UNIVERSITI PUTRA MALAYSIA

NEW QUASI-NEWTON EQUATION AND METHOD VIA HIGHER ORDER TENSOR MODELS

FAHIMEH BIGLARI GHOLILOU

FS 2010 20

NEW QUASI-NEWTON EQUATION AND METHOD
VIA
HIGHER ORDER TENSOR MODELS

By

FAHIMEH BIGLARI GHOLILOU

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

June 2010

DEDICATION

To

Hazrat Zahra salamollah alayha

My husband and my son

Golamreza and Moeinoddin

My Father

and

My Dear Teachers

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

NEW QUASI-NEWTON EQUATION AND METHOD VIA HIGHER ORDER TENSOR MODELS

By

FAHIMEH BIGLARI GHOLILOU

June 2010

Chair: Professor Malik Hj Abu Hassan, PhD

Faculty: Science

This thesis introduces a general approach by proposing a new quasi-Newton (QN) equation via fourth order tensor model. To approximate the curvature of the objective function, more available information from the function-values and gradient is employed. The efficiency of the usual QN methods is improved by accelerating the performance of the algorithms without causing more storage demand.

The presented equation allows the modification of several algorithms involving QN equations for practical optimization that possess superior convergence property. By using a new equation, the BFGS method is modified. This is done twice by employing two different strategies proposed by Zhang and Xu (2001) and Wei et al. (2006) to generate positive definite updates. The superiority of these methods compared to the standard BFGS and the modification proposed by Wei et al. (2006) is shown. Convergence analysis that gives the local and

global convergence property of these methods and numerical results that shows the advantage of the modified QN methods are presented.

Moreover, a new limited memory QN method to solve large scale unconstrained optimization is developed based on the modified BFGS updated formula. The comparison between this new method with that of the method developed by Xiao et al. (2008) shows better performance in numerical results for the new method. The global and local convergence properties of the new method on uniformly convex problems are also analyzed.

The compact limited memory BFGS method is modified to solve the large scale unconstrained optimization problems. This method is derived from the proposed new QN update formula. The new method yields a more efficient algorithm compared to the standard limited memory BFGS with simple bounds (L–BFGS–B) method in the case of solving unconstrained problems. The implementation of the new proposed method on a set of test problems highlights that the derivation of this new method is more efficient in performing the standard algorithm.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PERSAMAAN QUASI-NEWTON BAHARU DAN KAEDAH MELALUI MODEL TENSOR BERPERINGKAT TINGGI

Oleh

FAHIMEH BIGLARI GHOLILOU

June 2010

Pengerusi: Professor Malik Hj Abu Hassan, PhD

Fakulti: Sains

Tesis ini memperkenalkan suatu pendekatan am dengan mencadangkan suatu persamaan kuasi-Newton (QN) baharu melalui model tensor berperingkat tinggi. Untuk menghampirkan kelengkungan bagi fungsi objektif, banyak lagi maklumat sedia ada daripada nilai fungsi dan kecerunanan digunakan. Kecekapan bagi kaedah QN biasa dipertingkatkan dengan mempercepatkan prestasi algoritma tanpa permintaan storan yang lebih.

Persamaan yang dikemukakan membenarkan ubahsuaian beberapa algoritma yang melibatkan persamaan QN untuk pengoptimuman praktik yang mempunyai sifat penumpuan superior. Dengan menggunakan suatu persamaan baharu, kaedah BFGS boleh diubahsuaikan. In dibuat dua kali dengan menggunakan dua strategi berlainan yang dicadangkan oleh Zhang dan Xu (2001) dan Wei et al. (2006) untuk menjana kemaskinian tentu positif. Kesuperioran bagi kaedah-kaedah ini dibandingkan dengan BFGS piawai dan ubahsuaian yang dicadangkan oleh Wei et al. (2006) ditunjukkan. Analisis penumpuan yang memberi sifat penumpuan setempat dan sejagat bagi kaedah-kaedah ini dan keputusan

berangka yang menunjukkan kebaikan kaedah-kaedah QN terubahsuai di kemukakan.

