UNIVERSITI PUTRA MALAYSIA

ON THE MODIFICATIONS OF A BROYDEN'S SINGLE PARAMETER RANK-TWO QUASI-NEWTON METHOD FOR UNCONSTRAINED MINIMIZATION

LEONG WAH JUNE

FSAS 1999 7

ON THE MODIFICATIONS OF A BROYDEN'S SINGLE PARAMETER RANK-TWO QUASI-NEWTON METHOD FOR UNCONSTRAINED MINIMIZATION

By

LEONG WAH JUNE

Thesis Submitted in Fulfilment of the Requirements for the Degree of Master of Science in the Faculty of Science and Environmental Studies
Universiti Putra Malaysia

May 1999

ACKNOWLEDGEMENTS

I would like to express my most gratitude and sincere appreciation to my chairman, Associate Professor Dr. Malik Hj. Abu Hassan and Dr. Mansor b. Monsi for their untiring guidance, valuable advice, support and comments. Their patience and persistent encouragement during the course of my research is instrumental to the completion of this thesis. I am also grateful to Dr. Leow Soo Kar for serving in the supervisory committee.

Special thanks is given to the Head of Department, Associate Professor Dr. Isa b. Daud, academic and general staff of the Department of Mathematics, Universiti Putra Malaysia, for their assistance in various capacities.

Last but not least, I would like to thank my family and friends for their understanding support and encouragement throughout the course of this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

1. $\Re^n$ denotes lineared n dimensional Real space.

2. g is the n×1 gradient vector of $f(x)$, that is

$$g_i = \frac{\partial f(x)}{\partial x_i}, \quad i = 1, 2, ..., n.$$

3. G is the n×n Hessian matrix of $f(x)$, that is the (i,j)th element of G, $G_{i,j}$ is given by

$$G_{i,j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad \begin{array}{l} i = 1, 2, ..., n. \\ j = 1, 2, ..., n. \end{array}$$

4. $B^{(k)}$ is an n×n matrix that is a kth approximation to G.

5. $H^{(k)}$ is an n×n matrix that is a kth approximation to G^{-1}.

6. $x^{(k)}$ is the kth approximation to x^*, a minimum of $f(x)$.

7. $f^{(k)} = f(x^{(k)})$.

8. $g^{(k)}$ is the gradient vector of $f(x)$ at $x^{(k)}$.

9. A superfix ^{T} on a matrix or vector denotes transpose.

10. C^r denotes the class of functions whose rth deriative is continuous.

11. $\|y\|$ denotes an arbitrary norm of y.

12. min denotes the minimum.

13. max denotes the maximum.

13. det (A) means determinant of A.

14. Tr (A) denotes the trace of A.

Abstract of thesis submitted to the Senate of Universiti Putra Malaysia
in fulfilment of the requirements for the degree of Master of Science.

ON THE MODIFICATIONS OF A BROYDEN'S SINGLE PARAMETER RANK-TWO QUASI-NEWTON METHOD FOR UNCONSTRAINED MINIMIZATION

By

LEONG WAH JUNE

May 1999

Chairman : Associate Professor Malik Hj. b. Abu Hassan, Ph.D.

Faculty : Science and Environmental Studies

The thesis concerns mainly in finding the numerical solution of non-linear unconstrained problems. We consider a well-known class of optimization methods called the quasi-Newton methods, or variable metric methods. In particular, a class of quasi-Newton method named Broyden's single parameter rank two method is focussed.

We also investigate the global convergence properties for some step-length procedures. Immediately from the investigations, a global convergence proof of the Armijo quasi-Newton method is given.

Some preliminary modifications and numerical experiments are carried out to gain useful numerical experiences for the improvements of the quasi-Newton updates.

We then derived two improvement techniques : the first we employ a switching criteria between quasi-Newton *Broyden-Fletcher-Goldfrab-Shanno* or BFGS and steepest descent direction and in the second we introduce a reduced trace-norm condition BFGS update. The thesis includes results illustrating the numerical performance of the modified methods on a chosen set of test problems.

Limitations and some possible extensions are also given to conclude this thesis.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk ijazah Master Sains.

PENGUBAISUAIAN KE ATAS KAEDAH KUASI-NEWTON PARAMETER TUNGGAL BERPANGKAT DUA BROYDEN UNTUK PEMINIMUMAN TAK BERKEKANGAN

Oleh

LEONG WAH JUNE

Mei 1999

Pengerusi : Profesor Madya Malik Hj. b. Abu Hassan, Ph.D.

