**UNIVERSITI PUTRA MALAYSIA**

# GLOBALIZATION OF BARZILAI AND BORWEIN METHOD FOR UNCONSTRAINED OPTIMIZATION

**MAHBOUBEH FARID**

**IPM 2009 10**

**GLOBALIZATION OF BARZILAI AND BORWEIN METHOD FOR UNCONSTRAINED OPTIMIZATION**

**By**

**MAHBOUBEH FARID**

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

**December 2009**

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

**GLOBALIZATION OF BARZILAI AND BORWEIN METHOD FOR UNCONSTRAINED OPTIMIZATION**

By

**MAHBOUBEH FARID**

**December 2009**

**Chairman: Professor Malik Hj. Abu Hassan, PhD**

**Faculty: Institute for Mathematical Research**

The focus of this thesis is on finding the unconstrained minimizer of a function. Specifically, we will focus on the Barzilai and Borwein (BB) method that is a famous two-point stepsize gradient method. First we briefly give some mathematical background. Then we discuss the (BB) method that is important in the area of optimization. A review of the minimization methods currently available that can be used to solve unconstrained optimization is also given.

Due to BB method's simplicity, low storage and numerical efficiency, the Barzilai and Borwein method has received a good deal of attention in the optimization community but despite all these advances, stepsize of BB method is computed by means of simple approximation of Hessian in the form of scalar multiple of identity and especially the BB method is not monotone, and it is not easy to generalize the method to general nonlinear functions. Due to the presence of these deficiencies, we introduce new gradient-type methods in the frame of BB method including a new

gradient method via weak secant equation (quasi-Cauchy relation), improved Hessian approximation and scaling the diagonal updating.

The proposed methods are a kind of fixed step gradient method like that of Barzilai and Borwein method. In contrast with the Barzilai and Borwein approach's in which stepsize is computed by means of simple approximation of the Hessian in the form of scalar multiple of identity, the proposed methods consider approximation of Hessian in diagonal matrix. Incorporate with monotone strategies, the resulting algorithms belong to the class of monotone gradient methods with globally convergence. Numerical results suggest that for non-quadratic minimization problem, the new methods clearly outperform the Barzilai- Borwein method.

Finally we comment on some achievement in our researches. Possible extensions are also given to conclude this thesis.

Abstrak tesis untuk dibentangkan kepada Senat Universiti Putra Malaysia bagi memenuhi syarat Ijazah Doktor Falsafah.

**GLOBALISASI KAEDAH BARZILAI DAN BORWEIN BAGI PENGOPTIMUMAN TAK BERKEKANGAN**

Oleh

**MAHBOUBEH FARID**

**Disember 2009**

**Pengerusi: Professor Malik Hj. Abu Hassan, PhD**

**Fakulti: Institute Penyelidikan Matematik**

Fokus tesis ini adalah untuk mencari suatu minimum tak berkekangan bagi sesuatu fungsi. Secara khusus, kami akan memberi tumpuan kepada kaedah Barzilai dan Borwein (BB) yang terkemuka iaitu kaedah kecerunan saiz langkah dua titik. Pertama sekali, kami membincangkan tentang latar belakang matematik secara ringkas. Kemudian, kami menumpukan perbincangan kepada kaedah BB yang memainkan peranan penting dalam bidang pengoptimuman. Suatu sorotan tentang kaedah peminimuman semasa bagi menyelesaikan masalah pengoptimuman tak berkekangan juga diberi.

Disebabkan oleh kemudahan, storan rendah dan kecekapan berangka kaedah BB, ia telah menerima perhatian dalam komuniti pengoptimuman tetapi walaupun dengan kemajuannya, saiz langkah bagi kaedah BB dikira secara penganggaran mudah Hessan dalam bentuk gandaan skalar matriks identiti dan terutamanya kaedah BB tidak ekanada, dan ia tidak mudah diitlakan ini kepada fungsi tak linear am.

