UNIVERSITI PUTRA MALAYSIA

MULTIPLE ALTERNATE STEPS GRADIENT METHODS FOR UNCONSTRAINED OPTIMIZATION

LEE SUI FONG

IPM 2009 11

MULTIPLE ALTERNATE STEPS GRADIENT METHODS FOR UNCONSTRAINED OPTIMIZATION

By

LEE SUI FONG

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Science

September 2009

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

MULTIPLE ALTERNATE STEPS GRADIENT METHODS FOR UNCONSTRAINED OPTIMIZATION

By

LEE SUI FONG

September 2009

Chairman: Dr. Leong Wah June, PhD

Faculty: Science

The focus of this thesis is on finding the unconstrained minimizer of a function by using the alternate steps gradient methods. Specifically, we will focus on the well-known classes of gradient methods called the steepest descent (SD) method and Barzilai-Borwein (BB) method. First we briefly give some mathematical background on unconstrained optimization as well as the gradient methods. Then we discuss the SD and BB methods, the fundamental gradient methods which are used in the gradient method alternately to solve the problems of optimization. Some general and local convergence analyses of SD and BB methods are given, as well as the related so-called line search method.

A review on the alternate step (AS) gradient method with brief numerical results and convergence analyses are also presented.

The main practical deficiency of SD method is the directions generated along the line tend to two different directions, which causes the SD method performs poorly and requires more computational work. Though BB method does not guarantee a descent in the objective function at each iteration due to it non-monotone behavior, it performs better than SD method in this case. Motivated by these limitations, we introduce a new gradient method for improving the SD and BB method namely the Multiple Alternate Steps (MAS) gradient methods. The convergence of MAS method is investigated. Analysis on the behavior of MAS method is also performed. Furthermore, we also presented the numerical results on quadratics test problems in order to compare the numerical performance of MAS method with SD, BB and AS methods.

The purpose of this research is to study a working knowledge of optimization theory and methods. We hope that the new MAS gradient method can give significant research contribution in our daily life application. For example, in maximizing the profit of a manufacturing operation or improving a system in certain ways to reduce the effective runtime in computer science.

Finally we comment on some achievements in our researches. 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

KAEDAH KECERUNAN SELANG-SELI BERBILANG LANGKAH UNTUK PENGOPTIMUMAN TAK BERKEKANGAN

Oleh

LEE SUI FONG

September 2009

Pengerusi: Dr. Leong Wah June, PhD

Fakulti: Sains

Tumpuan tesis ini adalah mencari peminimum tak berkekangan bagi suatu fungsi dengan menggunakan kaedah kecerunan selang-seli langkah. Khususnya, kami akan menumpu kepada suatu kelas kaedah kecerunan terkenal yang dipanggil kaedah penurunan tercuram (SD) dan kaedah Barzilai dan Borwein (BB). Pertama, kami memberi secara ringkas tentang latarbelakang matematik dalam pengoptimuman tak berkekangan dan juga kaedah kecerunan. Kemudian kami membincangkan kaedah SD dan kaedah BB, iaitu kaedah kecerunan asas yang digunakan dalam kaedah kecerunan secara selang-seli bagi menangani masalah pengoptimuman tak berkekangan. Analisis tentang penumpuan tempatan secara am terhadap SD dan BB diberikan, bersama dengan sesuatu

yang berkait dengan kaedah gelintaran garis. Satu sorotan bagi kaedah kecerunan selang-seli langkah (AS) dengan keputusan berangka yang ringkas dan analisis penumpuan juga dibentangkan.