Fakulti : Sains dan Pengajian Alam Sekitar

Penumpuan utama tesis ini adalah dalam usaha mencari penyelesaian berangka untuk masalah pengoptimuman tak linear tak berkekangan. Kami mempertimbangkan suatu kelas kaedah pengoptimuman yang terkenal disebut kaedah kuasi-Newton atau keadah metrik berubah. Khususnya, satu kelas kaedah kuasi-Newton bernama kaedah parameter tunggal berpangkat dua Broyden akan difokuskan.

Kami juga menyiasat sifat penumpuan sejagat untuk beberapa prosedur mencari panjang langkah. Sejurus dari penyiasatan ini, satu pembuktian penumpuan sejagat untuk kaedah kuasi-Newton Armijo diberi.

Beberapa pengubahsuaian pendahuluan dan ujikaji-ujikaji berangka dilaksanakan untuk memperolehi pengalaman berangka yang berguna dalam usaha

memperbaiki rumus kemaskini kuasi-Newton. Kami seterusnya menerbitkan dua teknik pembaikan: yang pertama kami memggunakan satu kriteria penukaran di antara arah kuasi-Newton *Broyden-Fletcher-Goldfrab-Shanno* atau BFGS dan penurunan tercuram dan untuk yang kedua, kami memperkenalkan satu rumus kemaskini BFGS bersyarat terturun dalam norma surihan. Tesis ini akan meliputi keputusan-keputusan untuk menghuraikan persembahan berangka untuk kaedah-kaedah terubahsuai ke atas suatu set masalah ujian yang dipilih.

Kekurangan-kekurangan dan beberapa kemungkinan kajian lanjutan akan diberi untuk mengakhiri tesis ini.

CHAPTER I

INTRODUCTION

Organization of the Studies

In this thesis we are concerned mainly in finding the numerical solution of *non-linear optimization* problems We begin in this chapter by giving a brief introduction to the definition of non-linear optimization Chapter I also describes some properties of real-valued functions which are most frequently used in the numerical solution of optimization problems and in particular in the minimization of real-valued functions

In Chapter II, we first consider a well-known class of optimization methods called the *quasi-Newton* methods, or variable metric methods Theory, conditions and motivation of this method are also presented In particular, a class of quasi-Newton methods named Broyden`s (1965) single parameter rank-two method is reviewed in detail A description of so-called line search procedures and its algorithms is included Finally, the theory of convergence for quasi-Newton updates is discussed

In Chapter III, we extend our analysis to the line search procedures as described briefly in Chapter II. We begin by discussing three commonly used classic inexact line search procedures namely Armijo (1966)'s, Goldstein (1965)'s and Wolfe (1969, 1971)'s line search. Chapter III ends with a global convergence proof derived for a bounded condition quasi-Newton updates using the classic Armijo's line search.

In Chapter IV, we perform a series of numerical experiments and modification to the quasi-Newton update. Numerical results corresponding to the experiments and its modification for three test examples are given to show the effect of such efforts.

Chapters V and VI contain a description of three modified quasi-Newton algorithms for unconstrained minimization in which we use numerical experiences gained in Chapter IV to improve the available quasi-Newton updates. Numerical results are given to show the efficiency of the new algorithms. Finally, we give the convergence proof for each modified method.

The thesis closes with a summary of the achievements of the previous chapters. Limitations and possible extensions of the work are also considered.

Problem Review

Non-linear optimization is concerned with methods for locating the least value (the *minimum*) or the greatest value (the *maximum*) of a non-linear function of any number of independent variables, referred to as the *objective function*. The least value problem is called *minimization* and the greatest value problem *maximization*. Any maximization problem can be converted into a minimization problem by multiplying the objective function by a factor of -1. It is therefore not necessary to consider these two aspects of the problem separately. We adopt the more usual convention of discussing the entire subject in terms pertaining to minimization. When the problem is to be solved subject to no special conditions upon the values that the independent variables are allowed to assume, it is said to be a problem of *unconstrained optimization*.

Basic Definitions and Theorems

The aim of this section is to present the properties of real-value functions that are most frequently used in the numerical solution of optimization problems and in particular in minimization of a real-valued function, which will be the objective function. In the following, we shall use the standard notations $\|x\|$, $\|x-y\|$, $S(x,\varepsilon)$, $S[x,\varepsilon]$ to denote the Euclidean norm of $x\in\Re^n$, the Euclidean distance of a pair x, $y\in\Re^n$, the open and closed sphere with center $x\in\Re^n$ and radius $\varepsilon>0$ respectively. We are essentially interested in the two following problems:

Problem 1.1. Global Minimization

Let $X\subseteq\Re^n$ be a closed set and let $f: X\rightarrow\Re$. Find a point $x^*\in X$ such that

$$f(x^*) \leq f(x)$$

for all $x\in X$.