Disebabkan oleh kehadiran kelemahan-kelemahan tersebut, kami memperkenalkan kaedah kecerunan baru dalam rangka kaedah BB termasuk suatu kaedah kecerunan baru melalui persamaan sekan lemah (perhubungan kuasi-Cauchy), menambahbaik penghampiran Hessan dan menskalar pengemaskinian pepenjuru.

Kaedah yang dicadangkan adalah suatu jenis kaedah kecerunan langkah tetap sepertimana kaedah Barzilai dan Borwein. Berbeza dari pendekatan Barzilai dan Borwein di mana saiz langkah dikira melalui penghampiran mudah Hessan dalam bentuk gandaan skalar matriks identiti, kaedah yang dicadangkan mempertimbangkan penghampiran Hessan dalam bentuk matriks pepenjuru. Bergabung dengan strategi ekanada pada setiap lelaran, algoritma yang terhasil adalah anggota kepada kelas kaedah kecerunan ekanada dengan penumpuan sejagat. Keputusan berangka mencadangkan bahawa untuk masalah peminimuman bukan kuadratik, kaedah baharu yang dicadangkan secara jelasnya adalah lebih baik daripada kaedah Barzilai-Borwein.

Akhir sekali, kami mengulas tentang pencapaian dalam penyelidikan kami. Perlanjutan yang mungkin juga diberi bagi mengakhiri tesis ini.

## ACKNOWLEDGEMENTS

First and foremost I offer my sincerest gratitude to my chairman, Professor Dr. Malik Hj. Abu Hassan who has supported me throughout my thesis from the initial to the final level with his patience and knowledge whilst allowing me the room to work in my own way. I would like to express my deep and sincere gratitude to my co-supervisor, Dr. Leong Wah June. His wide knowledge and his logical way of thinking have been of great value for me. His encouraging, detailed and constructive comments have enabled me to develop an understanding of the subject. I am also grateful to Dr. Mansor Monsi for serving in the supervisory committee.

I offer my regards to the Head of Institute, academic and general staffs of the Institute for Mathematical Research, University Putra Malaysia, who supported me in any respect during the completion of the project. The financial support of the University Putra Malaysia under Special Graduate Research is gratefully acknowledged.

I owe my loving thanks to my parents who have lost a lot due to my research abroad. Without their encouragement, understanding and support it would have been impossible for me to finish this work. My special gratitude is due to my brother, for his loving support and personal guidance.

## APPROVAL SHEET 1

I certify that Examination Committee has met on date of viva voce to conduct the final examination of Mahboubeh Farid on her Degree of Doctor of Philosophy thesis entitled " Globalization of Barzilai and Borwein Method 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 Degree of Doctor of Philosophy.

Member of the Examination committee were as follows

**ISA DAUD, PhD**
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

**DATO' MOHAMED SULEIMAN, PhD**
Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

**NORIHAN MD. ARIFFIN, PhD**
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

**ISMAIL MOHD, PhD**
Professor
Department of Mathematics, Faculty of Science and Technology
Universiti Malaysia Terengganu
Malaysia
(External Examiner)

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

Date:

**APPROVAL SHEET 2**

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 Supervisory committee were as follows:

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

**LEONG WAH JUNE, PhD**
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: 11 February 2010

**DECLARATION**

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

MAHBOUBEH FARID

Date: 17 February 2010

## TABLE OF CONTENETS

# LIST OF TABLES

**LIST OF FIGURES**

## LIST OF NOTATIONS

1. $R^n$ denotes the linear $n-$dimensional Real space.

2. $g$ is the $n \times 1$ gradient vector of function $f$, with components

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

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

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

4. $x_k$ is the $k$ th approximation to $x^*$, a minimum of $f$.

5. $g_k$ is the gradient vector of $f$ at $x_k$.

6. $D_k$ is an $n \times n$ $k$ th diagonal matrix approximation to $G$.

7. $U_k$ is an $n \times n$ $k$ th diagonal matrix approximation to $G^{-1}$.