Kekurangan utama secara praktik kaedah SD adalah arah yang ditujukan sepanjang garis cenderung kepada dua arah yang berlainan, di mana menyebabkan kaedah penurunan tercuram menunjukkan prestasi yang lemah dan memerlukan kerja komputasi yang berlebihan. Di samping itu, kaedah BB pula tidak dapat menjamin penurunan dalam fungsi objektif bagi setiap lelaran yang disebabkan oleh sifat penurunan yang tidak seragam, akan tetapi kaedah BB menunjukkan prestasi yang lebih baik berbanding dengan kaedah SD dalam kes ini. Dengan kelemahan-kelemahan itu, kami memperkenalkan satu kaedah kecerunan yang dinamakan sebagai kaedah kecerunan selang-seli berbilang langkah (MAS) bagi memperbaiki kaedah SD dan BB. Penumpuan kaedah MAS akan diselidikkan. Analisis tentang kelakuan kaedah MAS juga akan dilaksanakan. Di samping itu, kami juga membentangkan keputusan berangka tentang peminimuman masalah kuadratik untuk tujuan perbandingan prestasi berangka kaedah MAS terhadap kaedah SD, BB dan AS.

Tujuan kajian ini adalah untuk mengkaji teori dan kaedah pengoptimuman. Kita berharap dengan kaedah MAS yang baru ini dapat memberi sumbangan bermakna dalam aplikasi aktiviti harian kita. Contohnya, maximumkan penguntungan dalam operasi pengeluaran ataupun memperbaiki sesuatu system dalam sains computer untuk memendikkan masa efektif.

Akhirnya kami memberi komen tentang beberapa pencapaian dalam penyelidikan kami. Kemungkinan lanjutan juga diberi untuk mengakhiri tesis ini.

ACKNOWLEDGEMENTS

I am grateful to several people for their help during the course of this research. In particular, I wish to express my infinite gratitude and sincere appreciation to my chairman, Dr. Leong Wah June for his valuable comments, support, advice, suggestions. I am also grateful to Professor Dr. Malik B. Hj. Abu Hassan and Dr. Mansor B. Monsi for serving in the supervisory committee. Their patience and persistent encouragement during the course of my research is instrumental to the completion of this thesis

Special thanks is given to the Head of Department and general staffs of the Institute For Mathematical Research, University Putra Malaysia, for their assistance in various capacities. I am also acknowledged the financial support given to me by University Putra Malaysia under the Graduate Research Fellowship. I also thank to miss Mahboubeh Farid for her untiring guidance in writing MATLAB code.

Finally, I would like to thank my family and friends for their support and encouragement during the course of this study.

I certify that an Examination Committee has met on **data of viva voce** to conduct the final examination of **Lee Sui Fong** on her **Master of Science** thesis entitled **"Multiple Alternate Steps Gradient Methods For Unconstrained Optimization'** in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the student be awarded the (Name of relevant degree).

Members of the Examination Committee were as follows:

Name of Chairperson, PhD
Title (e.g. Professor/Associate Professor/Ir) – omit if not relevant
Name of Faculty
Universiti Putra Malaysia
(Chairman)

Name of Examiner 1, PhD
Title (e.g. Professor/Associate Professor/Ir) – omit if irrelevant
Name of Faculty
Universiti Putra Malaysia
(Internal Examiner)

Name of Examiner 2, PhD
Title (e.g. Professor/Associate Professor/Ir) – omit if irrelevant
Name of Faculty
Universiti Putra Malaysia
(Internal Examiner)

Name of External Examiner, PhD
Title (e.g. Professor/Associate Professor/Ir) – omit if irrelevant
Name of Department and / or Faculty
Name of Organisation (University/ Institute)
Country
(External Examiner)

HASANAH MOHD. GHAZALI, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

LEONG WAH JUNE, PhD
Faculty of Science
Universiti Putra Malaysia
(Chairman)

MALIK B. HJ. ABU HASSAN, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

MANSOR B. MONSI, PhD
Faculty of Science
Universiti Putra Malaysia
(member)

HASANAH MOHD GHAZALI, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 14th January 2010

DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

LEE SUI FONG

Date:

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF NOTATIONS

1. R^n denotes the linear n- dimensional real space.

2. g is the $n \times 1$ gradient vector of $f(\bar{x})$, that is

$$g^{(i)} = \frac{\partial f(\bar{x})}{\partial x^{(i)}}, i = 1,2, \dots, n.$$

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

$$G^{(i,j)} = \frac{\partial^2 f(\bar{x})}{\partial x^{(i)} \partial x^{(j)}}, i = 1,2, \dots, n \;\; j = 1,2, \dots, n.$$

4. x_k is the kth approximation to x^*, a minimum of $f(\bar{x})$.

