-superlinearly, \(\{B_k\}\) and \(\{B_k^{-1}\}\) are bounded and \(\frac{|\Delta x_k|}{|\Delta \lambda_k|} > M > 0\) for \(k\) sufficiently large. Then

(i) \(\{x_k\}\) converges to \(x_*\) \(q\)-superlinearly.

(ii) \(\{\lambda_k\}\) converges to \(\lambda_*\) \(q\)-superlinearly or \(q\)-sublinearly with unbounded \(q_1\) factor.

**Proof.** The proof follows from Theorem 3.2.3 and Theorem 3.3.1. \(\square\)

In general the condition \(\frac{|\Delta x_k|}{|\Delta \lambda_k|} > M > 0\) is restrictive. In particular if there exists a subsequence \(\{n_k\}\) of \(k\) such that \(\lim_{k \to \infty} \frac{|\lambda_{n_k} - \lambda_*|}{|x_{n_k} - x_*|} = \infty\), then the condition \(\frac{|\Delta x_k|}{|\Delta \lambda_k|} > M > 0\) cannot hold. Recall that this was the situation with the example (1) studied by Byrd (1985) when we chose the quasi-Newton update for the full Hessian appropriately. We now give a formal proof of this fact.

**Theorem 3.1** If there exists a sequence of positive integer \(\{n_k\}\) such that

\[
\lim_{k \to \infty} \frac{|\lambda_{n_k} - \lambda_*|}{|x_{n_k} - x_*|} = \infty\quad \text{and} \quad \{B_k\}\quad \text{and} \quad \{B_k^{-1}\}\quad \text{are bounded, then} \quad \lim_{k \to \infty} \frac{|\Delta x_{n_k}|}{|\Delta \lambda_{n_k}|} = 0.
\]
Proof. From Lemma 3.1.1 we have

$$x_{n_k+1} - x_{n_k} = -B^{-1}_{n_k} \Gamma_{n_k} (x_{n_k} - x_*) - B^{-1}_{n_k} \nabla h_{n_k} (\lambda_{n_k+1} - \lambda_*).$$

It follows that

$$\frac{x_{n_k+1} - x_{n_k}}{\lambda_{n_k+1} - \lambda_{n_k}} = \frac{-B^{-1}_{n_k} \Gamma_{n_k} (x_{n_k} - x_*)}{\lambda_{n_k+1} - \lambda_{n_k}} - B^{-1}_{n_k} \nabla h_{n_k} \frac{(\lambda_{n_k+1} - \lambda_*)}{\lambda_{n_k+1} - \lambda_{n_k}}. \quad (3.4.4)$$

Now

$$\frac{\lambda_{n_k+1} - \lambda_{n_k}}{x_{n_k} - x_*} = \frac{\lambda_{n_k+1} - \lambda_*}{x_{n_k} - x_*} - \frac{\lambda_{n_k} - \lambda_*}{x_{n_k} - x_*}. \quad (3.4.5)$$

By Theorem 3.1.1 we have $\frac{\lambda_{k+1} - \lambda_*}{x_{n_k} - x_*}$ is bounded and by assumption we have

$$\lim_{k \to \infty} \frac{\lambda_{n_k} - \lambda_*}{x_{n_k} - x_*} = \infty. \text{ Hence from (3.4.5) we have}$$

$$\lim_{k \to \infty} \frac{\lambda_{n_k+1} - \lambda_{n_k}}{x_{n_k} - x_*} = \infty,$$

or

$$\lim_{k \to \infty} \frac{x_{n_k} - x_*}{\lambda_{n_k+1} - \lambda_{n_k}} = 0. \quad (3.4.6)$$

Also the facts that $\lim_{k \to \infty} \frac{\lambda_{n_k} - \lambda_*}{x_{n_k} - x_*} = \infty$ and $\frac{\lambda_{n_k+1} - \lambda_*}{x_{n_k} - x_*}$ is bounded will give

$$\lim_{k \to \infty} \frac{\lambda_{n_k+1} - \lambda_*}{\lambda_{n_k} - \lambda_*} = 0. \text{ Therefore}$$

$$\lim_{k \to \infty} \frac{\lambda_{n_k+1} - \lambda_*}{\lambda_{n_k+1} - \lambda_{n_k}} = \lim_{k \to \infty} \left( \frac{\lambda_{n_k} - \lambda_*}{\lambda_{n_k+1} - \lambda_{n_k}} \cdot \frac{\lambda_{n_k+1} - \lambda_*}{\lambda_{n_k} - \lambda_*} \right)$$