8. $B_k$ is an $n \times n$ $k$ th matrix approximation to $G$.

9. $A^T$ denotes the transpose of matrix $A$.

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

11. min denotes the minimum.

# CHAPTER 1

# INTRODUCTION

## 1.1 Introduction

Optimization Theory and Methods is a young subject in applied mathematics, computational mathematics and operations research which has wide applications in science, engineering, business management, military and space technology. The subject is involved in optimal solution of problems which are defined mathematically, i.e., given a practical problem, the ‘‘best’’ solution to the problem can be found from many schemes by means of scientific methods and tools.

## 1.2 General Form of Optimization Problems

The general form of optimization problems is

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

$$\text{s.t} \quad x \in R^n$$

where $x \in R^n$ is a decision variable, $f(x)$ an objective function, $X \subset R^n$ a constraint set or feasible region. Particularly, if the constraint set $X = R^n$, the optimization problem (1.1) is called an unconstrained optimization problem:

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

The constrained optimization problem can be written as follows:

$$\min_{x \in R^n} f(x)$$

$$\text{s.t.}\quad c_i(x) = 0, \quad i \in E, \tag{1.3}$$

$$c_i(x) \geq 0, \quad i \in I,$$

where $E$ and $I$ are, respectively, the index set of equality constraints and inequality constraints, $c_i(x)$, $(i = 1,\ldots,m \in E \cup I)$ are constrained functions. When the objective function and constrained functions are linear functions, the problem is called linear programming. Otherwise, the problem is called nonlinear programming.

**Definition 1.1.** A point $x^*$ is a *global minimizer* if $f(x^*) \leq f(x)$ for all $x$, where $x$ ranges over all of $R^n$.

**Definition 1.2.** A point $x^*$ is a *local minimizer* if there is neighbourhood $N$ of $x^*$ such that $f(x^*) \leq f(x)$ for $x \in N$.

This thesis studies solving unconstrained optimization problem (1.2) from the view points of both theory and numerical methods where $f$ is continuously differentiable function and a local minimizer provides a satisfactory solution.

Additional information about this topic can be found in Nocedal and Wright (1999) and Sun and Yuan (2006).

## 1.3 Function and Differential

A continuous function $f : R^n \to R$ is said to be continuously differentiable at $x \in R^n$, if $(\frac{\partial f(x)}{\partial x_i})$ exists and is continuous, $i = 1,2,\ldots,n$. The gradient of $f$ at $x$ is defined as

$$\nabla f(x) = [\frac{\partial f(x)}{\partial x_1},\ldots,\frac{\partial f(x)}{\partial x_n}]^T. \tag{1.4}$$

If $f$ is continuously differentiable at every point of an open set $D \subset R^n$, then $f$ is said to be continuously differentiable on $D$ and is denoted by $f \in C^1(D)$. A continuously differentiable function $f : R^n \to R$ is called twice continuously differentiable at $x \in R^n$, if $(\frac{\partial^2 f(x)}{\partial x_i \partial x_j})$ exists and is continuous, $i = 1,2,...,n$. The Hessian of $f$ is defined as the $n \times n$ symmetric matrix with elements

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

If $f$ is twice continuously differentiable at every point of an open set $D \subset R^n$, then $f$ is said to be twice continuously differentiable on $D$ and is denoted by $f \in C^2(D)$.

Let $f : R^n \to R$ be continuously differentiable on an open set $D \subset R^n$. Then for $x \in D$ and $d \in R^n$, the directional derivative of $f$ at $x$ in the direction $d$ is defined as

$$f'(x,d) = \lim_{\theta \to 0} \frac{f(x+\theta d) - f(x)}{\theta} = \nabla f(x)^T d, \tag{1.5}$$

where $\nabla f(x)$ is the gradient of $f$ at $x$, an $n \times 1$ vector.