5. g_k is the gradient vector of $f(\bar{x})$ at x_k.

6. H_k is an $n \times n$ matrix that is a kth approximation to G^{-1}.

7. A superfix ^{T} on a matrix or vector denotes tanspose.

8. $\|g\|$ denotes an arbitrary norm of g.

9. min denotes the minimum.

CHAPTER I

INTRODUCTION

1.1 Preliminaries

Optimization is a very active research area. It is focus on problems involving decision making to make the ‘best’ choice. Optimization problems are widely spread among the area of engineering, decision sciences and operation research. These applications include structural optimization, digital processing, engineering design, database design, and processing, chemical process control and mechanical engineering. Better design always results in lower implementation and more effective operation under a variety of operating conditions. Optimal solution often brings significant economical and social impact. Over the past few decades, there has been rapid progress in the development of optimization method and software. Along with the tremendous development of high-performance computers technology and computational method, more and more large-scale optimization have been studied and solved.

1.2 Minimization Problem

The general form of minimization problem is defined as follows:

Given a set D and a function $f: D \to P$, find at least one point $x^* \in D$ that satisfies $f(x^*) \leq f(x)$ for all $x \in D$, or show the non-existence of such a point.

The minimization problem can be expressed mathematically as follows:

$$\text{Minimize } f(x), \tag{1.1}$$

$$\text{subject to } x \in D$$

In this formulation, the function f is called the objective function of the problem, $x \in D$ is a decision variable where $x = (x_1, x_2, \ldots, x_n)^T$ is an n-dimensional vector of unknowns, and D is the feasible domain of x specified by constraints. So, the above minimization problem is a general form of a constrained optimization problem, where the decision variables are constrained to be in the constraint set D. In this case, it involves finding the best solution, x of the decision variables over all possible vectors in D, which is the smallest value of the objective function. Specially, the minimization problem (1.1) is called unconstrained minimization if the constraint set $D = R^n$.

Definition 1.1 A vector $x^* \in D$ is called a global minimizer of f over D if it satisfies $f(x^*) \leq f(x)$ for all $x \in D$. The corresponding value of f is called a global minimum.

Definition 1.2 A vector $x^* \in D$ is called a local minimizer of f over D if it satisfies $f(x^*) \leq f(x)$ for all $x \in D$ closed to x^*. The corresponding value of f is called a local minimum.

Because of the theory, minimizing $-f$ is equivalent to maximizing f, maximization problems can be transformed into minimization problem shown in (1.1). We use optimization and minimization interchangeably in this thesis.

The optimization problems can be classified into constrained optimization and unconstrained optimization based on the presence of constraints. This thesis is limited to unconstrained optimization problems, in which we assume that $f(x)$ is continuous differentiable, the variables can take any real values, and a local minimizer provides a satisfactory solution. This probably reflects the available of optimization software as well as the needs of practical applications of optimization problems.

A number of books give substantial amount of material concerning unconstrained optimization and are recommended to readers who desire additional information on this topic. These include Fletcher (1980), Gill et al. (1981), Dennis and Schnabel (1983), Nocedal and Wright (1999) and Ortega and Rheinboldt (1970).

1.3 Existence and Uniqueness of Solutions

Let $f(x)$ be a differentiable function and $g(x)$ denotes the n component gradient column vector of the first partial derivatives of $f(x)$; where

$$[g(x)]_j = \frac{\delta f(x)}{\delta x_j}, j = 1, \dots n \qquad (1.2)$$

Next, let $G(x)$ denotes the $n \times n$ Hessian matrix of second partial derivatives of $f(x)$; for which

$$[G(x)]_{ji} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, i = 1, .. n, j = 1, \dots n. \qquad (1.3)$$

$G(x)$ is also called the Hessian matrix of $f(x)$. If $f(x)$ is twice continuously differentiable, then $G(x)$ is symmetric for all x.

