GRADUATE SEMINAR REPORT

ON

RELAXATION METHODS IN NONLINEAR OPTIMIZATION PROBLEMS

(Submitted for Partial Fulfillment of M.Sc. Degree in Mathematics)

BY

ABIYOT AYELE

ADVISOR

Prof.Dr.R.DEUMLICH

SCHOOL OF GRADUATE STUDIES
ADDIS ABABA UNIVERSITY

JULY 2004

ADDIS ABABA

Preface

Widely used techniques of solving optimization problems are penalty and Lagrange methods. The methods indicate candidates for the solution depending on properties of objective function and a feasible set .In some conditions numerical comparisons among the candidates is the only way to determine the solution.

Relaxation method is an iterative method for approximating the solution of optimization problems numerically.

This seminar paper consist three chapters. The first chapter introduces the notion of relaxation process and explains the behavior of convex and strongly convex functions with respect to relaxation sequences. The second chapter is mainly about estimation of relaxation process of convex and strongly convex functions. Chapter three comprises different techniques of constructing relaxation sequences.

Finally, I would like to thank ***Prof. Dr. R. Deumlich***, my advisor, for his willing to provide materials and for many valuable discussions and suggestions with regard to various improvement of the paper.

Abiyot Ayele

Addis Ababa, Ethiopia
July 2004

Table of Content

Chapter 1

Introduction

1.1. The Notion of Relaxation Process:

Consider the problem of minimizing $\varphi : X \to \Re$. The procedure of constructing a sequence of points (x_k) is called a relaxation process if

$$x_k \in X \text{ and } \varphi(x_{k+1}) \le \varphi(x_k) \quad k \in N_0.$$

In what follows we shall always assume that the solution set of the original problem is non empty. When investigating convergency of relaxation process we shall also assume that (x_n) is an infinite sequence and consequently, it's all points lie out side of the solution set.

1.2.Auxilary Materials

We shall consider several auxiliary assertions as a necessary apparatus for investigating the convergence of a relaxation procure.

Definition 1.2.1: A differentiable function f on a set X is said to be of lass $C^{1,1}(X)$ iff there is a constant $L > 0$ such that for any $x, y \in X$ with $[x, y] \subset X$,

$$\|f'(x) - f'(y)\| \le L \, \| x - L \| \text{ is holds.}$$

Example 1.2.1.

Let $X := \Re^n$, $f: U \to \Re$, $U \subseteq \Re^n$ be convex

$$f(x) := \langle Ax, x \rangle + \langle p , x \rangle$$

where A is a symmetrical positive definite $n \times n$ matrix,

$$f'(x) = \sum_{n=1}^{n} \langle 2Ax + p, e_i \rangle e_i$$

$$\|f'(x) - f'(y)\| = \left\| \sum_{j=1}^{n} \langle 2Ax - p, e_i \rangle e_i - \sum_{j=1}^{n} \langle 2Ay + p, e_i \rangle e_i \right\|$$

$$= \left\| \sum_{j=1}^{n} \langle 2A(x-y)\, e_i \rangle e_i \right\|$$

$$= \sum_{j=1}^{n} \| \langle 2A(x-y), e_i \rangle e_i \|$$

$$= \sum_{j=1}^{n} | 2\langle A(x-y), e_i \rangle |$$

$$\leq \sum_{j=1}^{n} 2||A(x-y)||$$

$$\leq 2n||A||\ ||x - y||.$$

There fore $f \in C^{1,1}(X)$

Lemma 1.2.1: If $f \in C^{1,1}(X)$, then for any x, y elements of X such that $[x, y] \subset X$

$$f(x) - f(y) \geq \langle f'(x), x - y\rangle - \frac{L}{2}||x - y||^2, \text{ holds.}$$

Proof: Since $f \in C^{1,1}(X)$.

Let $F(t) = f(y + t(x - y))$, $t \in [0, 1]$. So F is differentiable, by mean value Theorem for a function F of one variable,

$$F(1) - F(0) = F'(t) \qquad \text{for some } t \in (0, 1),$$

this implies $f(x) - f(y) = \langle f'(y + t(x - y)), x - y\rangle$.

So $f(x) - f(y) = \int_0^1 f(x) - f(y)]dt = \int_0^1 \langle f'(y+t(x-y)), x-y\rangle \, dt$

$$= \langle f'(x), x - y\rangle + \int_0^1 \langle f'(y+t(x-y)) - f'(x), x-y\rangle dt$$

$$\geq \langle f'(x), x - y\rangle - \int_0^1 \|f'(y+t(x-y) - f'(x)\| \|x-y\| dt$$

By Couchy – Swartz inequality and by definition of class $C^{1,1}(x)$;

$$f(x) - f(y) \geq \langle f'(x), x - y\rangle - \int_0^1 L\|(y+t(x-y) - x\| \|x-y\| dt$$

$$= \langle f'(x), x - y\rangle - A \int_0^1 L|t-1| \|x-y\|^2 dt$$

$$= \langle f'(x), x - y\rangle - L||x - y|| \int_0^1 (1-t)dt$$

$$= \langle f'(x), x - y\rangle - \frac{L}{2} \|x-y\|.$$

Consider, $X \subseteq \Re^n$, $Y \subseteq \Re^n$ such that $Y \subseteq X$ and $X \neq Y$, Let $f: X \rightarrow \Re$ satisfying $f(y) = 0$ for all

$y \in Y$. Consider an arbitrary sequence $(x_k) \subset X$ such that $f(x_k) \xrightarrow{k \to \infty} 0$.

Given $\varepsilon > 0$, we introduce the set $U_\varepsilon := \{x \in X | d(x, Y) \leq \varepsilon\}$, were $d(x, Y)$ is the distance between the set Y and the paint x. i.e. $d(x, Y) = \min_{y \in Y} d(x, y)$. In our case $d(x, Y) = \min_{y \in Y} ||x - y||$.

Theorem 1.2.1: Let $X \subseteq \Re^n$, $Y \subseteq \Re^n$ such that $Y \subseteq X$ and $X \neq Y$. Let $f: X \subseteq \Re^n$, such that

1. $f(y) = 0$ for all $y \in Y$
2. there is an arbitrary sequence $\{x_k\} \subseteq X$ such that $f(x_k) \xrightarrow{k \to \infty} 0$
3. for any $\varepsilon > 0$ there is $\delta = \delta(\varepsilon) > 0$ such that $f(x_k) \geq \delta$, for all $x_k \in X \backslash U_\varepsilon$.

Then $d(x_k, Y) \xrightarrow{k \to \infty} 0$.

Proof:

From 3. of the supposition , for any $\varepsilon > 0$ there is $\delta = \delta(\varepsilon) > 0$ such that $f(x_k) \geq \delta$ whenever, $x_k \in X \backslash U_\varepsilon$.

It's contra positive is stated as ,

$x_k \in X \backslash U_\varepsilon$ whenever $f(x_k) < \delta$,

hence $d.(x_k, Y) \leq \varepsilon$ whenever $f(x_k) < \delta$ and also from 2. for $\delta(\varepsilon) > 0$, there is $k_0(\varepsilon) > 0$ such that $f(x_k) \leq \delta$, for all $k \geq k_0(\varepsilon)$.

Hence we see that $d(x_k, Y) \leq \varepsilon$, for all $k \geq k_0(\varepsilon)$ this shows $d(x_k, Y) \xrightarrow{k \to \infty} 0$.

Remark1.2.1: We consider the application of Theorem 1.2.1 to the study of convergence of relaxation for minimizing a function on a set X. Consider the set,

$X_0 := \{x \in X \mid h(x) \leq h(x)\}$ and the set $X^* \subset X_0$ of points where the necessary conditions for local minimum of the function h are fulfilled.

Assume we know the function f such that $f(x) > 0$ for all $x \in X_0 \backslash X^*$ and $f(x^*) = 0$, for all $x^* \in X^*$.

Let the inequality $h(x_k) - h(x_{k+1}) \geq f(x_k)$ for all $x_k \in X_0 \backslash X^*$ hence

$h(x_k) - h(x_{k+1}) \geq f(x_k) > 0$ $k \in N_0$ implies $h(x_0) > h(x_1) > \ldots > h(x_k) > h(x_{k+1}) > \ldots$.

