UNIVERSITI PUTRA MALAYSIA

SYMMETRIC RANK-ONE METHOD AND ITS MODIFICATIONS FOR UNCONSTRAINED OPTIMIZATION

ALIYU USMAN MOYI

FS 2014 44

SYMMETRIC RANK-ONE METHOD AND ITS MODIFICATIONS FOR UNCONSTRAINED OPTIMIZATION

By

ALIYU USMAN MOYI

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

June 2014

DEDICATIONS

To the memory of my late father,

Alhaji Moyi Hannu da Zuma

and my beloved mother

Hajiya Zulaihatu Ibrahim

May Allah bestow His mercy upon them and make paradise their final place of aboard.

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

SYMMETRIC RANK-ONE METHOD AND ITS MODIFICATIONS FOR UNCONSTRAINED OPTIMIZATION

By

ALIYU USMAN MOYI

June 2014

Chair: Associate Professor Leong Wah June, PhD

Faculty: Science

The attention of this thesis is on the theoretical and experimental behaviors of some modifications of the symmetric rank-one method, one of the quasi-Newton update for finding the minimum of real valued function f over all vectors $x \in \mathbb{R}^n$. Symmetric rank-one update (SR1) is known to have good numerical performance among the quasi-Newton methods for solving unconstrained optimization problems. However, it is well known that the SR1 update may not preserve positive definiteness even when updated from a positive definite approximation and can be undefined with zero denominator. Thus, it is our aim in this thesis to provide effective remedies aimed toward dealing with these well known shortcomings and improve the performance of the update.

A new inexact line search strategy in solving unconstrained optimization problems is proposed. This method does not require the evaluation of the objective function. Instead, it forces a reduction in gradient norm on each direction, hence it is suitable for problems when function evaluation is very costly. The convergence properties of this strategy is shown using the Lyapunov function approach. Similarly, we proposed some scaling strategies to overcome the challenges of the SR1 update. Under some mild assumptions, the convergence of these methods is proved. Furthermore, in order to exploit the good properties of the SR1 update in providing quality Hessian approximations, we introduced a three-term conjugate gradient method via the symmetric rank-one update in which a conjugate gradient line search direction is constructed without the computation and storage of matrices and possess the sufficient descent property. Extensive computational

experiments performed on standard unconstrained optimization test functions and some real-life optimization problems in order to examine the impact of the proposed methods in comparison with other existing methods has shown significant improvement on the performance of the SR1 method in terms of efficiency and robustness.

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

KAEDAH PANGKAT-SATU SIMETRI DAN PENGUBAHSUAIANNYA UNTUK PENGOPTIMUMAN TAK BERKEKANGAN

Oleh

ALIYU USMAN MOYI

Jun 2014

Pengerusi: Professor Madya Leong Wah June, PhD

Fakulti: Sains

Tumpuan tesis ini adalah mengenai tingkah laku secara teori dan eksperimen beberapa pengubahsuaian kaedah pangkat-satu simetri, salah satu daripada kemas kini kuasi-Newton(QN) untuk mencari minumum bagi fungsi bernilai nyata f ke atas semua vektor $x \in \mathbb{R}^n$. Kemaskini pangkat-satu simetri (SR1) diketahui mempunyai prestasi berangka yang baik di antara kaesah kuasi-Newton untuk menyelesaikan masalah pengoptimuman tak berkekangan. Walau bagaimanapun, adalah diketahui bahawa kemaskini SR1 tidak boleh memelihara ketentu-positifan walaupun dikemaskini dari anggaran yang tentu positif dan boleh jadi tak tertakrif dengan pembahagi sifar. Oleh itu, adalah menjadi matlamat kami di dalam tesis ini untuk menyumbangkan penyelesaian yang berkesan bertujuan mengatasi kelemahan tersebut dan meningkatkan prestasi kemaskini ini.

Satu strategi carian garis tidak tepat baru dalam menyelesaikan masalah pengoptimuman tak berkekangan dicadangkan. Strategi ni tidak memerlukan penilaian fungsi objectif. Sebaliknya, ia memaksa pengurangan dalam norma kecerunan pada setipa lellaran. Oleh itu, ia sesuai untuk masalah apabila penilaian fungsi adalah sangat mahal. Sifai-sifat penumpuan strategi ini telah ditunjuk dengan menggunakan pendekatan fungsi Lyapunov. Begitu juga, kami mencadangkan beberapa strategi penskalaan untuk mengatasi cabaran kamiskini SR1. Penumpuan bagi kaedah-kaedah ini terbukti di bawah beberapa andaian ringan . Tambahan pula, untuk mengekploitasi sifat-sifat baik daripada kemas kini SR1 dalam memperkenalkan kaedah kecerunan konjugat tiga sebutan melalui kemaskini pangkat-satu di mana suatu arah carian garis konjugat kecerunan dibina tanpa pengiran

dan penyimpanan matriks dan memiliksi sifat penurunan yang mencukupi. Pengiraan ujikaij dijalakan je atas fungsi ujian pengoptimuman tak berkekangan piawai dan beberapa masalah pengoptimuman dunia sebenar dikaji. Stategi yang dicadangkan telah menunjukkan peningkatan yang sangat ketara prestaisi kaedah SR1 dari segi kecakapan dan kekukuhan berbanding strategi dengan yang sedia ada.

ACKNOWLEDGEMENTS

In the name of Allah, the Beneficent, the Most Merciful, all thanks and praises be to Him, the Lord of the worlds Who has taught man that which he knew not. First and foremost, I begin by expressing my profound gratitude to Almighty Allah for sparing my life to witness this important opportunity of starting and completing this task successfully. His bounties are enormous, for there were times when things were very tough and He gave me the strength and ability to persevere.

I can not search for the right words to express the debt of gratitude I owe the chairman of my supervisory committee, Associate Professor Dr. Leong Wah June, whose excellent supervision, precious discussions, insightful comments, helpful guidance and many more too numerous to mention here, enabled me to develop a good understanding of optimization, it is to his credit that I am able to reach this point in time, I am very happy and grateful for having the opportunity to work under his supervision and the least I can say is may God reward him, bestow more knowledge unto him and protect him throughout his future endeavors.