Problem 1.2. Local Minimization

Let $X \subseteq \Re^n$ be a closed set and let $f: X \to \Re$. Find a point $x^* \in X$ and a real number $\varepsilon > 0$ such that

$$f(x^*) \leq f(x)$$

for all $x \in S(x, \varepsilon) \cap X$.

Definition 1.1. Each point $x^* \in \Re^n$ which is either a solution of Problem 1.1 or of Problem 1.2 will be called global or local minimizer, respectively, and the corresponding function value $f(x^*)$, global or local minimum. A minimizer $x^* \in \Re^n$ will be called isolated if there do not exist any other minimizers in a sufficiently small neighbourhood of x^*.

Remark 1.1. While the global minimizer of $f(x^*)$ may not be unique, there exists only one global (or absolute) minimum function value $f(x^*)$.

Remark 1.2. As a particular case of both Problem 1.1 and Problem 1.2, we consider only the case $X = \Re^n$ in which the so-called *unconstrained minimization* of $f(x)$. Therefore in the rest of the chapter, we consider only $X = \Re^n$.

The necessary and sufficient conditions for the existence and uniqueness of solutions of the minimization Problems 1.1 and 1.2 are given by the properties of the function $f(x)$. For further development, it is worthwhile to define the largest set of points for which property in Problem 1.2 holds with respect to isolated minimum x^*, the so-called region of attraction of x^*.

Definition 1.2. Consider the isolated local minimum $x^* \in X$ of the real-valued function $f(x), f: X \rightarrow \mathfrak{R}$, then the largest open connected neighbourhood of x^*, $M(x^*)$ for which

$$f(x^*) \leq f(x) \quad \text{for all } x \in M(x^*) \subset \mathfrak{R}^n$$

is called the region of descent to x^*.

Definition 1.3. For each isolated local minimum $x^* \in \mathfrak{R}^n$ of $f(x)$ and each real number k>0, consider the connected component $M_k'(x^*)$ containing x^* of the set

$$f(x^*) \leq f(x) < f(x^*)+k$$

where,

$$M'(x^*) \subset M(x^*).$$

Consider next the open connected set $N(x^*) \subset \mathfrak{R}^n$ defined as

$$N(x^*) = \bigcup_{k>0} M_k'(x^*);$$

the set $N(x^*)$ is called the region of attraction of x^* and identifies all points which are associated with x^*.

It is also important to discuss a few properties of $f(x)$ namely continuity, Lipschitz continuity and differentiability.

Definition 1.4. A function $f(x)$, $f: X \rightarrow \Re$ is continuous in $x \in X$ if $S(x,\delta) \cap X \neq \{x\}$ for all $\delta > 0$ and in addition for all $\varepsilon > 0$ there exists a $\delta > 0$ such that $y \in S(x,\delta)$ implies $|f(y)-f(x)| < \varepsilon$. If $f(x)$ is continuous for all points $x \in X$, it will be said to be continuous in X.

Remark 1.3. If $\alpha \in \Re$ and $f(x)$, $f: X \rightarrow \Re$ be continuous in $\Re^n$. Then the set

$$N(\alpha) = \{x \in \Re^n : f(x) < \alpha\} \quad (1.1)$$

is open; if $N(\alpha) \cap X$ is not empty and bounded for some $\alpha \in \Re$, there exists then at least one solution of the global minimization Problem 1.1 which is contained in $N(\alpha) \cap X$. The set $N(\alpha)$ will be called level set of the function $f(x)$.

If the set $N(\alpha)$ is not connected i.e. it can be partitioned into union of disjoint open sets, called components, then each component of $N(\alpha)$ with a non-empty and bounded intersection with X contains at least one solution of the local minimization Problem 1.2.

This result allows us to establish the existence and the minimum number of solutions of any optimization problem. The key condition for that is the existence of bounded components of $N(\alpha)$.

It may frequently occur that continuity of the objective function above does not ensure a positive outcome of the numerical approximation methods for the solution of optimization problems. The stronger condition which may be required in this case is that the objective function $f(x)$ be Lipschitzian.

Definition 1.5. A function $f(x)$, $f: X \rightarrow \Re$ is said to be globally Lipschitzian on X if there exists a real number L>0 called the Lipschitz constant, such that

$$|f(x)-f(y)| \leq L\|x-y\| \quad \text{for all } x, y \in X.$$

The function $f(x)$ is said to be locally Lipschitzian in X if each $x \in X$ there exists a sphere $S(x,\varepsilon)$ and a real number $L(x)>0$ such that

$$|f(x)-f(y)| \leq L\|x-y\| \quad \text{for all } x,y \in S(x,\varepsilon) \cap X.$$

Remark 1.4. The knowledge of the Lipschitz constant L of a Lipschitzian function on a compact set $\Omega \subset \Re^n$ is useful in some global minimization problems. Indeed when the objective function $f(x)$ is Lipschitzian on the compact set $\Omega \subset \Re^n$ and its Lipschitz constant L is known, one can apply some particular methods which are based upon the construction of a piecewise linear approximation from below $f_1(x)$ of $f(x)$, with

$$f_1(x) \leq f(x) \quad \text{for all } x \in \Omega.$$

Where the approximating function $f_1(x)$ is constructed via the sequential evaluation of $f(x)$ on all points of a reticle while the knowledge of L allows the identification of the reticle.