Hence $\{h(x_k)\}$ is monotone and bounded sequence,

$\lim_{k \to \infty} [h(x_k) - h(x_{k+1}] = \lim_{k \to \infty} h(x_k) - \lim_{k \to \infty} h(x_{k+1}) = 0$

$h(x_k) - h(x_{k+1}) \geq f(x_k) > 0$

So, $\lim_{k \to \infty} f(x_k) = 0$.

If f satisfies condition 3. of Theorem 1.2.1, we have $d(x_k, X^*) = 0$, i.e. the sequence (x_k) converges to a local minima of the function h on X .

Theorem 1. 2.2: If

1) φ is convex function of class $C^{1,1}(\Re^n)$,
2) $X^* = \{x^* \mid \varphi^*(x^*) = 0\} \neq \varnothing$,
3) for any $\varepsilon > 0$ there is $\delta = \delta(\varepsilon) > 0$ such that $\|\varphi'(x)\| \geq \delta$ for all x such that $d(x, X^*) = \|x - p^*\| \geq \varepsilon$,

then, any sequence $(x_k) \subset X_0 = \{x \mid \varphi(x) \leq \varphi(x_0)\}$ satisfies $\|x_h - p_k^*\| \| \leq \eta < \infty$, $k \in N_0$, where p^* and p_k^* are best approximations of x and x_k with respect to X^*, respectively.

Proof: φ has a minimum value on $\Re^n$,

$$x^* \in X^* \text{ implies } \varphi'(x^*) = 0$$

Subgradient inequality imply, for all $x \in \Re^n$

$$\varphi(x) - \varphi(x^*) \geq \langle \varphi'(x^*), x - x^* \rangle = 0,$$

hence $\varphi(x) \geq \varphi(x^*)$, for all $x \in \Re^n$

Put $\varphi(x^*) := \varphi^*$ (The minimum valve of φ on $\Re^n$)

Let $C := \varphi(x_0) - \varphi^*$, for the sequence (x_k) C X_0 ,

$$\varphi(x_k) - \varphi^* \leq \varphi(x_0) - \varphi^* = c < \infty, \quad k \in N_0 ,$$

by the condition 3. of the Theorem, for an $\varepsilon > 0$ and index k such that

$$\|x_k - p_k^*\| \geq \varepsilon \text{ implies } \frac{\varepsilon}{\|x_k - p_k^*\|} \leq 1,$$

Let $y_k = \dfrac{\varepsilon}{\|x_1 - p_k^*\|} x_k + \left(1 - \dfrac{\varepsilon}{\|x_k - p_k^*\|}\right) p_k^*$.

Obviously $y_k \in [p_k^* - x_k]$.

$$d(y_k, X^*) = \|y_k - p_k^*\| = \left\| \left(\frac{\varepsilon}{\|x_k - p_k^*\|} x_k + \left(1 - \frac{\varepsilon}{\|x_h - p_k^*\|}\right) p_k^* - p_k^* \right) \right\|$$

$$= \left\| \left(\frac{\varepsilon}{\left\| x_k - p_k^* \right\|} \right) \left(x_h - p_k^* \right) \right\| = \varepsilon \quad ,$$

hence by he conditions 3. of the Theorem $\|\varphi'(y_k)\| \geq \delta(\varepsilon)$.

$$\varphi(y_k) - \varphi^* \geq \varphi(y_k) - \varphi(y_k - \frac{\delta}{L} \cdot \frac{\varphi'(y_k)}{\|\varphi'(y_k)\|}), \quad k \in N$$

Applying Lemma 1.2.1 to the right side of inequality we obtain,

$$\varphi(y_k) - \varphi^* \geq \left\langle \varphi'(y_k), \frac{\delta}{L} \frac{\varphi'(y_k)}{\|\varphi'(y_k)\|} \right\rangle - \frac{1}{2} \left\| \frac{\delta}{L} \frac{\varphi'(y_k)}{\|\varphi'(y_k)\|} \right\|^2$$

$$= \frac{\delta}{L} \underbrace{\|\varphi'(y_k)\|}_{\geq \delta} - \frac{\delta^2}{2L} \geq \frac{\delta^2}{L} - \frac{\delta^2}{2L} = \frac{\delta^2}{2L} > 0 \qquad \text{(a).}$$

Assume the Theorem is false. i.e. there is a sequence$(x_k) \subseteq X$ such that $d(x_k, X^*) \to \infty$.

Since $\varphi^* = \varphi\left(p_k^*\right)$ [as φ is convex and $p_k^* \in X^*$].

$$\varphi(y_k) \leq \frac{\varepsilon}{\left\| x_k - p_k^* \right\|} \varphi(x_k) + \left(1 - \frac{\varepsilon}{\left\| x_k - p_k^* \right\|} \right) \varphi\left(p_k^*\right)$$

$$= \varphi^* + \underbrace{(\varphi(x_k) - \varphi^*)}_{\leq c}, \frac{\varepsilon}{\left\| x_k - p_k^* \right\|}$$

$$\leq \varphi^* + \frac{\varepsilon.c}{\left\| x_k - p_k^* \right\|} \quad .$$

Hence $\lim\limits_{k \to \infty} \varphi(y_k) \leq \varphi^* + \lim\limits_{k \to \infty} \dfrac{\varepsilon.c}{\left\| x_k - p_k^* \right\|} = \varphi^* \quad ,$

which is a contradiction to the result in (a) of the proof. //

Remark 1.2.2:

The set X^* in the above Theorem is closed and convex.

i) convexity:

Let $x \in X^*$, $y \in X^*$, $z := \lambda x + (1 - \lambda)y \qquad \lambda \in (0,1)$

$\varphi(z) \leq \lambda\varphi(x) + \varphi(y)\ (1 - 1) = \varphi(x)$, since $\varphi(x) = \varphi(y)$

hence φ(z) + φ(y), which implies $z \in X^*$

so, convexity of X* follows

ii) Closeness:

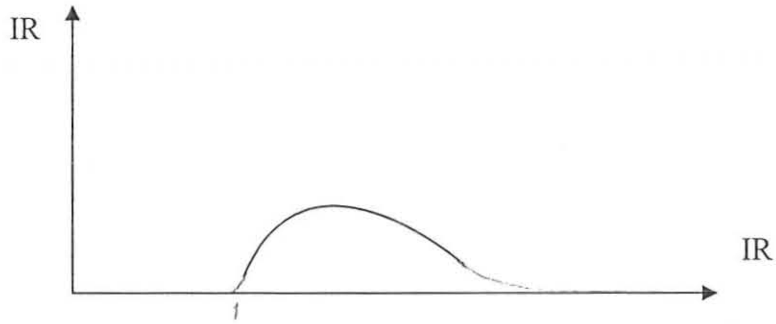
Let $\left(x_k^*\right) \subseteq X^*$ be a convergent sequence such that $x_h^* \to x$.

Now $\varphi(x) = \varphi\left(\lim_{k\to\infty} x_k\right) = \lim_{k\to\infty} \varphi(x_k) = \varphi^*$

hence $x \in X^*$, closeness of X*, follows,

The above theorem plays an important role in justifying the rate of convergence of minimization method for convex functions. Since these estimates are essentially based on conditions that $\|x_k - p_h^*\|$ is bounded.

Consider the graph of function of one-variable as shown below



graph

$X := [0, \infty)$, $Y; = [0, 1)$, $f: [0, \infty) \to \Re$ such that $f(y) = 0$, for all $y \in Y$.

Let (x_k) be a sequence in X such that $1 < x_k < x_{k+1}, k \in N$ and $f(x_k) \xrightarrow{k\leftarrow\infty} 0$, but $d(x_k, Y) = |x_k - 1| \xrightarrow{k\to\infty} \infty$,which violates the condition 3.. of Theorem 1.2.1 . Hence, bounded ness of X is material.

We now investigate the properties of the sets X and Y, and the function f that guarantee the fulfillment of the condition 3. of Theorem 1.2.1.

Theorem 1.2.3:

If 1. the set $X \subseteq \Re^n$ is bounded and closed,

2. $f: X \to \Re$ is continuous and $f(y) = 0$ for all $y \in Y$, $f(x) > 0$ for all $x \in X \backslash Y \neq \varnothing$,

then for $\varepsilon > 0$ there $\delta = \delta(\varepsilon) > 0$ such that $x_k \in X \backslash U_\varepsilon$ implies $f(x_k) \geq \delta$.