Definition 1.3 A function f of the n-vector x is said to be Lipschitz continuous with constant γ in an open neighbourhood $D \subset R^n$, written $f \in Lip_\gamma(D)$, if there exists a real constant $\gamma \geq 0$ such that for all $x, y \in D$,

$$\|f(x) - f(y)\|_Y \leq \gamma \|x - y\|_X \qquad (1.4)$$

where $\|.\|$ is an approximate norm.

Taylor's formula is the basis for many numerical methods and models for optimization. Most methods for optimizing nonlinear differentiable functions of continuous variables rely heavily upon Taylor series expansions of these functions. We will briefly review the Taylor series expansions used in unconstrained optimization and a few mathematical properties of these expansions.

Theorem 1.1 Taylor's theorem.

If a function $f: R \rightarrow R$ is n times continuously differentiable on an interval [a,b], then

$$f(b) = f(a) + \frac{h}{1!}f'(a) + \frac{h^2}{2!}f''(a) + \cdots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(a) + R_n(f),$$

where $f^{(i)}$ is the ith derivative of f, $h = b - a$, and

$$R_n(f) = \frac{h^n(1-\theta)^{n-1}}{(n-1)!}f^{(n)}(a + \theta h)$$

$$= \frac{h^n}{n!}f^{(n)}(a + \theta' h),$$

with $\theta, \theta' \epsilon\ (0,1)$.

Proof. We have

$$R_n(f) = f(b) - f(a) - \frac{h}{1!}f'(a) - \frac{h^2}{2!}f''(a) - \cdots - \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(a).$$

Denoting by $g_n(x)$ an auxiliary function obtained from $R_n(f)$ by replacing a by x. Hence,

$$g_n(x) = f(b) - f(x) - \frac{b-x}{1!}f'(x) - \frac{(b-x)^2}{2!}f''(x) -$$

$$\cdots - \frac{(b-x)^{n-1}}{(n-1)!}f^{(n-1)}(x).$$

Differentiating $g_n(x)$ yields

$$g_n'(x) = -f'(x) + \left[f^{(1)}(x) - \frac{b-x}{1!}f''(x)\right]$$

$$+ \left[2\frac{b-x}{2!}f''(x) - \frac{(b-x)^2}{2!}f'''(x)\right] + \cdots$$

$$+ \left[(n-1)\frac{(b-x)^{n-1}}{(n-1)!}f^{(n-1)}(x) - \frac{(b-x)^{n-1}}{(n-1)!}f^{(n)}(x)\right]$$

$$= -\frac{(b-x)^{n-1}}{(n-1)!}f^{(n)}(x).$$

Observe that $g_n(x)$ is continuous and differentiable on $[a, b]$ with $g_n(b) = 0$ and $g_n(a) = R_n$.

Applying the mean-value theorem yields

$$\frac{g_n(b) - g_n(a)}{b - a}$$

$$= g_n'(a + \theta h),$$

where $\theta \in (0,1)$.

The above equation is equivalent to

$$\frac{-R_n(f)}{h} = -\frac{(b - a - \theta h)^{n-1}}{(n-1)!} f^{(n)}(a + \theta h)$$

$$= -\frac{h^{n-1}(1-\theta)^{n-1}}{(n-1)!} f^{(n)}(a + \theta h)$$

hence,

$$R_n(f) = \frac{h^n(1-\theta)^{n-1}}{(n-1)!} f^{(n)}(a + \theta h).$$

Applying the Cauchy theorem yields

$$\frac{g_n(b) - g_n(a)}{u_n(b) - u_n(a)} = \frac{g_n'(a + \theta' h)}{u_n'(a + \theta' h)},$$

where $u_n(x) = (b - x)^n$.