To my co-supervisor Professor Dr. Fudziah Binti Ismail, a caring mother, a good listener, a visionary leader whom we can depend upon for her support, invaluable contributions throughout this journey, may Allah continue to guide and strengthen her courage to maintain the good work she is doing.

I would like to thank the other member of my supervisory committee, Associate Professor Mohd Rizam Bin Abu Bakar for his support and contributions in this thesis. Indeed he has been very helpful at all times.

My deepest gratitude to my mother for her support, prayers and encouragement during this study period and in all times of my life and by extension to my family in ensuring that my dream becomes a reality. I thank God for the blessing of having her as my mother.

Similarly my sincere appreciation goes to my beloved wife Khadija Muazu Lawal and our children Abdurrahman, Aisha and Zainab for being with me here in Malaysia, and for having to share the pains and gains of this struggle, may Allah bless, guide and reward them all.

Further appreciations goes to my mentor, my hope, I was brought up under his good care, he supported my education right from primary school up to the university level, without which it would have been the other side of the story, in person of Alhaji Usman Moyi Kaura and his wife Hajiya Saratu, others are Alhaji Ismail Moyi, Alhaji Muhammad Albada Moyi, Ahaji Mainasara Moyi, Malam Abubakar Lawal, Ahmad Lawal and the entire Barayar zaki family, I will always remain grateful for the various contributions and support I received from you at different

stages of my life.

Lastly, my appreciation goes to the former Rector of Abdu Gusau Polytechnic, Talata Mafara, Engineer Alhaji Sanusi Yau (Garkuwan Kaura), present Rector Alhaji Abubakar Liman Shinkafi, Malam Samaila Mohammad Yusuf Auki Bungudu, Dr. Abubakar Roko Shinkafi, Dr. Muhammad Waziri Yusuf, Dr. Ibrahim Saidu, staff of Abdu Gusau Polytechnic and all those who prayed for us in this struggle, may Allah reward you all.

I certify that a Thesis Examination Committee has met on 13 June 2014 to conduct the final examination of Usman Aliyu Moyi on his thesis entitled "Symmetric Rank-One Method and its Modifications for Unconstrained Optimization" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti 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:

Zarina Bibi binti Ibrahim, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Zanariah binti Abdul Majid, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Ibragimov Gafurjan, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Honglei Xu, PhD
Associate Professor
Curtin University
Australia
(External Examiner)

NORITAH OMAR, PhD
Associate Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 23 June 2014

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

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

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

Mohd Rizam Bin Abu Bakar, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

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

Date:

Declaration by Graduate Student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ______________________ Date: ______________________

Name and Matric No.: ____

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under the supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ____________________
Name of
Chairman of
Supervisory
Committee: ____________________

Signature: ____________________
Name of
Member of
Supervisory
Committee: ____________________

Signature: ____________________
Name of
Member of
Supervisory
Committee: ____________________

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

x	Vector of variables
f	Real-valued Objective function
$\mathbb{R}^n$	n-dimensional space
x_{k-1}, x_k	Previous iterate point, current iterate point
d, d_k	Search direction (at kth iteration)
x^*	Local minimizer or solution of an optimization problem
$N(x^*)$	A neighborhood of a solution x^*
$g_k = g(x_k) = \nabla f(x_k)$	First derivative (gradient) of f at x_k
$G_k = G(x_k) = \nabla^2 f(x_k)$	Second derivative (Hessian) of f at x_k
α, α_k	Steplength (at kth iteration)
$(\|\cdot\|_2)$	Norm of a vector or matrix (l_2-norm)
A^T	The transpose of a matrix A
B_k	$n \times n$ matrix approximation to G at kth iteration
H_k	$n \times n$ matrix approximation to G^{-1} at kth iteration
Itrn	Iteration count
Feval	Function Evaluation
SD	Steepest descent method
QN	Quasi-Newton method
CG	Conjugate gradient method
BFGS	Broyden, Fletcher, Goldfarb and Shanno method
SR1	Symmetric rank-one method
GD	Gradient descent line search method

CHAPTER 1

INTRODUCTION

1.1 Background

Optimization problems are the products of processes in the real world, from science and engineering to economics. They constantly emanate from our desire to choose the best among competing alternatives that require some decision making, and in solving these problems we attempt to find the degree of goodness of the alternative conveyed by an objective function. Undoubtedly, optimization is an important tool essential in any problem involving decision making. In general, methods concerned with finding the maximum or minimum of a given function of many real variables can be term as optimization methods. According to Walsh (1975) "the study of optimization techniques is attractive because of its very wide field of application arising from diverse discipline". One of the factor that gives a tremendous boom to the growth and application of optimization methods is the advent of computer, which saw numerous optimization techniques been developed and are constantly being applied in solving real-life problems, but the choice of method for solving a given problem is largely a matter of personal preference, since according to Nocedal and Wright (2006), "there is no universal optimization algorithm but rather a collection of algorithms, each of which is designed to a particular type of optimization problem". In practice, it was observed that a particular algorithm may be a good option in solving certain type of minimization problems, but its efficiency degenerates when applied to solve other categories of problems (see Phua (1997) for instance).

1.2 Fundamental Concepts and Basic Definitions

Consider the optimization problem

$$\min \ f(x) \tag{1.1}$$

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

where x is the vector of variables of the optimization problem, D is a subset of $\mathbb{R}^n$ called the constraint or feasible set and the function $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is the objective function. The best vector x of the decision variable over all possible vectors in D known as the optimum vector that solves the problem (1.1) is denoted by x^* with a corresponding optimum function value $f(x^*)$. Classifying problems with the general form (1.1) can be done by considering the nature of the objective function, the constraints (linear, nonlinear, convex), the number of variables (small, medium or large), the smoothness of the given function (differentiable or non differentiable) and so on. But basically, the main difference is between those problems having and those not having constraints on the variables. If $D = \mathbb{R}^n$, the problem is called unconstrained minimization problem,

$$\min_{x \in \mathbb{R}^n} f(x). \tag{1.2}$$

Otherwise it is a constrained minimization problem.