Proof:

Assume the theorem is false.

i.e. there is a sequence $(x_k) \subseteq X \backslash U_\varepsilon$ and an $\varepsilon > 0$ such that $f(x_k) \xrightarrow{k\to\infty} 0$.

Since $x_{k_i} \notin U_\varepsilon$ implies $d(x_{k_i}, Y) \geq \varepsilon$

So $d(x, Y) = d\left(\lim_{i\to\infty} x_{k_i}, Y\right) = \lim_{i\to\infty} d\left(x_{k_i} Y\right) \geq \varepsilon$,

hence $x \in X \backslash Y$. The continuity f implies

$$f(x_{k_i}) \xrightarrow{i\to\infty} f(x),$$

Since $(f(x_k))$ is convergent sequence, $f(x) = 0$, which implies $x \in Y$. Is a contradiction. //

Theorem 1.2.4: If the set Y ($Y \subset X$) of accumulation points of a sequence $(x_k) \subset X$ is finite ($Y = \{y_1, y_2, \ldots, y_m\}$), $\lim_{k\to\infty} \|x_k - x_{k+1}\| = 0$ and $\lim_{k\to\infty} d(x_k, Y) = 0$, then the sequence is convergent

Proof: $U_\delta(y) := \{x \in X \mid d(x, Y) \leq \delta\}$

Case 1. $m = 1$, i.e. $Y = \{y\}$

$$d(x_k, Y) = \|x_k - y\| \xrightarrow{k\to\infty} 0$$

hence $\lim_{k\to\infty} x_k = y$, so (x_k) is convergent.

Case 2. $m > 1$, Let $\delta = \min_{i\neq j} \|y_i - y_j\| > 0$, i, j = 1, 2, .., m from the condition

$d(x_k, Y) \xrightarrow{k\to\infty} 0$, we have an index $k_o(\delta) > 0$, such that $d(x_k, Y) \leq \frac{\delta}{4}$, for all $k \geq k_0$,

which implies $x_k \in U_{\frac{\delta}{4}}(y_i) = \{x \in X \mid d(x, y_i) \leq \frac{\delta}{4}\}$ for some $i \in \{1, 2, \ldots, m\}$ and hence

$x_k \in \bigcup U_{\frac{\delta}{4}}(y_i)\ \forall k \geq k_0(\delta_0)$.

From the condition $\|x_k - x_{k+1}\| \xrightarrow{k\to\infty} 0$ for $\frac{\delta}{4} > 0$ there is $k_0 > 0$ such that

$\|x_k - x_{k+1}\| < \frac{\delta}{4}$, for all $k \geq k'_0$.

WOLG $k_0 = \max \{k'_0, k_0\}$.

Since y_l is the accumulation point of the sequence , there is a subsequence (x_{k_i}) which converges to y_1. So, there is $x_{k_1} \in U_{\frac{\delta}{4}}(y_l)$ for some $k_1 \geq k_0$.

For any $j \in \{2, 3, \ldots, m\}$

$$\|y_j - x_{k_{i+1}}\| = \|(y_j - y_1) + (y_1 - x_{k_i}) + (x_{k_i} - x_{k_i} + 1)\|$$

$$\geq \|y_j - y_1\| - \|(y_1 - x_{k_i}) + (x_{k_i} - x_{k_i} + 1)\|$$

$$\geq \underbrace{\|y_i - y_1\|}_{\geq \delta} \underbrace{\|y_i - x_{k_i}\|}_{\leq \delta/4} \underbrace{\|x_{k_i} - x_{k_i} + 1\|}_{\leq \delta/4}$$

$$\geq \delta - \frac{\delta}{4} - \frac{\delta}{4} = \frac{\delta}{2}.$$

There for $\|y_j - x_{k_i} + 1\| \geq \frac{\delta}{2} > \frac{\delta}{4}$ for all $j \in \{2, 3, \ldots, m\}$.

Hence $x_{k_i} + 1 \notin U_{\frac{\delta}{4}}(y_j)$ $j \in \{2, 3, \ldots m\}$, and for all $k \geq k_0(\delta)$ $x_k \in U_{\frac{\delta}{4}}(y_l)$ for all $k_i \geq k_0$.

$x_k \in U_{\frac{\delta}{4}}(y_l)$ for all $k \geq k_0$ (by induction on k), hence $x_k \xrightarrow{k \to \infty} y_1$. //

Remark 1.2.3.

Observe the following property of convex function of real variable, which is differentiable.

Let $f: \Re \to \Re$, and $t_1 \in \Re$, $t_2 \in \Re$.

Since f is convex it satisfies subgradiant inequality,

$$\langle f'(t_1), t_2 - t_1 \rangle \leq f(t_2) - f(t_1) \qquad (1)$$

$$\langle -f'(t_2), t_2 - t_1 \rangle \leq f(t_1) - f(t_2) \qquad (2)$$

adding eq (1) and eq (2) we have

$$\langle f'(t_1) - f'(t_2), t_2 - t_1 \rangle \leq 0$$

Implies $\quad (f'(t_1) - f'(t_2))(t_2 - t_1) \leq 0$, since $t_1 < t_2$, $f'(t_1) \leq f'(t_2)$.

Theorem 1.2. 5. If f is convex and differentiable function on a convex set X and the set $Y := M(f, X)$ is bounded, then for all $\varepsilon > 0$ there is $\delta = \delta(\varepsilon) > 0$: $\forall x \in X_0 \setminus U_\varepsilon(y)$: $\|f'(x)\| \geq \delta$, where $U_\varepsilon(Y) = \{x | d(x, Y) \leq \varepsilon\}$ and $M(f,X) := \{x \mid f(x) \leq f(y) \forall y \in X\}$.

Proof: Let G be a boundary of $U_\varepsilon(Y)$.

Let $G_1 = G \cap X$.

(a) G_1 is bounded and closed set,

(i) To, show whether G_1 is bounded:

Let $x \in G_1$. Implies $x \in G$, hence $d(x, y) \leq \varepsilon$, for some $y^* \in Y$.

Since Y is bounded $\|x\| - \|y^*\| \leq \|x - y^*\| \leq \varepsilon$, this shows $\|x\| \leq \varepsilon + \|y^*\| \leq \varepsilon + M_0$, ($M_0$ is the bound of Y) so, G_1 is bounded.

(ii) To show whether G_1 is closed:

Let (x_k) be a sequence in G_1 such that $x_k \to x$.

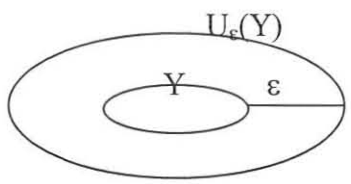
$d(x, Y) = d\left(\lim_{k\to\infty} x_k, Y\right) = \lim_{k\to\infty} d(x_k, Y) = \varepsilon$. Implies $x \in G_1$, hence G_1 is closed.

Since G_1 is closed and bounded [i.e. compact] and f is continuous, f possesses minimum on G_1. For arbitrary fixed $y \in Y$,

$$\xi := \min_{z\in G_1} (f(z) - f(y)) > 0 \ .$$

Let $x \in X\backslash U(Y)$ and Let $y \in Y$

$U_\varepsilon(Y)$



$d(x, G_1) = \|z - x\|$, for some $z \in G_1$.

$\|y - z\| \leq \| x - y\|$, since $Y \subseteq U_\varepsilon(Y)$ and $y \in Y$.

Implies $0 < \dfrac{\|y-z\|}{\|x-y\|} \leq 1$.

Let $z' := \dfrac{\|y-x\|}{\|x-y\|} x + \left(1 - \dfrac{\|y-x\|}{\|x-y\|}\right) y$.

$$d(z', Y) = \min_{y\in Y} \|z' - y\|$$

$$= \min_{y \in Y} \left\| \left(\frac{\|y-z\|}{\|x-y\|} x + \left(1 - \frac{\|y-z\|}{\|x-y\|} \right) y \right) - y \right\|$$

$$= \min_{y \in Y} \left\| \left(1 - \frac{\|y-z\|}{\|x-y\|} \right) (x-y) \right\|$$

$$= \min_{y \in Y} \|y - z\| = \varepsilon .$$

There for $z' \in G$.