For any $x, x+d \in D,$ if $f \in C^1(D),$ then

$$f(x+d) = f(x) + \int_0^1 \nabla f(x+td)^T d \mathrm{d}t$$

$$= f(x) + \int_x^{x+d} \nabla f(\xi) \mathrm{d}\xi. \tag{1.6}$$

Thus,

$$f(x+d) = f(x) + \nabla f(\xi) d, \quad \xi \in (x, x+d). \tag{1.7}$$

Similarly, for all $x, y \in D$, we have

$$f(y) = f(x) + \nabla f(x + t(y-x))^T (y-x), \quad t \in (0,1), \tag{1.8}$$

or

$$f(y) = f(x) + \nabla f(x)^T (y-x) + o(\|y-x\|). \tag{1.9}$$

It follows from (1.8) that

$$| f(y) - f(x) | \le \|y-x\| \sup_{\xi \in L(x,y)} \|f'(\xi)\|, \tag{1.10}$$

where $L(x,y)$ denotes the line segment with endpoint $x$ and $y$.

Let $f \in C^2(D)$. For any $x \in D,\ d \in R^n$, the second directional derivative of $f$ at $x$ in direction $d$ is defined as

$$f''(x,d) = \lim_{\theta \to 0} \frac{f'(x+\theta d, d) - f'(x,d)}{\theta}, \tag{1.11}$$

which is equal to $d^T \nabla^2 f(x) d$, where $\nabla^2 f(x)$ denotes the Hessian of $f$ at $x$. For any $x, \quad x + d \in D$, there exists $\zeta \in (x, x+d)$ such that

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

or

$$f(x+d) = f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d + o(\|d\|^2). \tag{1.13}$$

Let $h: R^n \to R,\ g: R^m \to R,\ f: R^n \to R^m$. Let $f \in C^1, g \in C^1,\ h(x_0) = g(f(x_0))$. Then the chain rule is

$$h'(x_0) = g'(f(x_0)) f'(x_0), \tag{1.14}$$

where $f'(x_0) = [\frac{\partial f_i(x_0)}{\partial x^j}]_{m \times n}$ is an $m \times n$ matrix. Also

$$h''(x_0) = \nabla f(x_0)^T \nabla^2 g[f(x_0)] \nabla f(x_0) + \sum_{i=1}^{m} \frac{\partial g[f(x_0)]}{\partial f_i} [f_i(x_0)]''. \tag{1.15}$$

Next, we discuss the calculus of vector-valued functions.

A continuous function $F : R^n \to R^m$ is continuously differentiable at $x \in R^n$ if each component function $f_i (i = 1,\ldots, m)$ is continuously differentiable at $x$. The derivative $F'(x) \in R^{m\times n}$ of $F$ at $x$ is called the Jacobian matrix of $F$ at $x$,

$$F'(x) = J(x),$$

with components

$$[F'(x)]_{ij} = [J(x)]_{ij} = \frac{\partial f_i}{\partial x^j}(x), \quad i = 1,\ldots, m; \quad j = 1,\ldots, n.$$

If $F : R^n \to R^m$ is continuously differentiable in an open convex set $D \subset R^n$, then for any $x, \quad x + d \in D$, we have

$$F(x+d) - F(x) = \int_0^1 J(x+td)ddt = \int_x^{x+d} F'(\xi)\mathrm{d}\xi. \tag{1.16}$$

In many of our considerations, we shall single out the different types of continuities.

**Definition 1.3.** $F : D \subset R^n \to R^m$ is *Holder continuous* on $D$ if there exists constants $\gamma \geq 0$ and $p \in (0,1]$ so that for all $x.y \in D$,

$$\|F(y) - F(x)\| \leq \gamma \|y - x\|^p. \tag{1.17}$$

If $p = 1$ then $F$ is called *Lipschitz continuous* on $D$ and $\gamma$ is a *Lipschitz constant*.

## 1.4 Convex Set and Convex Function

Convex sets and convex functions play an important role in the study of optimization.

**Definition 1.4.** Let the set $S \subset R^n$. If for any $x_1, x_2 \in S$, we have $\alpha x_1 + (1-\alpha)x_2 \in S, \quad \forall \alpha \in [0,1]$, then $S$ is said to be *convex set*.