In unconstrained optimization, we minimize an objective function that depends on real variables and the values of these variable have no restrictions at all. It is worth mentioning that there is no difference, for the values of the variables that minimize $f(x)$ also maximize $-f(x)$. Therefore, we are going to restrict our attention in this thesis to minimizing unconstrained optimization problems only. Many modern optimization techniques are designed to solve specifically the general unconstrained optimization problems. For a given constrained optimization problem, techniques for unconstrained optimization problems can be used to solve these problems since they constitute the foundation for the constrained-problems, in which the constraints are substituted by penalization terms in the objective function to cushion the effect of constraint violation. To have a better understanding of the methods and to follow the development described in this thesis, the following definitions which we used throughout are briefly outlined (See Nocedal and Wright (2006) for details):

Definition 1.2.1 A vector $x^* \in D$ is a *global minimizer* of f over D if $f(x^*) \leq f(x)$ for all x, where x ranges over all of D.

Definition 1.2.2 A vector $x^* \in D$ is called a *local minimizer* of f over D if there is an open neighborhood N of x^* such that $f(x^*) \leq f(x)$ for all $x \in N$.

Definition 1.2.3 A norm is any mapping $\| \cdot \|$ from $\mathbb{R}^n$ to the non-negative real numbers, such that the following conditions are satisfied for all $x, y \in \mathbb{R}^n$ and all $\alpha \in R$.
(i) $\|x\|$=0 if and only if $x = 0$
(ii) $\|x + y\| \leq \|x\|$+$\|y\|$
(iii) $\|\alpha x\|$=$|\alpha|\|x\|$

Definition 1.2.4 Let A, B $\in \mathbb{R}^{m \times n}$. A mapping $\| \cdot \|$: $\mathbb{R}^{m \times n} \longrightarrow \mathbb{R}$ is said to be a matrix norm if the following properties are satisfied for all $A, B \in \mathbb{R}^{m \times n}$ and all $\alpha \in R$:
(i) $\|A + B\| \leq \|A\|$+$\|B\|$
(ii) $\|\alpha A\|$=$|\alpha|\|A\|$.

Let $g(x)$ represents the gradient of f at x whose i-th component is defined as

$$[g(x)]_i = \frac{\partial f(x)}{\partial x_i}, \quad i = 1, ..., n \tag{1.3}$$

and also $G(x)$ to denote the $n \times n$ Hessian matrix of f at x whose ij-th component is defined as

$$[G(x)]_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i = 1, ..., n, \quad j = 1, ..., n \tag{1.4}$$

Definition 1.2.5 A function f is said to be *Lipschitz continuous* in an open neighborhood $D \subset \mathbb{R}^n$ if there is a constant $\gamma > 0$ such that

$$|f(x) - f(y)| \leq \gamma \|x - y\| \quad \text{for all } x, y \in D \tag{1.5}$$

where $\| \cdot \|$ is a selected norm and γ is called Lipschitz constant.

Definition 1.2.6 Let A be a square matrix, we say that A is *symmetric* if $A = A^T$, and a symmetric matrix A is *positive definite* if

$$x^T A x > 0 \quad \text{for all } x \in \mathbb{R}^n, \quad x \neq 0 \tag{1.6}$$

Similarly by applying the weak inequality to (1.6), A is called *positive semidefinite* that is when

$$x^T A x \geq 0. \tag{1.7}$$

Definition 1.2.7 Two vectors $x, y \neq 0$ are said to be *orthogonal* if the scalar product $x^T y = (x, y) = 0$. In the same degree, two vectors $x, y \neq 0$ are termed to be *mutually conjugate* with respect to the matrix A if $x^T A y = (x, Ay) = 0$, where A is a positive definite symmetric matrix.

Definition 1.2.8 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function that is continuously differentiable, we say that $x^* \in \mathbb{R}^n$ is a *stationary or critical point*, if $\nabla f(x^*) = 0$.

Definition 1.2.9 If a stationary or critical point is neither a local minimizer nor a local maximizer such a point is called a *saddle point*.

Definition 1.2.10 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function at x. We say that the direction $d \in \mathbb{R}^n$ is a *direction of decrease* if there exists a constant $\sigma > 0$ such that $f(x + \alpha d) < f(x)$ for all $\alpha \in (0, \sigma)$.

Definition 1.2.11 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function that is differentiable at $x^* \in \mathbb{R}^n$. A direction $d \in \mathbb{R}^n$ is said to be a *descent direction* of f at x if

$$d^T \nabla f(x) < 0. \tag{1.8}$$

Lemma 1.2.1 (Nocedal and Wright (2006))
Let U and V be matrices in $\mathbb{R}^{n \times p}$ for some p between 1 and n and let $L(\mathbb{R}^n)$ define the linear space of all matrices of order n. The rank-k update of a nonsingular

matrix $A \in L(\mathbb{R}^n)$ of the form

$$\tilde{A} = A + UV^T \tag{1.9}$$

is nonsingular if and only if $\mu = I + V^T A^{-1} U \neq 0$. In particular if $\mu \neq 0$, then the inverse of $\tilde{A}$ is given by

$$\tilde{A}^{-1} = A^{-1} - A^{-1}U(I + V^T A^{-1} U)^{-1} V^T A^{-1} \tag{1.10}$$

Equation (1.10) is known as the Sherman-Morrison formula, a very explicit way of expressing the inverse of a matrix. For the original proof of the formula, see Sherman and Morrison (1950)

Definition 1.2.12 The condition number $k(A)$ of an $n \times n$ nonsingular matrix defined by

$$k(A) = \frac{\lambda_1}{\lambda_n}$$

is the ratio of its largest and smallest eigenvalue.