So we have $z \in G \cap [y, x]$ for all $x \in X \backslash U_\varepsilon(Y)$ and $y \in Y$. Let $\|z - y\| = \eta < \infty$, since $X \cap U_\varepsilon(Y)$ is bounded and f is convex,

$F(\lambda) := f(z + \lambda(z - y))$ is convex and differentiable hence,

$\lambda_1 < \lambda_2$ implies $F'(\lambda_1) \le F'(\lambda_2)$ (Remark 1.2.3), which implies

$\langle f'(z + \lambda_1(z - y)), (z - y) \rangle \le f'(z + \lambda(z-y))\rangle$.

If $\lambda_1 = 0 < \lambda < 1$, we have $\langle f'(z), (z - y)\rangle \le \langle f'(z + \lambda(x - y)), z - y\rangle$, hence

$x = z + \lambda(z - y)$. Since $z \in [y, x]$,

$$\xi \le f(z) - f(y) \le \langle f'(z), z - y \rangle \le \langle f'(x), z - y\rangle$$
$$\le \|f'(x)\| \, \|z - y\| \le \eta \|f'(x)\| .$$

Implies $\frac{\xi}{\eta} \le \|f'(x)\|$,there forchoose $\delta = \frac{\xi}{\eta}$. //

Assume that for a relaxation process in Theorem 1.2.2 it is proved that $\lim_{k\to\infty} d(x_k, X^*) = 0$.

If the set X^* of points of local minima of the function φ is finite and the condition $\lim_{k\to\infty} \|x_k - y_{x+1}\| = 0$ is fulfilled, then Theorem1.2.5.implies the entire sequence converges to a point x^* of local minimum (depends on choice of x_0). Now we consider some conditions guaranteeing the fulfillment of the result $\|x_k - x_{k+1}\| \xrightarrow{k\to\infty} 0$.

Definition 1.2. 2: A function f called linearly non-constant on a set X if there are no two different point's $x \in X$, $y \in X$ such that $[x, y] \subset X$ and $f(z) = f(x)$ for all $z \in [x, y]$.

Example 1.2.2: Let $f: X \to \Re$, be strictly convex and $X \subseteq \Re^n$ be convex.

For any $x, y \in X$, $[x, y] \subseteq X$,

$z := x\lambda + (1 - \lambda)y, \lambda \in (0, 1]$

$f(z) = f(\lambda x+(1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$.

WOLG Assume $f(z) = f(y)$, so $f(z) \leq \lambda f(x) + (1-\lambda)f(z)$, implies $f(z) < f(x)$.

Hence f is linearly non-constant.

Example 1.2.2: $f: X \rightarrow \Re$ be a function satisfying the condition $f(z) < \max \{f(x), f(y)\}$ for all $z \in (x, y)$.

Definition1.2. 3: A sequence $(x_k) \subset X$ is said to be strongly lowering for a function f on the set X if $[x_k, x_{k+1}] \subset X$ and $f(x_k) \geq f(z) \geq f(x_{k+1})$, for all $z \in [x_k, x_{k+1}]$.

Theorem1.2.6: If a function f is continuous and linearly – non constant on a closed and bounded set X, then any strongly lowering sequence (x_k) satisfies the condition

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

Proof : Assume the Theorem is false, i.e. there is $\varepsilon > 0$ and a subsequence (x_{k_i}) of (x_k) such that

$$\| x_{k_{i+1}} - x_{k_i} \| \geq \varepsilon > 0, \text{ for all } i > 0$$

Since X is bounded (x_{k_i}) has a convergent subsequence (Bolzano-wearistrass theorem).

WOLG we take (x_{ki}) as it is a convergent subsequence and we have $\lim_{i\to\infty} x_{ki} =: x$ and $\lim_{i\to\infty} x_{ki+1} =: y$.

$$\|x - y\| = \| \lim_{i\to\infty} x_{ki} - \lim_{i\to\infty} x_{ki+1} \| = \lim_{i\to\infty} \| x_{ki} - x_{ki+1} \| \geq \varepsilon .$$

Since X is closed and $[x_{ki} - x_{ki+1}] \subset X$ implies $[x, y] \subset X$.

As f is continuous on X [compact -set] f has both minimum and maximum value and moreover (x_k) is strongly lowering sequence for f, hence $(f(x_k))$ is a monotone and bounded sequence.

$$\lim_{k\to\infty} (f(x_k) - f(x_{k+1})) = \lim_{k\to\infty} f(x_k) - \lim_{k\to\infty} f(x_{k+1}) = 0 .$$

Consequently for the subsequence $\left(f\left(x_{k_i}\right)\right)$,

$$\lim_{i\to\infty} (f(x_{ki}) - f(x_{ki+1})) = 0 .$$

$0 = \lim_{i\to\infty} (f(x_{ki}) - f(x_{k_{i+1}})) = f(\lim_{j\to\infty} x_{ki}) - f(\lim_{j\to\infty} x_{ki+1}) = f(x) - f(y)$ which contradicts the linearly – non constant ness of f on X .

Lemma1.2.2. If a sequence (μ_k) is such that $\mu_k - \mu_{k+1} \geq \tau_k \geq 0$, then

$$\mu_m \leq \mu_0 \left[1+\mu_0\sum_{k=0}^{m-1}\tau_k\right]^{-1}, \qquad m\in N.$$

Proof: $\mu_k > 0$, $\mu_k - \mu_{k+1} \geq \tau_k \mu_k^2$ for $k\in N$

$$\frac{\mu_k}{\mu_{k+1}} - 1 \geq \tau_k. \frac{\mu^2 k}{\mu_{k+1}} - 1$$

implies $\frac{\mu k}{\mu_{k+1}} - 1 \geq 1 + \tau_k \frac{\mu^2 k}{\mu_{k+1}} \geq 1$.

Therefore $\frac{\mu_k - \mu_{k+1}}{\mu_k\mu_{k+1}} = \frac{1}{\mu_{k+1}} - \frac{1}{\mu_k} \geq \frac{\tau_k \mu_{k^2}}{\mu_k\mu_{k+1}} = \tau_k \frac{\mu_k}{\mu_{k+1}} \geq \tau_k$.

Implies $\sum_{k=0}^{m-1}\left[\frac{1}{\mu_{k+1}} - \frac{1}{\mu_k}\right] \geq \sum_{k=0}^{m-1}\tau_k$,

implies $\frac{1}{\mu_m} - \frac{1}{\mu_0} \geq \sum_{k=0}^{m-1}\tau_k$

$$\frac{1}{\mu_m} \geq \frac{1}{\mu_0} + \sum_{k=0}^{m-1}\tau_k = \frac{1}{\mu_0}\left[1+\mu_0\sum_{k=0}^{m-1}\tau_k\right],$$

imlies $\frac{1}{\mu_m} \geq \frac{1}{\mu_0}\left[1+\mu_0\sum_{k=0}^{m-1}\tau_k\right]$.

Hence $\mu_m \leq \mu_0 \left[1+\mu_0\sum_{k=0}^{m-1}\tau_k\right]^{-1}$. //

Remark 1.2.4: h: =[0, 1) $\to\Re$ by h (x) = ln(1-x) + x, x $\in$[0,1)

$$h'(x) = \frac{-1}{1-x} + 1 = \frac{-x}{1-x}$$

x = 0 and x = 1 are critical points and also end paints of the interval .h attains maximum value at, x = 0 i.e. h(0) = 0 is the maximum value of h for all x $\in$(0, 1). There for ln (1 – x) $\leq$ - x, for all x $\in$[0, 1).

Lemma 1.2. 3: If a number sequence (μ_k) is such that $\mu_k - \mu_{k+1} \geq \tau_k\mu_k$, $\mu_k \geq 0$, then

$$\mu_m \leq \mu_0 \exp \left[-\sum_{k=0}^{m-1} \tau_h \right], m \in N .$$

Proof: $\mu_k - \mu_{k+1} \geq \tau_k\mu_k$ implies $1 - \dfrac{\mu_{k+1}}{\mu_k} \geq \tau_k$, $k \in N_0$.