Tambahan pula, berdasarkan rumus kemaskini BFGS terubahsuai suatu kaedah ingatan terhad QN baharu untuk menyelesaikan pengoptimuman tak berkekangan berskalar besar dibangunkan. Perbandingan antara kaedah baharu ini dengan kaedah yang dibangunkan oleh Xiao et al. (2008) menunjukkan prestasi lebih baik dalam keputusan berangka untuk kaedah baharu. Sifat penumpuan sejagat dan setempat bagi kaedah baharu ke atas masalah cembung seragam juga dianalisiskan.

Dalam kajian ini, kaedah ingatan terhad BFGS yang padat adalah juga diubahsuaikan untuk menyelesaikan masalah pengoptimuman tak berkekangan berskalar besar. Kaedah ini diterbitkan daripada rumus kemaskini QN baharu yang dicadangkan. Kaedah baharu ini memberikan satu algoritma yang lagi cekap berbanding kaedah ingatan terhad BFGS piawai dengan batas ringkas (L–BFGS–B) dalam kes menyelesaikan masalah tak berkekangan. Implementasi kaedah baharu yang dicadangkan ke atas satu set masalah ujian menggariskan yang terbitan kaedah baharu in lebih cekap dalam prestasi algoritma piawai.

ACKNOWLEDGEMENTS

First and foremost, I wish to express my genuine praise to the sublimate and almighty ALLAH for His blessings and auspices.

I am sincerely grateful to my supervisor, Prof. Dr. Malik Hj Abu Hassan, for giving me the opportunity to work under his supervision, for his genuine interest in my research and career, for never being too busy to set regular time aside for getting together, for stimulating conversations and valuable advice on many topics, for his patience and for making sure that I stayed on track when the crunch was on. He is everything that one could hope for in a supervisor, and more.

I had the honor to learn from Assoc. Prof. Dr. Leong Wah June and Dr. Mansor Monsi the members of my thesis committee are gratefully acknowledged.

I would like to thank my husband Gholamreza Mansourfar and my son, Moeinoddin for their support, patience and understanding during this time. It is to them that I dedicate this work, with love and gratitude.

Additional thanks to my friend Farzin Modarres who contributed towards this thesis by providing me with valuable feedback about my research and my technical writings.

Thanks to all who without their support and encouragement, I could not have arrived at the goal of doctoral study at UPM.

I certify that an Thesis Examination Committee has met on 28 June 2010 to conduct the final examination of Fahimeh Biglari Gholilou on her thesis entitled "New Quasi-Newton Equation and Methods Via Higher Order Tensor Models" in accordance with Universities and University Colleges Act 1971 and the Constitution of the University Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Fudziah Binti Ismail, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairperson)

Mohamed Suleiman, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Norihan Binti Md. Arifin, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Kok Lay Toe, PhD
Professor
Department of Mathematics and Statistics
University Curtin
Australia
(External Examiner)

BUJANG KIM HUAT, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 23 July 2010

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

Malik Hj Abu Hassan, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Leong Wah June, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Mansor Monsi, PhD
Faculty of Science
Universiti Putra Malaysia
(Member)

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

Date: 12 August 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.