This ratio is a measure that tells the extent of the difficulty to solve the associated linear system $Ax = b$ or alternatively the condition number $k(A)$ quantifies the sensitivity of the problem $Ax = b$. Matrices with large condition numbers are said to be *ill-conditioned* while those with small condition numbers are said to be *well-conditioned.*

Taylor series expansion serves as the building block in the techniques and methods for locating the minimizer of a nonlinear differentiable function. for this reason we now state it in the following theorem :

Theorem 1.2.1 Taylor's Theorem (See, for example Nocedal and Wright (2006)) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function that is continuously differentiable and that $d \in \mathbb{R}^n$ then we have that

$$f(x + d) = f(x) + \nabla f(x + \tau d)^T d, \tag{1.11}$$

for some $\tau \in (0, 1)$. In addition, if f is twice continuously differentiable, it will leads to

$$\nabla f(x + d) = \nabla f(x) + \int_0^1 \nabla^2 f(x + \tau d)^T d \, d\tau, \tag{1.12}$$

and that

$$f(x + d) = f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x + \tau d)^T d, \tag{1.13}$$

for some constant $\tau \in (0, 1)$.

Since most algorithms for unconstrained optimization are iterative, they will generate a sequence of iterates that are intended to converge. Convergence rate is of interest to measure how quickly the iterates of an algorithm converges to the solution x^* of a given function. Understanding the convergence rate of an algorithm is very important since according to Dennis and More (1977) important as a method is its convergence rate, if the method converges slowly we may not be able to see it converge. Therefore, in agreement with the above assertion we outlined the following definitions which will give an insight into how convergence rate is basically characterized:

Definition 1.2.13 Let $\{x_k\}$ denote a sequence in $\mathbb{R}^n$ that converges to the final solution x^*, for some norm $\|\cdot\|$. Given a constant $\xi \in (0,1)$, we say that the convergence is *q-linear* if

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq \xi, \quad \text{for all } k \text{ sufficiently large} \tag{1.14}$$

Definition 1.2.14 The convergence is termed *q-superlinear* if

$$\lim_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0 \tag{1.15}$$

and it is called *q-quadratic* convergence if there exist a positive constant G, not necessarily less than 1, such that

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} \leq G, \quad \text{for all } k \text{ sufficiently large} \tag{1.16}$$

Remark:
"*Any sequence that converges q-quadratically also converges q-superlinearly and any sequence that converges q-superlinearly also converges q-linearly.*" (Nocedal and Wright (2006))

Definition 1.2.15 The rate of convergence is termed *r-linear* where ('*r*' stands for 'root') if there is a sequence of nonnegative scalers $\{\nu_k\}$ such that

$$\|x_k - x^*\| \leq \nu_k \quad \text{for all } k \tag{1.17}$$

and $\{\nu_k\}$ converges q-linearly to zero. The above definition is a weaker type of convergence rate in which the overall error decrease rate is significant than the decrease over each individual step of the algorithm Dennis and Schnabel (1983).

Conventionally, we call for the termination of an algorithm whenever the condition $\nabla f(x_k) = 0$ is attained. However, this condition is seldom found in practice because the numerical evaluation of the gradient hardly becomes identically equal to zero. Therefore, we need a more practical stopping criterion to ascertain whether a method has converges to the required solution or not. In order to promise the convergence of a given method, we impose the condition that either

$$|f(x_k) - f(x^*)| \leq \epsilon \quad \text{or} \quad \|x_k - x^*\| \leq \epsilon$$

where ϵ is a user specified. In practical realities, these conditions are also not obtainable since we require the information of x_*, therefore we opt to use the following criteria frequently use to stop iterative methods for unconstrained optimization:

$$\|\nabla f(x_k)\| \leq \epsilon \tag{1.18}$$

$$\|\nabla f(x_k)\| \leq \epsilon \times \max(1, \|x_k\|) \tag{1.19}$$

where ϵ is the machine precision.

1.3 Optimality Conditions for Unconstrained Optimization

Optimality conditions constitutes one of the basis for locating the solution of an unconstrained optimization problem in the algorithms we are considering in this thesis. These conditions are derived by letting the solution point x^* to be the local minimizer of a continuously differentiable function f. In the light of the above, we are here giving a brief review on these conditions for which their proofs can be found in Nocedal and Wright (2006).

Theorem 1.3.1 (First-Order Necessary Optimality Conditions)
Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable in an open neighborhood $D \subset \mathbb{R}^n$, if $x^* \in D$ is a local solution to the problem (1.2) then

$$\nabla f(x^*) = 0 \tag{1.20}$$

Theorem 1.3.2 (Second-Order Necessary Conditions for optimality)
Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable in an open neighborhood $D \subset \mathbb{R}^n$, if $x^* \in D$ is a local solution to the problem (1.2) then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semidefinite.

Theorem 1.3.3 (Second-Order Sufficient Conditions for Optimality)
Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable in an open neighborhood $D \subset \mathbb{R}^n$, and $x^* \in D$, if $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite, then x^* is a strict local minimizer of f.

1.4 Convexity

Convexity plays a major role in the theory of minimization algorithms, since many objective functions are convex within some neighborhood of a local minimizer and convergence analysis of numerical methods for locating local minimizers can easily be launched for convex objective functions (Wolfe (1978)). The term convex can both be applied to sets and functions. Convex set and convex function do feature most frequently in numerous areas of applied science and Mathematics. Thus we give the following definitions:

Definition 1.4.1 A set $D \subset \mathbb{R}^n$ is *convex* if for any two points $x_1, x_2 \in D$

and for any constant $\lambda \in [0, 1]$, we have

$$\lambda x_1 + (1 - \lambda)x_2 \in D. \tag{1.21}$$

Definition 1.4.2 A function f is said to be a *convex function* over a nonempty set $D \subset \mathbb{R}^n$ if for any two points $x_1, x_2 \in D$ and for all $\lambda \in [0, 1]$, we have

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2). \tag{1.22}$$

Similarly, the function f is called *strictly convex*, if we replace the inequality in (1.22) by a strict inequality such that $x_1 \neq x_2$ and the constant λ is in the open interval (0,1).