**Definition 1.5.** Let $S \subset R^n$ be a nonempty convex set. Let $f : S \subset R^n \to R$. If for any $x_1, x_2 \in S$ and all $\alpha \in (0,1)$, we have

$$f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2), \tag{1.18}$$

then $f$ is said to be *convex* on $S$. If the inequality (1.18) is strict inequality for all $x_1 \ne x_2$, i.e.,

$$f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha) f(x_2), \tag{1.19}$$

then $f$ is called a *strict convex function* on $S$. If there is constant $c > 0$ such that for any $x_1, x_2 \in S$,

$$f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2) - \frac{1}{2} c\alpha(1-\alpha)\|x_1 - x_2\|^2, \tag{1.20}$$

then $f$ is called a *uniformly (or strongly) convex function* on $S$.

If $-f$ is a convex (strictly convex, uniformly convex) function on $S$, then $f$ is said to be concave (strictly concave, uniformly concave) function. Next, we give some properties of convex functions.

**Theorem 1.1.** 1. Let $f$ be a convex function on convex set $S \subset R^n$ and $\alpha \ge 0$ is a real number , then $\alpha f$ is also a convex function on $S$.

2. Let $f_1, f_2$ be convex functions on convex set $S$ , then $f_1 + f_2$ is also a convex function on $S$.

3. Let $f_1, f_2, ..., f_m$ be convex function on a convex set $S$ and $\alpha_1, \alpha_2, ..., \alpha_m \ge 0$ are real numbers, then $\sum_{i=1}^{m} \alpha_i f_i$ is also a convex function on $S$.

**Proof.** We only prove the second statement. The others are similar.

Let $x_1, x_2 \in S$ and $0 < \alpha < 1$, then

$$f_1(\alpha x_1 + (1-\alpha)x_2) + f_2(\alpha x_1 + (1-\alpha)x_2)$$

$$\leq \alpha[f_1(x_1)+f_2(x_2)]+(1-\alpha)[f_1(x_2)+f_2(x_2)]. \qquad \square$$

Continuity is an important property of convex function. However, it is not sure that convex function whose domain is not open is continuous.

The following theorem shows that a convex function is continuous on an open convex set or the interior of its domain.

**Theorem 1.2.** Let $S \subset D$ be an open convex set and $f : D \subset R^n \to R$ be convex. Then $f$ is continuous on $S$.

**Proof.** Let $x_0$ be an arbitrary point in $S$. Since $S$ is an open convex set, we can find $n+1$ points $x_1,\ldots,x_{n+1} \in S$ such that the interior of the convex hull

$$C = \{x \mid x = \sum_{i=1}^{n+1} \alpha_i x_i, \alpha_i \geq 0, \sum_{i=1}^{n+1} \alpha_i = 1\}$$

is not empty and $x_0 \in \operatorname{int} C$.

Now let $\alpha = \max_{1 \leq i \leq n+1} f(x_i)$, then

$$f(x) = f(\sum_{i=1}^{n+1} \alpha_i x_i) \leq \sum_{i=1}^{n+1} \alpha_i f(x_i) \leq \alpha, \forall x \in C, \qquad (1.21)$$

so that $f$ is bounded over $C$. Also, since $x_0 \in \operatorname{int} C$, there is a $\delta > 0$ such that $B(x_0,\delta) \subset C$, where $B(x_0,\delta) = \{x \mid \|x - x_0\| \leq \delta\}$. Hence for arbitrary $h \in B(0,\delta)$ and $\lambda \in [0,1]$, we have

$$x_0 = \frac{1}{1+\lambda}(x_0+\lambda h) + \frac{\lambda}{1+\lambda}(x_0 - h). \qquad (1.22)$$

Since $f$ is convex on $C$, then

$$f(x_0) = \frac{1}{1+\lambda} f(x_0+\lambda h) + \frac{\lambda}{1+\lambda} f(x_0 - h). \qquad (1.23)$$