It is also necessary to discuss the existence of derivatives, differentiability, twice differentiability, continuous differentiability, twice continuous differentiability and convexity of a function $f(x)$ and its relationships between critical (or stationary) points.

Definition 1.6. Given a function $f(x)$, $f: X \rightarrow \Re$, we say that at the point x there exist the n first partial derivatives of f if all the n limits

$$\lim_{h \rightarrow 0} \frac{(f(x+he_i)-f(x))}{h}, \quad i = 1, 2, \ldots, n \qquad (1.2)$$

exist and they are finite. In (1.2) we denote by e_i, as the vector of the ith coordinate axis. If at the point $x \in X$, there exist the n first partial derivatives of $f(x)$ the n-vector $g(x)$ defined component-wise through the n relationships

$$\nabla f(x) = g_i(x) = \frac{\partial f(x)}{\partial x_i}, \quad i = 1, 2, \ldots, n.$$

is called the gradient of f.

Definition 1.7. A function $f(x)$, $f: X \rightarrow \Re$ is said to be differentiable at the point $x \in X$ if at the point x all n first partial derivatives of f exist and in addition

$$\lim_{y \rightarrow x} \frac{(f(y)-f(x)-(y-x)^T g(x))}{\|y-x\|} = 0.$$

Definition 1.8. A function $f(x)$, $f: X \rightarrow \Re$ is said to be continuously differentiable at the point $x \in X$ if all n first partial derivatives exist in a neighborhood of x and they are continuous in x.

Definition 1.9. If $f(x)$ is continuously differentiable in X, it is said to belong to function class C^1.

Remark 1.5. The following three properties that we have presented are connected by the relationships:

Continuous differentiability → differentiability → local Lipschitz → continuity.

The inverse relationships do not hold in general.

If a real-valued function is continuously differentiable in the neighbourhood of a minimum, it has a remarkably regular behavior in this neighbourhood. This regularity allows the following strong characterization of minima.

Definition 1.10. Let f: $X \rightarrow \Re$ be continuously differentiable in all interior points of X. The points $x^* \in X$ such that

$$g(x^*) \qquad (1.3)$$

are called critical (or stationary)

Remark 1.6. If $x^* \in \Re^n$ is a local minimum of a real valued function $f(x)$, f: $\Re^n \rightarrow \Re$ which is continuously differentiable, then x^* is a critical point of $f(x)$, i.e. (1.3)

Thus critical points play a major role in the unconstrained minimization problems. Indeed not only the topological structure of the function in the neighbourhood of the minima is very regular and help their identification, but the vast majority of local minimization algorithms is based upon the search for the critical points of $f(x)$, i.e. on the solution of the system of n algebraic equations in n unknowns

$$g(x) = 0$$

which identifies all critical points of $f(x)$.

Consider the differential equation

$$x = g(x) \tag{1.4}$$

when the real-valued function $f(x)$ is continuously differentiable provides a very useful tool to characterize the properties of the critical points of $f(x)$. Before discussing the additional conditions that must be imposed on $f(x)$ in order to be able to use equation (1.4), it is important to point out that the critical points of $f(x)$ coincide with the equilibrium points of (1.4), i.e. the points x^* such that

$$x = g(x^*) = 0. \tag{1.5}$$

In addition, if we consider the solutions $x = x(x^0,t)$ of equation (1.5), we see that the critical points x* which are minima of $f(x)$ are also asymptotically stable equilibrium points of (1.5) (i.e. stable and such that $\lim_{t\to+\infty} x(x^0,t)\to x^*$ for all $x^0 \in S(x^*,\varepsilon)$, $\varepsilon>0$, while critical points which are saddle points of $f(x)$ are also equilibrium points which are saddle points of (1.5).

The last problem that we want to recall is that of the region of attraction of a minimum and its boundary. The region of attraction of a (isolated) local minimum was defined in Definition 1.2 when the function $f(x)$ is continuously differentiable. The region of attraction can be more conveniently defined in another way which is related to the properties of the differential equation (1.4) and explains the name given to this set.

Definition 1.11. If $x^* \in \Re^n$ is an isolated local minimum of the continuously differentiable real-valued function $f(x)$, $f\colon \Re^n \to \Re$, consider the ordinary differential equation (1.4), then the set of all points x^* such that