Definition 1.4.3 A function f is called a *uniformly (or strongly) convex function* on D, if there is a constant $h > 0$ such that for any two points $x_1, x_2 \in D$;

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) - \frac{1}{2}h\lambda(1 - \lambda)\|x_1 - x_2\|^2. \tag{1.23}$$

Theorem 1.4.1 Let $D \subset \mathbb{R}^n$ be a nonempty set and $f : \mathbb{R}^n \to \mathbb{R}$ a convex function on D, it follows that any local minimizer x^* is a global minimizer of f and if in addition f is continuously differentiable, then any stationary point x^* is a global minimizer of f.

proof. See Nocedal and Wright (2006)

In the next part, we give some theorems on first and second order conditions of differentiable convex functions with the following characterization, the proof which can be seen as advanced by Andrei (2007).

Theorem 1.4.2 Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function on D, where $D \subset \mathbb{R}^n$ is an open nonempty set, then we have:
(i) f is convex if and only if , for any two points $x_1 \neq x_2 \in D$:

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T(x_2 - x_1). \tag{1.24}$$

(ii) f is strictly convex on D if and only if , for any two points $x_1, x_2 \in D$, $x_1 \neq x_2$:

$$f(x_2) > f(x_1) + \nabla f(x_1)^T(x_2 - x_1). \tag{1.25}$$

(iii) f is uniformly (or strongly) convex function on D, if and only if , for any two points $x_1, x_2 \in D$, there is a constant $h > 0$ such that:

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T(x_2 - x_1) + \frac{1}{2}h\|x_1 - x_2\|^2. \tag{1.26}$$

Theorem 1.4.3 Let $f : D \subset \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable function on D, where $D \subset \mathbb{R}^n$ is an open nonempty set, then:

(i) f is convex if and only if its Hessian matrix is positive semidefinite at each point in D.
(ii) f is strictly convex if and only if its Hessian matrix is positive definite at each point in D.
(iii) f is uniformly (or strongly) convex function if and only if its Hessian matrix is uniformly positive definite at each point in D, that is for a constant $a > 0$ we have:

$$a\|\nu\|^2 \leq \nu^T \nabla^2 f(x)\nu. \qquad \forall x \in D,\ \nu \in \mathbb{R}^n.$$

1.5 Statement of the Problem

The major shortcomings of the SR1 update are the background of this research which includes:
1. The approximate inverse Hessian generated by the SR1 update may not preserve positive definiteness even when updated from a positive definite approximation, and thus d_{k+1} may not be a descent direction.
2. Sometimes the denominator $y_k^T(s_k - H_k y_k)$ in the SR1 update may become zero or undefined leading to numerical instabilities.

1.6 Scope of the thesis

Quasi-Newton methods are distinguished by their use of approximate Hessian matrices. These approximate matrices are evaluated with respect to some iterative update formula as the algorithm progresses. The update procedure only requires the gradient of the objective function in each iteration, these methods differ by the formula they use for updating the approximate Hessian matrix. In this thesis our focus is only on the line search implementation of the symmetric rank-one update to find the optimal solution of the general unconstrained optimization problem (1.2).

1.7 Objectives of the Thesis

In this thesis, we focus on various modifications of symmetric rank-one (SR1) method for solving small, medium and large-scale unconstrained optimization problems. These modified schemes of update not only possesses good numerical performance of the original SR1 but an improved numerical efficiency and stability, in which the major shortcomings of the SR1 of not maintaining positive definite approximations and becoming undefined with zero denominator are overcome by some simple strategies.

Specifically the objectives of the thesis are:
- To derive effective inexact line search strategy for the implementation of SR1 method.
- To develop new scaling strategies in avoiding the loss of positive definiteness of the SR1 update.

• To derive a three-term conjugate gradient method via the symmetric rank-one formula for solving large-scale unconstrained optimization problems.
• To establish the convergence results of the proposed methods.
• To present extensive numerical results on some benchmark optimization problems for evaluation of the performance of the proposed methods compared with some existing schemes.
• To apply the proposed methods in solving some real-life optimization problems.

1.8 Outline of the Thesis

The thesis is arranged accordingly into 7 chapters as follows:
In chapter 1 we present an overview of unconstrained optimization algorithm, some basic mathematical background related to the research work and the objectives of the research are highlighted.

A comprehensive review of related literature on existing line search methods for unconstrained optimization is given chapter 2 .

In chapter 3, we present a new line search strategy for the implementation of the SR1 updating scheme. Convergence results for this inexact line search method are shown. Numerical results are reported and discussed, and the chapter ends with a brief conclusion.

Some scaling strategies to overcome the major shortcoming of the SR1 update are derived in chapter 4. Convergence results of the improved SR1 methods are analyzed. Numerical experiments obtained from the improved SR1 methods with other existing variants of the SR1 method on some standard set of test problems are reported.

Chapter 5 suggests a three-term conjugate gradient method inspired by the SR1 update for solving large-scale unconstrained optimization problems. Convergence analysis for the proposed method is established. Conclusions are drawn based upon the computational evidence at the end of the chapter.

Chapter 6 will discuss some applications of the proposed methods in solving real-life optimization problems. Numerical results obtained from the applications are presented and discussed with the chapter ends with a conclusion.

Finally, in chapter 7, we conclude the thesis with summary of the achievements based on the earlier stated objectives. An outline of a number of possible directions of related future research is given with the findings in this thesis serving as a basis.

BIBLIOGRAPHY

Al-Baali, M. 1985. Descent property and global convergence of the Fletcher-Reeves method with inexact line search. *IMA Journal of Numerical Analysis* 5 (1): 121–124.

Al-baali, M. 2000. Extra updates for the BGFS method. *Optimization Methods and Software* 13 (3): 159–179.

Andrei, N. 2007. Convex functions. *Advancd Modeling and Optimization* 9 (2): 257–267.

Andrei, N. 2008. An unconstrained optimization test functions collection. *Advanced Modeling and Optimization* 10 (1): 147–161.

Andrei, N. 2011. A modified Polak-Ribire-Polyak conjugate gradient algorithm for unconstrained optimization. *Optimization* 60 (12): 1457–1471.

Andrei, N. 2012. A simple three-term conjugate gradient algorithm for unconstrained optimization. *Journal of Computational and Applied Mathematics* (241): 19–29.