$$= 0. \quad (3.4.7)$$

Combining (3.4.4), (3.4.6), and (3.4.7) with the fact that $\{B_k^{-1}\}$ is bounded we get

$$\lim_{k \to \infty} \frac{\Delta x_{n_k}}{\Delta \lambda_{n_k}} = 0. \quad \square$$
Theorem 3.2 Consider the SQP Broyden, PSB, DFP or BFGS secant method. In the case of DFP and BFGS assume that the matrix $A_*$ is positive definite. Then there exist positive numbers $\epsilon$ and $\delta$ such that whenever $|x_0 - x_*| \leq \epsilon$ and $|B_0 - A_*| \leq \delta$ we have

(i) \{$(x_k, \lambda_k)$\} converges to $(x_*, \lambda_*)$ $q$-superlinearly,

(ii) \{$(x_k)$\} converges to $x_*$ $q$-superlinearly,

(iii) \{$(\lambda_k)$\} converges to $\lambda_*$ $q$-superlinearly or $q$-sublinearly with unbounded $q_1$ factor,

(iv) $\lim_{k \to \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|x_k - x_*|} = 0.$

Proof. The proof of (i), (ii), and (iii) follows from known results, see Fontecilla, Steihaug and Tapia (1987), and Theorem 3.3.1. The proof of (iv) follows from (ii) and Lemma 3.1.1 with the fact that

$$\lim_{k \to \infty} \frac{|(B_k - A_*) \Delta x_k|}{|\Delta x_k|} = 0.$$  $\square$
CHAPTER 4

Summary and Concluding Remarks

In the following we give a brief summary and some concluding remarks. In general if we have $q$-superlinear convergence in the pair $(x, \lambda)$ we only have $r$-superlinear convergence in the variable $x$ and the multiplier $\lambda$. However, we have shown that for an SQP method we always have at least two-step $q$-superlinear convergence for the variable $x$. We also gave a condition that ensured one-step $q$-superlinear convergence for the variable $x$, and the well-known secant updates Broyden, PSB, DFP and BFGS all satisfy this condition. Furthermore the assumption made by Gill, Murray, Saunders and Wright (1986), i.e. $\frac{|\Delta x_k|}{|\Delta \lambda_k|}$ is bounded away from zero, also implies $q$-superlinear convergence in the variable $x$, but it is restrictive. As for the convergence of the multiplier $\lambda$, we showed that it was $q$-superlinear or $q$-sublinear with unbounded $q_1$ factor whenever the convergence for the pair $(x, \lambda)$ and the variable $x$ were $q$-superlinear. We also gave a condition that implies $q$-sublinear convergence for the pair $(x, \lambda), x,$ and $\lambda$. Unfortunately, if this condition holds, the assumption given by Gill, Murray, Saunders and Wright (1986) cannot hold. We do not know how often $q$-sublinear convergence with unbounded $q_1$ factor in the multiplier $\lambda$ will occur. However, we do believe that in most cases the Broyden
PSB, DFP and BFGS secant methods will give $q$-superlinear convergence for the pair $(x, \lambda)$, the variable $x$ and the multiplier $\lambda$; however the possibility of $q$-sublinear convergence in $\lambda$ is real.
Appendix

In this appendix we collect and catalog the various SQP secant methods. For more details see Tapia (1988).

ALGORITHM A.1: SQP Lagrangian Secant Method

By a successive quadratic programming (SQP) Lagrangian secant method for problem (1.1) we mean the iterative process

\[ x_+ = x + s \]  \hspace{1cm} (A.1a)

\[ \lambda_+ = \lambda + \Delta \lambda \]  \hspace{1cm} (A.1b)

\[ B_{lt} = B_t + SE\text{CANT}(s, y_t, B_t, v(s, y_t, B_t)) \]  \hspace{1cm} (A.1c)

with

\[
SE\text{CANT}(s, y, B, v(s, y, B)) = \frac{(y - Bs)v^T + v(y - Bs)^T}{v^T s} - \frac{(y - Bs)^T s v^T}{(v^T s)^2},
\]

\[ y_t = \nabla_x \ell(x_+, \lambda_+) - \nabla_x \ell(x, \lambda_+) \]  \hspace{1cm} (A.2)

and \(s\) and \(\Delta \lambda\) are respectively the solution and the multiplier associated with the solution of the quadratic program

\[
\text{minimize} \quad \nabla_x \ell(x, \lambda)^T s + \frac{1}{2} s^T B_{lt} s
\]

\[
\text{subject to} \quad \nabla h(x)^T s + h(x) = 0.
\]  \hspace{1cm} (A.3)
Notice that $B_t^+ s$ will satisfy the Lagrangian secant equation

$$B_t^+ s = y_t .$$  \hspace{1cm} (A.4)