Since $\mu_k \geq \mu_{k+1} > 0$, $1 \geq \dfrac{\mu_{k+1}}{\mu_k} > 0$ for all $k \in N_0$.

But $1 - \tau_k \geq \dfrac{\mu_{k+1}}{\mu_k} > 0$.

Therefore $0 < 1 - \tau_k \leq 1$ for all $k \in N_0$.

Implies $\mu_{k+1} \leq \mu_k - \tau_k\mu_k = (1 - \tau_k)\mu_k$, $k \in N_0$.

Hence $\mu_1 \leq (1 - \tau_0)\, \mu_0$

$$\mu_1 \leq (1 - \tau_1)\, \mu_1 \leq (1 - \tau_0)\, (1 - \tau_1)\, \mu_0$$

$$\vdots$$

$$\mu_m \leq (1 - \tau_0)\, (1 - \tau_1) \ldots (1 - \tau_{m-1})\, \mu_0 = \mu_0 \prod_{k=0}^{m-1} (1 - \tau_h)$$

$$= \mu_0 \exp \left\{ \ln \left[\prod_{k=0}^{m-1} (1 - \tau_h) \right] \right\}$$

$$= \mu_0 \exp \left\{ \sum_{k=0}^{m-1} ln(1 - \tau_h) \right\}$$

$$\leq \mu_0 \exp, \left\{ -\sum_{k=0}^{m-1} -\tau_h \right\} \; m \in N. \; //$$

Theorem1.2.7: Let $f: X \rightarrow \Re^n$ be convex and differentiable on a convex set X and Let the set $Y = N_\alpha(f)$ be bounded .

If a sequence $(x_k) \subset X$ is such that $\| f'(x_k) \| \xrightarrow{k\to\infty} 0$, then $\lim_{k\to\infty} d(x_k, Y) = 0$.

Proof:

(i) first we need to prove the bounded ness of (x_k). Assume it is not bounded, i.e. $||x_k|| \to \infty$. There is $\varepsilon > 0$, and an index k_0, such that $x_k \in X \backslash U_\varepsilon(Y)$.

To make clear;

Let M be a bound of Y

$||x_k - y|| \leq ||y|| + ||x_k|| \leq M + ||x_k|| =: \varepsilon$.

Since $||x_k|| \to \infty$, there is $k_0 > 0$: $\varepsilon < ||x_k||$, for all $k \geq k_0$, Hence $x_k \in X \backslash U_\varepsilon(Y)$, for all $k \geq k_0$.

There for $X \neq U_\varepsilon(Y)$.

By Theorem 1.2.5 there is $\delta = \delta(\varepsilon) > 0$: $\forall x_k \in X \backslash U_\varepsilon(Y) : ||f'(x_k)|| > \delta > 0 \ \forall_k \geq k_0$,

Which is a contradiction to the assumption, $||f'(x_k)|| \xrightarrow{k \to \infty} 0$.

Let p_k be the best approximation of x with respect to Y. Since both Y and (x_k) are bounded, there is $\eta > 0$ such that $||x_k - p_k|| \leq \eta < \infty$. As f is convex by subgradiant inequality,

$$0 \leq f(x_k) - f(y) = f(x_k) - f(p_k) \leq \langle f'(x_k), x_k - p_k \rangle$$
$$\leq ||f'(x_k)|| \, ||x_k - p_k|| \leq \eta \, ||f'(x_k)|| .$$

As $\|f'(x_k)\| \xrightarrow{k \to \infty} 0$,

implies $f(x_k) \xrightarrow{k \to \infty} f(y)$.

Since (x_k) is bounded, there is a convergent subsequence (x_{k_i})(Bolzano-Wearistrass Theorem) there for $f(x_{k_i}) \xrightarrow{i \to \infty} f(y)$. Since $(f(x_k))$ is a convergent sequence, more over

$f(x_k) \xrightarrow{k \to \infty} f(y)$ and is f continuous too,

$$\lim_{k \to \infty} d(x_k, Y) = 0. \ //$$

1.3 STRONGLY CONVEX FUNCTIONS

Definition 1.3.1 $\varphi := X \to \Re$ is said to be strongly convex if there is a constant $\rho>0$ such that for any x and y elements of X such that $[x,y]\subset X$ and any $\alpha\in[0,1]$,

$$\varphi(\alpha x + (1-\alpha)y) \leq \alpha\varphi(x) + (1-\alpha)\varphi(y) - \alpha(1-\alpha)\rho\|x-y\|^2 .$$

Example: 1.3.1

Let $\varphi : \Re^n \to \Re$

$\varphi(x) = \langle Ax, x\rangle + \langle p, x\rangle$, $p\in\Re^n$ and A is a positive definite symmetrical matrix .

Let $\lambda \in [0,1]$, $x\in\Re$, $y\in\Re$.

Now $\varphi(\lambda x + (1-\lambda)y) = \langle A(\lambda x + (1-\lambda)y), \lambda x + (1-\lambda)y\rangle + \langle p, \lambda x + (1-\lambda)y\rangle$

$= \lambda^2\langle Ax, x\rangle + 2\lambda(1-\lambda)\langle Ax, y\rangle + (1-\lambda)^2\langle Ay, y\rangle + \lambda\langle p, x\rangle + (1-\lambda)\langle p, y\rangle =$

$$[\lambda\langle Ax, x\rangle + \lambda\langle p, x\rangle] + [(1-\lambda)\langle Ay, y\rangle + (1-\lambda)\langle p, y\rangle] + \lambda^2\langle Ax, x\rangle - \lambda\langle Ax, x\rangle$$
$$+ (1-\lambda)^2\langle Ay, y\rangle - (1-\lambda)\langle Ay, y\rangle + 2\lambda(1-\lambda\langle Ax, y\rangle)$$
$$= \lambda\varphi(x) + (1-\lambda)\varphi(y) + \lambda(1-\lambda)\langle Ax, x\rangle + \lambda(\lambda-1)\langle Ay, y\rangle + 2\lambda(1-\lambda)\langle Ax, y\rangle$$
$$= \lambda\varphi(x) + (1-\lambda)\varphi(y) - \lambda(1-\lambda)\langle A(x-y), x-y\rangle .$$

Since A is appositive definite, there is $\rho > 0$ such that

$\langle A(x-y), x-y\rangle \geq \rho\|x-y\|^2$, implies

$$\varphi(\lambda x + (1-\lambda)y) \leq \lambda\varphi(x) + (1-\lambda)\varphi(x) - \lambda(1-\lambda)\rho\|x-y\|^2 .$$

Remark1.3.1: Strongly convex functions are strictly convex.

Lemma1.3.1: If 1. $\varphi := X \to \Re$ is strongly convex,

2. $X \subseteq \Re^n$ is closed and convex,

then for any point $y\in X$ the set $X_0 := \{x\in X \mid \varphi(x) \leq \varphi(y)\}$ is bounded.

Proof: Let $y\in X$ be arbitrary. Obviously φ is continuous as it is convex. Since φ is strongly convex there is $\rho \geq 0$ for any $x\in X$ and $y\in X$ such that,

$\varphi(\alpha x + (1-\alpha)y) \leq \alpha\varphi(x) + (1-\alpha)\varphi(y) - \alpha(1-\alpha)\|x-y\|^2$, for all $\alpha \in [0,1]$. By continuity of φ for $\rho > 0$, There is $\varepsilon > 0$, such that $|\varphi(x) - \varphi(y)| \leq \rho$, whenever $x\in U = \{x\in X / \|x-y\| < \varepsilon\}$.

Let $x \in X \setminus U$. Then $\|x-y\| > \varepsilon$, and hence $\alpha = \frac{\varepsilon}{\|x-y\|} < 1$. Strong convexity of φ implies $\alpha\varphi(x) \geq \varphi(y+\alpha(x-y)) - (1-\alpha)\varphi(y) + \alpha(1-\alpha)\rho\|x-y\|^2$.
(2)

$z := y + \alpha(x-y)$. So $\|x-y\| = \|y-\alpha(x-y)\| = \alpha\|x-y\| = \varepsilon$,

there for $z \in U$, and hence $\varphi(z) \geq \varphi(y) - \rho$. (3)

From (2) and (3) it follows that $\alpha\varphi(x) \geq -\rho + \alpha\varphi(y) + \alpha(1-\alpha)\rho\|x-y\|^2$.