Andrei, N. 2013. On three-term conjugate gradient algorithms for unconstrained optimization. *Applied Mathematics and Computation* 219 (11): 6316–6327.

Armijo, L. 1966. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of mathematics* 16 (1): 1–3.

Bard, Y. 1968. On a numerical instability of Davidon-like methods. *Mathematics of Computation* 22 (103): 665–666.

Beale, E. M. L. 1972. A derivation of conjugate gradients, in:F.A.Lootsma (Ed.). *Numerical methods for nonlinear optimization* Academic Press, London, pp. 39–43.

Broyden, C. G. 1967. Quasi-Newton methods and their application to function minimisation. *Mathematics of Computation* 21 (99): 368–381.

Broyden, C. G. 1970a. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics* 6 (1): 76–90.

Broyden, C. G. 1970b. The convergence of a class of double-rank minimization algorithms 2. The new algorithm. *IMA Journal of Applied Mathematics* 6 (3): 222–231.

Buckley, A. and LeNir, A. 1983. QN-like variable storage conjugate gradients. *Mathematical programming* 27 (2): 155–175.

Byrd, R. and Nocedal, J. 1989. A Tool for the Analysis of Quasi-Newton Methods with Application to Unconstrained Minimization. *SIAM Journal on Numerical Analysis* 26 (3): 727–739.

Byrd, R. H., Khalfan, H. F. and Schnabel, R. B. 1996. Analysis of a symmetric rank-one trust region method. *SIAM Journal on Optimization* 6 (4): 1025–1039.

Byrd, R. H., Liu, D. C. and Nocedal, J. 1992. On the behavior of Broyden's class of quasi-Newton methods. *SIAM Journal on Optimization* 2 (4): 533–557.

Byrd, R. H., Nocedal, J. and Yuan, Y. 1987. Global convergence of a cass of quasi-Newton methods on convex problems. *SIAM Journal on Numerical Analysis* 24 (5): 1171–1190.

Cauchy, A. 1847. Methode generale pour la resolution des systemes d'equations simultanees. *Comptes Rendus de L' Academia Des Sciences Paris* 25 (1847): 536–538.

Cheng, W. 2007. A two-term PRP-based descent method. *Numerical Functional Analysis and Optimization* 28 (11-12): 1217–1230.

Chong, E. K. and Zak, S. H. 2008. *An introduction to optimization.* , vol. 76. New York, John Wiley & Sons.

Conn, A. R., Gould, N. I. and Toint, P. L. 2000. *Trust region methods*. MPS-SIAM Series on Optimization, SIAM.

Conn, A. R., Gould, N. I. M. and Toint, P. L. 1991. Convergence of quasi-Newton matrices generated by the symmetric rank one update. *Mathematical Programming* 50 (1-3): 177–195.

Dai, Y. 2001. New properties of a nonlinear conjugate gradient method. *Numerische Mathematik* 89 (1): 83–98.

Dai, Y. 2002. A nonmonotone conjugate gradient algorithm for unconstrained optimization. *Journal of Systems Science and Complexity* 15 (2): 139–145.

Dai, Y. and Liao, L. 2001. New conjugacy conditions and related nonlinear conjugate gradient methods. *Applied Mathematics and Optimization* 43 (1): 87–101.

Dai, Y. and Ni, Q. 2003. Testing different conjugate gradient methods for large-scale unconstrained optimization. *Journal of Computational Mathematics-International Edition-* 21 (3): 311–320.

Dai, Y. and Yuan, Y. 1996a. Convergence Properties of the Conjugate Descent Method. *Advances in Mathematics* 26: 552–562.

Dai, Y. and Yuan, Y. 1996b. Convergence properties of the Fletcher-Reeves method. *IMA Journal of Numerical Analysis* 16 (2): 155–164.

Dai, Y. and Yuan, Y. 1999. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization* 10 (1): 177–182.

Dai, Y. and Yuan, Y. 2001. An efficient hybrid conjugate gradient method for unconstrained optimization. *Annals of Operations Research* 103 (1-4): 33–47.

Davidon, W. C. 1959. Variable metric algorithm for minimization. *Report, Argonne National Laboratory* .

Davidon, W. C. 1968. Variance algorithm for minimization. *The Computer Journal* 10 (4): 406–410.

Dennis, J. and More, J. J. 1977. Quasi-Newton methods, motivation and theory. *SIAM review* 19 (1): 46–89.

Dennis, J. and Wolkowicz, H. 1993. Sizing and least-change secant methods. *SIAM Journal on Numerical Analysis* 30 (5): 1291–1314.

Dennis, J. E. and More, J. J. 1974. A characterization of superlinear convergence and its application to quasi-Newton methods. *Mathematics of Computation* 28 (126): 549–560.

Dennis, J. J. E. and Schnabel, R. B. 1983. *Numerical methods for unconstrained optimization and nonlinear equations*. , vol. 16. Prentcie-Hall, Englewoods cliffs, NJ, Reprinted by SIAM.

Dixon, L. C. W. 1972. Variable metric algorithms: Necessary and sufficient conditions for identical behavior of nonquadratic functions. *Journal of Optimization Theory and Applications* 10 (1): 34–40.

Dolan, E. D. and More, J. J. 2002. Benchmarking optimization software with performance profiles. *Mathematical programming* 91 (2): 201–213.

Fiacco, A. V. and McCormick, G. P. 1968. *Nonlinear programming: sequential unconstrained minimization techniques*. SIAM.

Fletcher, R. 1970. A new approach to variable metric algorithms. *The computer journal* 13 (3): 317–322.

Fletcher, R. 1987. *Practical methods of optimization*. Chichester, England, John Wiley & Sons.

Fletcher, R. 1994, An overview of unconstrained optimization: in Algorithms for Continuous Optimization, Springer-Verlag, pp. 109-143.

Fletcher, R. and Powell, M. J. 1963. A rapidly convergent descent method for minimization. *The Computer Journal* 6 (2): 163–168.

Fletcher, R. and Reeves, C. M. 1964. Function minimization by conjugate gradients. *The computer journal* 7 (2): 149–154.