FAHIMEH BIGLARI GHOLILOU

Date: 7 July 2010

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

x	variables in an optimization problem
$f(x)$	objective function
$\Re^n$	n–dimensional space
x_{k-1}, x_k	last iterate, new iterate
p, p_k	search direction (on iteration k)
P^T, p^T	transpose
x^*	local minimizer or local solution
N	a neighborhood of x^*
l_b, u_b	lower bound, upper bound on the variables
g_k, G_k	derivative, second derivative (of f at x_k)
α, α_k	stepsize (on iteration k)
O(.), o(.)	let $\phi(n)$ is a positive value and steadily uniform function of n, then if there is a constant K such that $\vert f\vert \leq K\phi$ for $n \geq n_0$, we write $f = O(\phi)$. if $f/\phi \rightarrow 0$ when $n \rightarrow \infty$ we write $f = o(\phi)$ (see [30])
$\Vert.\Vert$	norm of a vector or matrix
$q(p)$, $m(p)$	quadratic function (3.19), (2.1)
B, B_k	matrix that approximate matrix G (on iteration k)
SPD	Symmetric Positive Definite
I	unit matrix
s, s_k	correction to x_k
c_1, c_2, η	fixed parameters in algorithms

QN	quasi Newton method
BFGS	method of Broyden, Fletcher, Goldfarb and Shanno
LBFGS	limited memory BFGS
L–BFGS–B	limited memory BFGS with simple bounds on the variables
BFGS–T	BFGS type method page 30
LBFGS–T	Limited memory BFGS type method
CLBFGS–T	Compact Limited memory BFGS type method
MBFGS–T	Modified BFGS type method
LMBFGS–T	Limited memory modified BFGS type method
CLMBFGS–T	Compact limited memory modified BFGS type method
BFGS–CU	BFGS with cautious updates method
MBFGS–CU	Modified BFGS with cautious updates method

CHAPTER 1

INTRODUCTION

1.1 Introduction

Secant modifications are proposed as a way to promote fast convergence of optimization algorithms which involve quasi Newton (QN) updates. They are designed to replace usual secant equation which uses information of gradient at the two most recent iterates by a new one that also employs the objective function values. The stated goal of these modifications is to get more curvature information in construction of quadratic model to the objective function to yield more effective minimization processes of algorithms.

The objectie of this thesis is to propose a general approach via new QN equation to derive new algorithms involving QN updating formulas and to improve these methods in efficiency to solve unconstrained nonlinear problems.

1.2 Basic concepts and methods

Definition 1.2.1. A square matrix A is positive definite if there is a positive scalar α such that

$$x^T Ax \geq \alpha x^T x, \qquad \text{for all x} \in \Re^{\text{n}}.$$

It is positive semidefinite if

$$x^T Ax \geq 0, \qquad \text{for all x} \in \Re^{\text{n}}.$$

Definition 1.2.2. A norm is any mapping $\|.\|$ from $\Re^n$ to the nonnegative real numbers that satisfies the following properties:

(i) $\|x+z\| \leq \|x\| + \|z\|$ for all x, z $\in \Re^n$;

(ii) $\|x\| \geq 0$, for all x $\in \Re^n$; $\|x\| = 0 \Rightarrow x = 0$;

(iii) $\|\alpha x\| = |\alpha| \|x\|$, for all $\alpha \in \Re$ and x $\in \Re^n$.

The well known examples of vector norm are as follows:

$$\begin{aligned} \|x\|_\infty &= \max_{1\leq i\leq n} |x_i|, \quad (l_\infty - norm) \\ \|x\|_1 &= \sum_{i=1}^{n} |x_i|, \quad (l_1 - norm) \\ \|x\|_2 &= (\sum_{i=1}^{n} |x_i|^2)^{1/2}, \quad (l_2 - norm) \end{aligned}$$

The above examples are particular cases of l_p-norm which is defined as

$$\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}, \ (l_p - \text{norm}).$$

For the Euclidean norm $\|.\| = \|.\|_2$ the Cauchy-Schwarz inequality holds, which states that

$$|x^T z| \leq \|x\| \|z\|,$$

with equality if and only if one of these vectors is a nonnegative multiple of the other. Another vector norm is the ellipsoid norm which is defined as

$$\|x\|_A = (x^T A x)^{1/2},$$

where $A \in \Re^{n\times n}$ is a symmetric and positive definite matrix.