Therefor $\varphi(x) \geq \varphi(y) + \rho\|x-y\|\left(\|x-y\| - \varepsilon + \frac{1}{\varepsilon}\right)$. *

If X_0 is unbounded, there is a sequence $(x_k) \subset X_0$ such that $\|x_k\| \xrightarrow{k\to\infty} \infty$.
Implies $\varphi(x_k) \xrightarrow{k\to\infty} \infty$, then we get $k_0(y) \in \mathbf{N}$ such that $\varphi(x_k) \geq \varphi(y)$ for $k \geq k_0(y)$.
This contradicts the assumption $x_k \in X_0$, for all $k \in \mathbf{N}$. //

Lemma1.3.2: If φ is strongly convex on a convex and closed set X,

then 1. for all $x \in X$, $\|x-x^*\|^2 \leq \frac{2}{\rho}(\varphi(x) - \varphi(x^*))$ is fulfilled,

2. if $\varphi \in C^1(X)$, then

i. for all x, $y \in X$, $\langle \varphi'(x) - \varphi'(y), x-y \rangle \geq \rho\|x-y\|^2$,

ii. $\|x-x^*\| \leq \frac{1}{\rho}\|\varphi'(x)\|$,

iii. $0 \leq \varphi(x) - \varphi(x^*) \leq \frac{1}{\rho}\|\varphi'(x)\|$,

Where $x^* \in M(\varphi, X)$.

Proof:

1. Strong convexity of φ and the assumption $x^* \in M(\varphi, X)$ imply

$$\varphi(x^*) \leq \varphi\left(\frac{x+x*}{2}\right) \leq \frac{1}{2}\varphi(x) - \varphi(x^*) - \frac{1}{4}\rho\|x-x*\|^2,$$

which implies $\|x-x^*\|^2 \leq \frac{2}{\rho}(\varphi(x) - \varphi(x^*))$

2. Strong convexity of φ imply, $\varphi\left(\frac{1}{2}(x-y)\right) \leq \frac{1}{2}\varphi(x) + \frac{1}{2}\varphi(y) - \frac{1}{4}\rho\|x-y\|^2$,

for some $\rho > 0$. Then $\frac{1}{4}\rho\|x-y\|^2 \leq \frac{1}{2}\varphi(x) + \frac{1}{2}\varphi(y) - \varphi\left(\frac{x+y}{2}\right)$. Hence

$$\frac{1}{2}\rho\|x-y\|^2 \le \varphi(x)+\varphi(y)-\varphi\left(\frac{x+y}{2}\right)$$

$$=\left[\varphi(x)-\varphi\left(\frac{x+y}{2}\right)\right]+\left[\varphi(y)-\varphi\left(\frac{x+y}{2}\right)\right].$$

From convexity and differentiability of φ it follows,

$$\frac{1}{2}\rho\|x-y\|^2 \le \left\langle \varphi'(x), \frac{x-y}{2}\right\rangle + \left\langle -\varphi'(y), \frac{x-y}{2}\right\rangle$$

$$=\frac{1}{2}\left\langle \varphi'(x)-\varphi'(y), \frac{x-y}{2}\right\rangle,$$

hence $\rho\|x-y\|^2 \le \langle \varphi'(x)-\varphi'(y), x-y\rangle.$

ii. From i. above $\langle \varphi'(x)-\varphi'(y)x-y\rangle \ge \rho\|x-y\|^2$ *.Since* $x^* \in X$ and $\varphi'(x^*)=0$,

$\left\langle \varphi'(x), x-x^*\right\rangle \ge \rho\left\|x-x^*\right\|^2$. By couch- Schwarz inequality,

$\rho\left\|x-x^*\right\|^2 \le \left\langle \varphi'(x), x-x^*\right\rangle \le \|\varphi'(x)\|\left\|x-x^*\right\|.$

By uniqueness of x*. $\|x-x\| \ne 0$.

Hence $\left\|x-x^*\right\| \le \frac{1}{\rho}\|\varphi'(x)\|$.

iii. $\varphi(x)-\varphi(x^*) \le \left\langle \varphi'(x), x-x^*\right\rangle \le \|\varphi'(x)\|\left\|x-x^*\right\|$. Combining with the result in ii.we get

$$0 \le \varphi(x)-\varphi(x^*) \le \frac{1}{\rho}\|\varphi'(x)\|. \;//$$

Definition 1.3.2: Let $\varphi : \Re^n \to \Re$. A function $\varphi^* := \Re^n \to \Re$ defined as

$\varphi^*(y) := \sup_{y\in\Re^n}\{\langle y,x\rangle - \varphi(y)\}$ is a conjugate of φ .

We will consider the following Theorem with out proof.

Theorem1.3.1: Le $\varphi : \Re^n \to \Re$ be convex and has a gradient mapping Lipschitzian with a constant $L > 0$ on $\Re^n$: For all $(x_1, x_2) \in \Re^n \times \Re^n$,

$\|\varphi'(x_1)-\varphi'(x_2)\| \le L\|x_1-x_2\|.$

Then φ^* is convex on each convex subset of domain of φ^*.

Example 1.3.2: $\varphi(x) = \|x\| + e^{-\|x\|}$, $x \in \Re^n$.

Now $\varphi'(x) = \frac{\left(1 - e^{-\|x\|}\right)}{\|x\|} x$, $x \in \Re^n$

$$\Rightarrow \quad \|\varphi'(x) - \varphi'(y)\| = \left[\sum_{i=1}^{n} \left(\frac{\left(1 - e^{-\|x\|}\right)}{\|x\|} \nu_i - \frac{\left(1 - e^{-\|y\|}\right)}{\|y\|} \upsilon_i \right)^2 \right]^{\frac{1}{2}}$$,$x := (\nu_1, \ldots, \nu_n)$ and $y := (\upsilon_1, \ldots, \upsilon_n)$.

Since $\forall x \in \Re^n \ \frac{1 - e^{-\|x\|}}{\|x\|} < 1$, $\left| \left(\frac{1 - e^{-\|x\|}}{\|x\|} \right) \nu_i - \left(\frac{1 - e^{-\|y\|}}{\|y\|} \right) \upsilon_i \right| \leq |\nu_i - \upsilon_i|$, $i \in \{1, \ldots, n\}$.

$$\text{Therefore } \|\varphi'(x) - \varphi'(y)\| = \left[\sum_{i=1}^{n} \left(\frac{\left(1 - e^{-\|x\|}\right)}{\|x\|} \nu_i - \frac{\left(1 - e^{-\|y\|}\right)}{\|y\|} \upsilon_i \right)^2 \right]^{\frac{1}{2}}$$

$$\leq \left[\sum_{i=1}^{n} |\nu_i - \upsilon_i|^2 \right]^{\frac{1}{2}}$$

$$= \|x - y\|.$$

Therefore $\varphi \in C^{1,1}(\Re^n)$.

Now $\varphi^*(y) = \sup_{x \in \Re^n} \left\{ \langle y, x \rangle - \|x\| - e^{-\|x\|} \right\}$, $y \in \Re^n$

Let h (x):$= \langle y, x \rangle - \|x\| - e^{-\|x\|}, x \in \Re^n$ and ($y \in \Re^n$ is arbitrary but fixed). So h is concave and differentiable .

$$h'(x) = y - \frac{1 - e^{-\|x\|}}{\|x\|} x$$

$$h'(x) = 0 \ \Rightarrow \ y = \frac{1 - e^{-\|x\|}}{\|x\|} x \tag{a}$$

$$\Rightarrow \ \langle \ y, x \ \rangle = \frac{1 - e^{-\|x\|}}{\|x\|} \langle \ x, x \ \rangle = \|x\| \left(1 - e^{-\|x\|}\right) \tag{b}$$

And $$\langle \ y, y \ \rangle = \frac{1 - e^{-\|x\|}}{\|x\|} \langle \ x, y \ \rangle = \left(1 - e^{-\|x\|}\right)^2 \tag{c}$$

h attains its maximum value for those x satisfying condition (a) above .Hence from condition (b) and (c) above and Theorem 1.3.1

$\varphi^*(y) = (\|y\| - 1)[1 - \ln(1 - \|y\|)]$ is strongly convex for $y \in \Re^n$ and $\|y\| < 1$

CHAPTER 2.