Gilbert, J. C. and Nocedal, J. 1992. Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on Optimization* 2 (1): 21–42.

Gill, P. E., Murray, W. and Wright, M. H. 1981. *Practical optimization*. London, Academic Press.

Goh, B. S. 2010. Convergence of algorithms in optimization and solutions of nonlinear equations. *Journal of optimization theory and applications* 144 (1): 43–55.

Goldfarb, D. 1970. A family of variable-metric methods derived by variational means. *Mathematics of computation* 24 (109): 23–26.

Goldstein, A. A. 1962. Cauchy's method of minimization. *Numerische Mathematik* 4 (1): 146–150.

Goldstein, A. A. 1965. On steepest descent. *Journal of the Society for Industrial & Applied Mathematics, Series A: Control* 3 (1): 147–151.

Hager, W. W. and Zhang, H. 2005. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM Journal on Optimization* 16 (1): 170–192.

Hager, W. W. and Zhang, H. 2006. A survey of nonlinear conjugate gradient methods. *Pacific journal of Optimization* 2 (1): 35–58.

Hestenes, M. R. and Stiefel, E. 1952. *Methods of conjugate gradients for solving linear systems*. Journal of Research of the National Bureau of Standards, 49: 409-436.

Hu, Y. F. and Storey, C. 1991. Global convergence result for conjugate gradient methods. *Journal of Optimization Theory and Applications* 71 (2): 399–405.

Kalman, R. E. and Bertram, J. E. 1960. Control system analysis and design via the second method of Lyapunov. *Journal of Basic Engineering* 82 (2): 371–393.

Kelley, C. T. and Sachs, E. W. 1998. Local convergence of the symmetric rank-one iteration. *Computational Optimization and Applications* 9 (1): 43–63.

Khalfan, H. F., Byrd, R. H. and Schnabel, R. B. 1993. A theoretical and experimental study of the symmetric rank-one update. *SIAM Journal on Optimization* 3 (1): 1–24.

Lemarechal, C. 1981. A view of line-searches in. *Optimization and Optimal Control,* Springer–Verlag, 59–78.

Leong, W. J. and Goh, S. B. 2013. Convergence and Stability of Line Search Methods for Unconstrained Optimization. *Acta Applicandae Mathematicae* 1–13.

Leong, W. J. and Hassan, M. A. 2009a. Convergence of a Positive Definite Symmetric Rank One Method with Restart. *Advanced Modeling and Optimization* 11: 423–433.

Leong, W. J. and Hassan, M. A. 2009b. A restarting approach for the symmetric rank one update for unconstrained optimization. *Computational Optimization and Applications* 42 (3): 327–334.

Leong, W. J. and Hassan, M. A. 2011. Positive-definite memoryless symmetric rank one method for large-scale unconstrained optimization. *Pacific Journal of Optimization* 7 (1): 29–41.

Leong, W. J., Hassan, M. A. and Farid, M. 2010. A monotone gradient method via weak secant equation for unconstrained optimization. *Taiwanese Journal of Mathematics* 14 (2): 413–423.

Levitin, E. S. and Polyak, B. T. 1966. Constrained minimization methods. *USSR Computational mathematics and mathematical physics* 6 (5): 1–50.

Liu, G., Jing, L. and Han, L. 2002. Convergence Properties of the DFP Algorithm for Unconstrained Optimization. *Optimization* 51 (5): 731–758.

Liu, Y. and Storey, C. 1991. Efficient generalized conjugate gradient algorithms, part 1: theory. *Journal of Optimization Theory and Applications* 69 (1): 129–137.

Luenberger, D. 1984. *Introduction to linear and nonlinear programming*. 2nd edn. Addison Wesley.

Lukan, L. and Spedicato, E. 2000. Variable metric methods for unconstrained optimization and nonlinear least squares. *Numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations* 124 (1-2): 61–95.

McCormick, G. P. and Pearson, J. D. 1969. Variable metric methods and unconstrained optimization in:. *Optimization,* Ed. R. Fletcher, London, Academic Press, 307–325.

Modarres, F., Hassan, M. A. and Leong, W. J. 2011a. Structured symmetric rank-one method for unconstrained optimization. *International Journal of Computer Mathematics* 88 (12): 2608–2617.

Modarres, F., Hassan, M. A. and Leong, W. J. 2011b. A symmetric rank-one method based on extra updating techniques for unconstrained optimization. *Computers & Mathematics with Applications* 62 (1): 392–400.

Modarres, F. and Leong, W. J. 2012. On the performance of a new symmetric rank-one method with restart for solving unconstrained optimization problems. *Computers & Mathematics with Applications* .

Modarres, F., Malik, A. H. and Leong, W. J. 2011c. Improved Hessian approximation with modified secant equations for symmetric rank-one method. *Journal of computational and applied mathematics* 235 (8): 2423–2431.

Mohammad Sofi, A. Z., Mamat, M., Mohd, I. and Dasril, Y. 2008. An alternative hybrid search direction for unconstrained optimization. *Journal of Interdisciplinary Mathematics* 11 (5): 731–739.

More, J. J. and Sorensen, D. C. 1983. Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing* 4 (3): 553–572.

More, J. J. and Thuente, D. J. 1994. Line search algorithms with guaranteed sufficient decrease. *ACM Transactions on Mathematical Software (TOMS)* 20 (3): 286–307.

Murtagh, B. A. and Sargent, R. W. H. 1969. A constrained minimization method with quadratic convergence in:. *Optimization* Ed. R. Fletcher, London, Academic Press, 215–246.

Narushima, Y., Yabe, H. and Ford, J. A. 2011. A three-term conjugate gradient method with sufficient descent property for unconstrained optimization. *SIAM Journal on Optimization* 21 (1): 212–230.

Nazareth, L. 1977. A conjugate direction algorithm without line searches. *Journal of Optimization Theory and Applications* 23 (3): 373–387.

Nocedal, J. and Wright, S. 2006. Numerical optimization, series in operations research and financial engineering. *Springer, New York* .

Oren, S. S. and Luenberger, D. G. 1974. Self-Scaling Variable Metric (SSVM) Algorithms Part I: Criteria and Sufficient Conditions for Scaling a Class of Algorithms. *Management Science* 20 (5): 845–862.