Similarly, a matrix norm can be defined

Definition 1.2.3. Let A, $B \in \Re^{m\times n}$. A mapping $\|.\| : \Re^{m\times n} \to \Re$ is said to be a matrix norm if it satisfies the properties

(i) $\|A+B\| \leq \|A\| + \|B\|$ for all A, B $\in \Re^{m\times n}$;

(ii) $\|A\| \geq 0$, for all A $\in \Re^{m\times n}$; $\|A\| = 0 \Rightarrow A = 0$;

(iii) $\|\alpha A\| = |\alpha| \|A\|$, for all $\alpha \in \Re$ and $A, \in \Re^{m\times n}$.

Corresponding to the above vector $l_p -$ norm, the matrix $l_p -$ norm is defined as

$$\|A\|_p = \sup_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \max_{\|x\|_p=1} \|Ax\|_p.$$

Explicit formulae for these norms are as follows:

$$\begin{aligned}
\|A\|_\infty &= \max_{i=1,\dots,m} \sum_{j=1}^{n} |A_{ij}|, \quad \text{(maximum row norm)} \\
\|A\|_1 &= \max_{j=1,\dots,n} \sum_{i=1}^{m} |A_{ij}|, \quad \text{(maximum column norm)} \\
\|A\|_2 &= (\lambda_{\max}(A^T A))^{1/2}, \quad \text{(spectral norm)}.
\end{aligned}$$

The most frequently used matrix norms also include the Frobenius norm

$$\|A\|_F = (\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2)^{1/2} = [tr(A^T A)]^{1/2},$$

where $tr(.)$ denotes the trace of a square matrix with $tr(A) = \sum_{i=1}^{n} A_{ii}$.

The weighted Frobenius norm and weighted l_2-norm are defined, respectively, as

$$\|A\|_{W,F} = \|W^{1/2} A W^{1/2}\|_F, \quad \|A\|_{W,2} = \|WAW\|_2,$$

where W is an $n \times n$ symmetric and positive definite matrix.

Definition 1.2.4. The condition number of a nonsingular matrix is defined as

$$\kappa(A) = \|A\| \|A^{-1}\|,$$

where any matrix norm can be used in the definition.

Definition 1.2.5. Suppose that $\{x_k\}$ is a sequence of points belonging to $\Re^n$. A sequence $\{x_k\}$ converges to some point x, written $\lim_{k\to\infty} x_k = x$, if for any $\epsilon \geq 0$, there is an index K such that

$$\|x_k - x\| \leq \epsilon, \quad \text{for all k} \geq \text{K}.$$

Definition 1.2.6. A point $x^* \in \Re^n$ is an accumulation point or limit point for $\{x_k\}$ if there is an infinite set of indices $k_1, k_2, k_3, \ldots$ such that the subsequence $\{x_{k_i}\}_{i=1,2,3,\ldots}$ converges to x^*; that is

$$\lim_{i\to\infty} x_{k_i} = x^*.$$

Definition 1.2.7. Let $\{x_k\}$ be a sequence in $\Re^n$ that converges to x^*. The convergence is said to be Q-linear if there is a constant $r \in (0, 1)$ such that

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq r, \quad \text{for all } k \text{ sufficiently large.}$$

The convergence is said to be Q-superlinear if

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0$$

and the convergence is called Q-quadratic if

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} \leq M \quad \text{for all } k \text{ sufficiently large,}$$

where M is a positive constant, not necessarily less than 1.

Definition 1.2.8. The convergence is R-linear if there is a sequence of nonnegative scalars $\{v_k\}$ such that

$$\|x_k - x^*\| \leq v_k \quad \text{for all } k, \text{ and} \{v_k\} \text{converges Q-linearly to zero.}$$

Likewise, $\{x_k\}$ converges R-superlinearly to x^* if $\{\|x_k - x^*\|\}$ is dominated by a sequence of scalars converging Q-superlinearly to zero, and $\{x_k\}$ converges R-quadratically to x^* if $\{\|x_k - x^*\|\}$ is dominated by a sequence converging Q-quadratically to zero.