THEOREMS ON ESTIMATES

In this section we consider estimates for the rate of convergence of relaxation minimization process of a general type irrespective of their specific realization. It is important that the estimates are valid only for convex optimization problems.

Thus, through out the section it is assumed that the $\varphi : = X \rightarrow \Re$ and the feasible region X are convex.

Notations. $\varphi^* = \min_{x \in X} \varphi(x)$, $X^* = \{x \in X \mid \varphi(x) \leq \varphi(y) \forall y \in X\}$, $\mu_k = \varphi(x_k) - \varphi^*$,

$d(x_k, X^*) = \|x_k - p_k^*\|$, where p_k^* is the beast approximation of the point x_k on the set X^*.

The convexity of φ implies $0 < \mu_k \leq \langle \varphi'(x_k), x_k - p_k^* \rangle$, (2.1)

For all x_k: $\varphi(x_k) \neq \varphi^*$.

Theorem 2.1.If

1. $\varphi := X \rightarrow \Re$ is convex and differentiable,
2. X is convex and closed,
3. $X^* \neq \varnothing$,
4. (x_k) is a relaxation sequence,

then

$$\varphi(x) - \varphi^* \leq \mu_0 \left[1 + \mu_0 \sum_{k=0}^{m-1} \frac{\varphi(x_k) - \varphi(x_{k+1})}{\langle \varphi'(x_k), x_k - p_k^* \rangle} \right]^{-1} \quad \text{holds.} \tag{2.2}$$

Proof: Since $\mu_k = \varphi(x_k) - \varphi^*$ and $\mu_{k+1} = \varphi(x_{k+1}) - \varphi^*$,

$$\begin{aligned} \mu_k - \mu_{k+1} &= \varphi(x_k) - \varphi(x_{k+1}) \\ &= \frac{\varphi(x_k) - \varphi(x_{k+1})}{\langle \varphi'(x_k), x_k - p_k^* \rangle^2} \langle \varphi'(x), x_k - p_k^* \rangle \\ &\geq \frac{\varphi(x_k) - \varphi(x_{k+1})}{\langle \varphi'(x_k), x_k - p_k^* \rangle^2} \mu_k^2 . \end{aligned}$$

The last inequality above follows from condition (2.1). By Lemma1.2.2. and

$$\tau_k := \frac{\varphi(x_k)-\varphi(x_{k+1})}{\langle \varphi'(x_k), x_k - p_k^* \rangle^2}, \mu_m := \varphi(x_m)-\varphi^*$$

the conclusion of the Theorem follows. //

The estimate (2.2) is of non-constructive character, since its right-hand side involves the unknown projection on to X^*.If the inequality (2.2), then the relaxation process is weakly convergent (i.e. $\varphi(x_m) \xrightarrow{m\to\infty} \varphi^*$). In this case the strong convexity of φ guarantees the strong convergence of the process ($x_m \xrightarrow{m\to\infty} x^*$). To investigate the convergence of relaxation process using formula (2.2) one should be able to estimate from below the ratio,

$$\frac{\varphi(x_k)-\varphi(x_{k+1})}{\langle \varphi'(x_k), x_k - p_k^* \rangle^2} .$$

Theorem.2.3:

If 1. $X:=\Re$,

2.$\varphi:=X\to\Re^n$ is convex and differentiable,

3. $X_0 := \{x|\varphi(x) \le \varphi(x_0)\}$ is bounded and diam $X_0 = \eta < \infty$,

4. (x_k) is a relaxation sequence,

then

$$\varphi(x_m)-\varphi^* \le \mu_0\left[1+\mu_0 \frac{1}{\rho}\sum_{k=0}^{m-1}\frac{\varphi(x_k)-\varphi(x_{k+1})}{\|\varphi'(x_k)\|}\right]^{-1}, \; m\in N, \qquad (2.3)$$

where $\mu_0 = \varphi(x_0)-\varphi^*$.

Proof: Since φ is differentiable and p_k^*, $x_k \in X_0$ for each $k\in N$, implies $\langle \varphi'(x_k), x_k - p_k^* \rangle \le \|\varphi'(x_k)\|\|x_k - p_k^*\| \le \eta\|\varphi'(x_k)\|$.

Hence $\langle \varphi'(x_k), x_k - p_k^* \rangle^2 \le \|\varphi'(x_k)\|^2\eta^2$. Applying condition (2.2) to the last inequality we get the result in (2.3). //

Remark .2.1: If there is an index m_0 such that $\varphi(x_{m_0}) > \varphi(x_{m_0+1})$, implies

$$\sum_{k=0}^{m-1}\frac{\varphi(x_k)-\varphi(x_{k+1})}{\|\varphi'(x_k)\|^2} > 0 \quad , \; m = m_0+1, m_0+2, \dots .$$

the inequality in (3.3) becomes,

$$\varphi(x_k)-\varphi^* \le \frac{\mu_0^2}{\eta^2}\left[\sum_{k=0}^{m-1}\frac{\varphi(x_k)-\varphi(x_{k+1})}{\|\varphi'(x_k)\|^2}\right]^{-1}, m = m_0+1, m_0+2, \ldots .$$

For convex functions of class $C^{1,1}(\Re^n)$ the condition that "X_0 is bounded" can be weakened by replacing it by the conditions of Theorem 2.5.

Theorem 2.4:

If 1. $X := \Re^n$,

2. $\varphi \in C^{1,1}(\Re^n)$ is convex,

3. $X^* \neq \emptyset$,

4. for any $\varepsilon > 0$ there is $\delta = \delta(\varepsilon) > 0$ such that $\|\varphi'(x)\| > \delta$ for all x with $d(x, X^*) = \|x - p^*\| \ge \varepsilon$,

5. (x_k) is a relaxation sequence,

then the estimate (2.3) holds.

Proof:

Let $X_0 := N_\alpha(\varphi), \alpha := \varphi(x_0), X^* = \{x | \varphi(x) = 0\}$ and $\{x_k\} \subseteq X_0$.
Since X^* is convex and closed,

$$\forall x_k \in X_0 \exists p_k \in X^* : d(x_k, X^*) = \|x_k - p_k^*\| .$$

By Theorem1.2.2. there is $\eta \in \Re_+ \setminus \{0\}$ such that $\|x_k - p_k\| \le \eta$, for all $k \in N_0$.

$$\langle \varphi'(x_k), x_k - p_k \rangle \le \|\varphi'(x_k)\| \|x_k - p_k\| \le \eta \|\varphi'(x_k)\| \quad , k \in N_0$$

implies $\langle \varphi'(x_k), x_k - p_k \rangle^2 \le \eta^2 \|\varphi'(x_k)\|^2$.

The assertion of the Theorem follows from Theorem 2.2 . //

Theorem 2.5 :

If 1. $\varphi := \Re^n \to \Re$ is strongly convex and differentiable,

2. (x_k) is a relaxation sequence,

then

$$\varphi(x_m) - \varphi^* \le \mu_0 \exp\left[-\rho \sum_{k=0}^{m-1}\frac{\varphi(x_k)-\varphi(x_{k+1})}{\|\varphi'(x_k)\|^2}\right]^{-1} \quad , m \in N$$

and

$$\|x_m - x^*\|^2 \le \frac{2}{\rho}\mu_0 \exp\left[-\rho \sum_{k=0}^{m-1}\frac{\varphi(x_k)-\varphi(x_{k+1})}{\|\varphi'(x_k)\|^2}\right] \quad , m \in N \quad ,$$

where ρ is a parameter of strong convexity .

proof : $\mu_k := \varphi(x_k) - \varphi^* \geq 0 \quad , \tau_k := \rho \dfrac{\varphi(x_k) - \varphi(x_{k+1})}{\|\varphi'(x_k)\|^2} \geq 0$

implies $\mu_k - \mu_{k+1} = \varphi(x_k) - \varphi(x_{k+1}) = \dfrac{\varphi(x_k) - \varphi(x_{k+1})}{\|\varphi'(x_k)\|^2} \|\varphi'(x_k)\|^2$. (2.4)

Since φ is strongly convex , $\mu_k = \varphi(x_k) - \varphi^* \leq \dfrac{1}{\rho} \|\varphi'(x_k)\|^2$

holds by Lemma 1.3.2 . From condition (2.4) and the last inequality we get

$$\mu_k - \mu_{k+1} \geq \rho \frac{\varphi(x_k) - \varphi(x_{k+1})}{\|\varphi'(x_k)\|} \mu_k \quad .$$

(2.5)

Again by Lemma 1.2.3. $\|x_k - x^*\|^2 \leq \dfrac{2}{\rho} \mu_k$ .The conclusion of the theorem follows from Lemma .