Osborne, M. R. and Sun, L. 1999. A new approach to symmetric rank-one updating. *IMA journal of numerical analysis* 19 (4): 497–507.

Pearson, J. D. 1969. Variable metric methods of minimisation. *The Computer Journal* 12 (2): 171–178.

Perry, J. M. 1977. A class of conjugate gradient algorithms with a two step variable metric memory. *Center for Mathematical Studies in Economies and Management Science. Evanston Illiois: Northwestern University Press* .

Phua, P. K. 1997. Eigenvalues and switching algorithms for quasi-Newton updates. *Optimization* 42 (3): 185–217.

Polak, E. and Ribiere, G. 1969. Note sur la convergence de mthodes de directions conjugues. *ESAIM: Mathematical Modelling and Numerical Analysis-Modelisation Mathematique et Analyse Numerique* 3 (R1): 35–43.

Polyak, B. T. 1969. The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics* 9 (4): 94–112.

Potra, F. A. and Shi, Y. 1995. Efficient line search algorithm for unconstrained optimization. *Journal of Optimization Theory and Applications* 85 (3): 677–704.

Powell, M. J. 1971a. Recent advances in unconstrained optimization. *Mathematical Programming* 1 (1): 26–57.

Powell, M. J. D. 1970. A survey of numerical methods for unconstrained optimization. *SIAM Review* 12 (1): 79–97.

Powell, M. J. D. 1971b. On the convergence of the variable metric algorithm. *IMA Journal of Applied Mathematics* 7 (1): 21–36.

Powell, M. J. D. 1976. Some global convergence properties of a variable metric algorithm for minimization without exact line searches. *Nonlinear programming* 9: 53–72.

Powell, M. J. D. 1977. Restart procedures for the conjugate gradient method. *Mathematical programming* 12 (1): 241–254.

Powell, M. J. D. 1986. How bad are the BFGS and DFP methods when the objective function is quadratic? *Mathematical Programming* 34 (1): 34–47.

Raydan, M. 1997. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM Journal on Optimization* 7 (1): 26–33.

Shanno, D. F. 1970. Conditioning of quasi-Newton methods for function minimization. *Mathematics of computation* 24 (111): 647–656.

Shanno, D. F. 1978. Conjugate gradient methods with inexact searches. *Mathematics of operations research* 3 (3): 244–256.

Sherman, J. and Morrison, W. J. 1950. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics* 21 (1): 124–127.

Shi, Z. 2004. Convergence of line search methods for unconstrained optimization. *Applied Mathematics and Computation* 157 (2): 393–405.

Shi, Z. 2006. Convergence of quasi-Newton method with new inexact line search. *Journal of mathematical analysis and applications* 315 (1): 120–131.

Smith, B. and Nazareth, J. L. 1997. Metric-Based Symmetric Rank-One Updates. *Computational Optimization and Applications* 8 (3): 219–244.

Spellucci, P. 2001. A modified rank one update which converges Q-superlinearly. *Computational Optimization and Applications* 19 (3): 273–296.

Sun, W. and Yuan, Y. 2006. *Optimization theory and methods: nonlinear programming.* , vol. 98. springer.

Touati-Ahmed, D. and Storey, C. 1990. Efficient hybrid conjugate gradient techniques. *Journal of Optimization Theory and Applications* 64 (2): 379–397.

Vrahatis, M. N., Androulakis, G. S., Lambrinos, J. N. and Magoulas, G. D. 2000. A class of gradient unconstrained minimization algorithms with adaptive stepsize. *Journal of Computational and Applied Mathematics* 114 (2): 367–386.

Walsh, G. R. 1975. *Methods of optimization*. London, John Wiley & Sons.

Wolfe, M. A. 1978. *Numerical methods for unconstrained optimization: An introduction*. Van Nostrand Reinhold New York.

Wolfe, P. 1968, Another variable metric method, Tech. rep., working paper.

Wolfe, P. 1969. Convergence conditions for ascent methods. *Siam Review* 11 (2): 226–235.

Wolkowicz, H. 1994. Measures for symmetric rank-one updates. *Mathematics of Operations Research* 19 (4): 815–830.

Yuan, G. and Lu, X. 2008. A new line search method with trust region for unconstrained optimization. *Communications on Applied Nonlinear Analysis* 15 (1): 35.

Yuan, G. and Wei, Z. 2009. New line search methods for unconstrained optimization. *Journal of the Korean Statistical Society* 38 (1): 29–39.

Yuan, G. and Wei, Z. 2013. Non Monotone Backtracking Inexact BFGS Method for Regression Analysis. *Communications in Statistics-Theory and Methods* 42 (2): 214–238.

Yuan, Y. 1993. Analysis on the conjugate gradient method. *Dynamical Systems* 2 (1): 19–29.

Zhang, L. 2005. A globally convergent BFGS method for nonconvex minimization without line searches. *Optimization Methods and Software* 20 (6): 737–747.

Zhang, L., Zhou, W. and Li, D. 2006a. A descent modified Polak-Ribiere-Polyak conjugate gradient method and its global convergence. *IMA Journal of Numerical Analysis* 26 (4): 629–640.

Zhang, L., Zhou, W. and Li, D. 2006b. Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search. *Numerische Mathematik* 104 (4): 561–572.

Zhang, L., Zhou, W. and Li, D. 2007. Some descent three-term conjugate gradient methods and their global convergence. *Optimisation Methods and Software* 22 (4): 697–711.

Zhang, Y. and Tewarson, R. P. 1988. Quasi-Newton algorithms with updates from the preconvex part of Broyden's family. *IMA Journal of Numerical Analysis* 8 (4): 487–509.

Zheng, X., Liu, H. and Lu, A. 2011. Sufficient descent conjugate gradient methods for large-scale optimization problems. *International Journal of Computer Mathematics* 88 (16): 3436–3447.

Zoutendijk, G. 1970. Nonlinear programming, computational methods. *Integer and nonlinear programming* 143 (1): 37–86.