By (1.21) and (1.23), we have

$$f(x_0+\lambda h)-f(x_0)\geq \lambda(f(x_0)-f(x_0-h))\geq -\lambda(\alpha-f(x_0)). \tag{1.24}$$

On the other hand,

$$f(x_0+\lambda h)=f(\lambda(x_0+h)+(1-\lambda)x_0)\leq \lambda f(x_0+h)+(1-\lambda)f(x_0),$$

which is

$$f(x_0+\lambda h)-f(x_0)\leq \lambda(f(x_0+h)-f(x_0))\leq \lambda(\alpha-f(x_0)). \tag{1.25}$$

Therefore (1.24) and (1.25) give

$$|f(x_0+\lambda h)-f(x_0)|\leq \lambda|f(x_0)-\alpha|. \tag{1.26}$$

Now, for given $\varepsilon>0$, choose $\delta^{'}\leq\delta$ so that $\delta^{'}|f(x_0)-\alpha|\leq\varepsilon\delta$. Set $d=\lambda h$ with $\|h\|=\delta$, then $d\in B(0,\delta)$ and

$$|f(x_0+d)-f(x_0)|\leq\varepsilon. \qquad \square$$

If convex function is differentiable, we can describe the characterization of differential convex functions. The following theorem gives the first order characterization of differential convex functions.

**Theorem 1.3.** Let $S\subset R^n$ be a nonempty open convex set and let $f:S\subset R^n\to R$ be a differentiable function. Then $f$ is convex if and only if

$$f(y)\geq f(x)+\nabla f(x)^T(y-x),\ \ \forall x,y\in S. \tag{1.27}$$

Similarly, $f$ is strictly convex on $S$.

$$f(y)>f(x)+\nabla f(x)^T(y-x),\quad \forall x,y\in S, y\neq x. \tag{1.28}$$

Furthermore, $f$ is strongly (or uniformly) convex if and only if

$$f(y)\geq f(x)+\nabla f(x)^T(y-x)+\frac{1}{2}c\|y-x\|^2,\quad \forall x,y\in S, \tag{1.29}$$

where $c>0$ is a constant.

**Proof.** Necessity: Let $f(x)$ be a convex function, then for all $\alpha$ with $0<\alpha<1$,

$$f(\alpha y+(1-\alpha)x)\le \alpha f(y)+(1-\alpha)f(x).$$

Hence,

$$\frac{f(x+\alpha(y-x))-f(x)}{\alpha}\le f(y)-f(x).$$

Setting $\alpha\to 0$ yields

$$\nabla f(x)^T(y-x)\le f(y)-f(x).$$

Sufficiency: Assume that (1.27) holds. Choose any $x_1,x_2\in S$ and set $x=\alpha x_1+(1-\alpha)x_2, 0<\alpha<1.$ Then

$$f(x_1)\ge f(x)+\nabla f(x)^T(x_1-x),$$

$$f(x_2)\ge f(x)+\nabla f(x)^T(x_2-x).$$

Hence

$$\alpha f(x_1)+(1-\alpha)f(x_2)\ge f(x)+\nabla f(x)^T(\alpha x_1+(1-\alpha)x_2-x)$$

$$= f(\alpha x_1+(1-\alpha)x_2),$$

which indicates that $f(x)$ is a convex function.

Similarly, we can prove (1.28) and (1.29) by use of (1.27). For example, from the definition of strictly convex, we have

$$f(x+\alpha(y-x))-f(x)<\alpha(f(y)-f(x)).$$

Then, using (1.27) and the above inequality, we have

$$\langle\nabla f(x),\alpha(y-x)\rangle\le f(x+\alpha(y-x))-f(x)<\alpha(f(y)-f(x)),$$

which is the required (1.28).

To obtain (1.29), it is enough to apply (1.27) to the function $f-\frac{1}{2}c\|\,.\,\|^2.$ □

Definition 1.4 of convex function indicates that function value is below the chord,