**ALGORITHM A.2: SQP Augmented Lagrangian Secant Method**

By a successive quadratic programming (SQP) augmented Lagrangian secant method for problem (1.1) we mean the iterative process

$$x_+ = x + s$$  \hspace{1cm} (A.5a)

$$\lambda_+ = \lambda + \Delta \lambda$$  \hspace{1cm} (A.5b)

$$B_t^+ = B_t + \text{SECANT} (s, y_L, B_L, v(s, y_L, B_L))$$  \hspace{1cm} (A.5c)

with $\text{SECANT} (s, y, B, v(s, y, B))$ given by (A.1d),

$$y_L = \nabla_z L(x_+, \lambda_+) - \nabla_z L(x, \lambda_+),$$  \hspace{1cm} (A.6)

$s$ and $\Delta \lambda$ are respectively the solution and the multiplier associated with the solution of the quadratic program

$$\begin{align*}
\text{minimize} & \quad \nabla_z L(x, \lambda)^T s + \frac{1}{2} s^T B_L s \\
\text{subject to} & \quad \nabla h(x)^T s + h(x) = 0 .
\end{align*}$$  \hspace{1cm} (A.7)

Notice that $B_L^+ s$ will satisfy the augmented Lagrangian secant equation

$$B_L^+ s = y_L .$$  \hspace{1cm} (A.8)
ALGORITHM A.3: SQP Augmented Scale Lagrangian Secant Method

By an SQP augmented scale Lagrangian secant method we mean an SQP Lagrangian secant method, i.e., Algorithm A.1, where the choice for $B_t^+$ in (A.1c) is

$$B_t^+ = B_t + \text{SECANT}(s, y_t, B_t, v_L) \quad (A.9)$$

with $\text{SECANT}(s, y, B, v)$ given by (A.1d) and for a given choice of scale $v = v(s, y, B)$ the augmented scale $v_L$ is defined to be

$$v_L = v(s, y_L^t, B_L) \quad (A.10)$$

where

$$y_L^t = y_t + \rho \nabla h(x_+)\nabla h(x_+)^T s \quad (A.11)$$

and

$$B_L^t = B_t + \rho \nabla h(x_+)\nabla h(x_+)^T \quad (A.12)$$

for some $\rho \geq 0$.

ALGORITHM A.4: SQP Structured Augmented Lagrangian Secant Method

By an SQP structured augmented Lagrangian secant method we mean an algorithm of the form of Algorithm A2, where the choice for $B_L^+$ in (A.5c) is

$$B_L^+ = B_L + \text{SECANT}(s, y_L^t, B_L, v(s, y_L^t, B_L)) \quad (A.13)$$
with \( \text{SECANT}(s, y, B, v(s, y, B)) \) given by (A.1d) and \( y^*_L \) given by (A.11), i.e.,

\[
y^*_L = \nabla_x \ell(x_+, \lambda_+) - \nabla_x \ell(x, \lambda_+) + \rho \nabla h(x_+) \nabla h(x_+)^T s.
\]  

(A.14)

In this case \( B^+_L \) satisfies the structured augmented Lagrangian secant equation

\[
B^+_L s = y^*_L
\]  

(A.15)

which is not the same as the augmented Lagrangian secant equation (A.8) due to the fact that \( y^*_L \) in (A.14) employs structure and differs from \( y_L \) in (A.8).
BIBLIOGRAPHY


BYRD, R.H. [1984]. On the convergence of constrained optimization methods with accurate Hessian information on a subspace. TR, CU-CS-270-84, Department of Computer Science, University of Colorado, Boulder, Co.


43


GILL, P.E., MURRAY, W., SAUNDERS, M.A. and WRIGHT, M.H. [1986]. Some theoretical properties of an augmented Lagrangian merit function. TR SOL 86-6, Department of Operations Research, Stanford University, Stanford, CA.