The estimates above allow one to accumulate, during the computation process, information about the rate of convergence of $\varphi(x_m)$ to φ^* or in case of strong convexity x_m to x^* .We note that as long as the order of magnitude of the ratio

$$\frac{\varphi(x_k) - \varphi(x_{k+1})}{\|\varphi'(x_k)\|^2}$$

is not lower than o(1/k) ,the relaxation minimization process remains convergent .

CHAPTER 3

DESCENT METHODS

In chapter 2. and chapter 1. we have seen the general properties of functions, which allow the convergence of relaxation sequence and give an estimate for them. Here we will construct such sequences using descent techniques.

Thus we will consider the solution of the problem of minimizing a function $\varphi : \Re^n \to \Re$.
The descent method consists the following procedure of constructing the sequence (x_k)

(i) $x_o \in \Re^n$ is arbitrarily chosen.

(ii) x_k, $k \geq 1$ is constructed recursively as follows.

1. a direction of descent $-s_k$ at the point x_k is selected.
2. the $(k+1)^{th}$ approximation is found by the formulas

$$x_{k+1}=x_k-\beta_k s_k, \forall k \geq 0 \quad (3.1)$$

In relaxation methods the choice of $-s_k$ and the parameter β_k is so as to satisfies the inequality $\varphi(x_k) \geq \varphi(x_{k+1}), k \in N_0$. The parameter β_k usually called the step length or simply a step.

3.1. Choosing the step length. From the formula we can see β_k determines the distance between x_k and x_{k+1} .In this section we assume $-s_k$ as it was already selected. Now we will see three methods for selecting the step at each iteration.

3.1.1. Duplication method. The procedure of the method is as follows.

(i) we take $\beta_k=\beta_{k-1}$,if $\varphi(x_{k+1})< \varphi(x_k)$,we either pass to the next iteration s or $\beta_k=\beta_{k+1}$.if $\varphi(x_k)> \varphi(x_{k+1})$,the duplication process continues until the decrease stops.

(ii) If $\varphi(x_{k+1})\geq \varphi(x_k)$, then we take $\beta_k=\frac{1}{2}\beta_{k-1}$.

(iii) If $\varphi(x_k-\frac{1}{2}\beta_{k-1}s_k)< \varphi(x_k)$, $x_{k+1}:=x_k-\frac{1}{2}\beta_{k-1}s_k$ and pass to the next iteration.

(iv) If $\varphi(x_k-\frac{1}{2}\beta_{k-1}S_k)\geq \varphi(x_k)$, $\beta_k=\frac{1}{4}\beta_{k-1}$ and the process continues.

3.1.2:One dimensional optimization method for selecting a step.

$$\psi(\beta) := \varphi(x_k - \beta s_k), \beta \in [0,\infty).$$

Since $-s_k$ is a direction of descent of φ at x_k we expect $\beta_k \in M(\psi,(0,\infty))$. Hence $\beta_k>0$ and $x_{k+1}:= x_k-\beta_k s_k$.

We introduce a new parameter λ in order to strengthen the inequality $\varphi(x_{k+1})\leq \varphi(x_k)$ with respect to the chosen β_k as follows:

$$\varphi(x_k-\beta_k s_k)\leq (1-\lambda_k)\varphi(x_k)+\lambda_k\omega_k \quad (3.2)$$

where $\omega_k = \inf_{\beta\geq 0} \varphi(x_k - \beta s_k)$

and $\lambda_k \le \dfrac{\varphi(x_k)-\varphi(x_{k+1})}{\varphi(x_k)-\omega_k} \in (0,1]$.

So, condition 2 and theorem 3.3 imply

$$\varphi(x_m)-\varphi(x^*) \le \mu_o\left[1+\frac{\mu_o}{\eta^2}\sum_{k=0}^{n-1}\frac{\varphi(x_k)-\varphi(x_{k+1})}{\left\|\varphi'(x_k)\right\|^2}\right]^{-1}$$

$$\le \mu_o\left[1+\frac{\mu_o}{\eta^2}\sum_{k=0}^{n-1}\lambda_k\frac{\varphi(x_k)-\omega_k}{\left\|\varphi'(x_k)\right\|^2}\right]^{-1} .$$

The last inequality shows the increase or decrease of the parameter λ_k in the feasible manner affects the rate of convergence of the sequence $(\varphi(x_k))$ to the minimum value $\varphi(x^*)$.

3.1.3.Numerical implimentable condition for selecting the step (note in the sense of sec.3.1.2)

Let Q be a proposition such that

$$Q(\beta):=\varphi(x_k)-\varphi(x_k-\beta S_k)\ge q\beta\langle\ \varphi'(x_k),s_k\ \rangle, q>0, \beta>0. \tag{3.3}$$

As $-s_k$ is a direction of descent of φ at x_k, $\langle\ \varphi'(x_k),s_k\ \rangle>0$ and consequently $\varphi(x_k)-\varphi(x_k-\beta S_k)\ge 0$ for all $\beta\ge 0$, which satisfies the condition (3.3).

$$\beta_k := \max_{\beta\ge 0}\{\beta | Q(\beta)\} \quad ,$$

and hence $x_{k+1}:=x_k+\beta_k s_k$.

The question is that weather a real number $\beta\ge 0$, such that $Q(\beta)$ is true exists .The following corollary assures the existence of a positive number satisfying condition (3.3)

Corollary 3.1: Let $\varphi\in C^{1,1}(\Re^n)$,let $-s_k$ be the direction of descent of φ at x_k and $q\in(0,1)$, then

$$\overline{\beta}_k := 2(1-q)\frac{\langle\varphi'(x_k),s_k\rangle}{L\|s_k\|^2} \quad ,$$

satisfies condition (3.3) .

proof : Applying Lemma 3.(chap. 2.) obtain

$$\varphi(x_k)-\varphi(x_k-\overline{\beta}_k s_k)\ge\overline{\beta}_k\langle\varphi'(x_k),s_k\rangle-\frac{L}{2}\overline{\beta}_k^2\|s_k\|^2$$

$$=\overline{\beta}_k\left[\langle\varphi'(x_k),s_k\rangle-\frac{L}{2}\overline{\beta}_k\|s_k\|^2\right]$$

$$=q\,\overline{\beta}_k\langle\varphi'(x_k),s_k\rangle .$$

Moreover $\overline{\beta}_k>0$ and $q>0$. //

Condition (3.3) can be rewritten for certain $q_k \in (0,1)$ and the corresponding calculated β_k as follows:

$$\varphi(x_k)-\varphi(x_k-\beta_k s_k)\geq q_k\beta_k\langle\varphi'(x_k),s_k\rangle \quad .$$

(3.4)

By $\langle\varphi'(x_k),s_k\rangle > 0$ we get

$$\langle\varphi'(x_k),s_k\rangle=\cos\theta_k\,\|\varphi'(x_k)\|\|s_k\| \text{ , where } \quad 0\leq\theta_k\leq\frac{\pi}{2}.$$

$$\alpha_k:=\cos\theta_k=\frac{\langle\varphi'(x_k),s_k\rangle}{\|\varphi'(x_k)\|\|s_k\|} \quad .$$

Lemma 3.1:

If 1. $\varphi\in C^{1,1}(\Re^n)$ and $c_0\leq\varphi(x)$, $\forall x\in\Re^n$,

2. the sequence (x_k) constructed by the formula (3.1) and satisfies condition (3.2),

then $\varphi(x_k)-\varphi(x_{k+1})\geq\frac{1}{2L}\lambda_k\alpha_k^2\|\varphi'(x_k)\|^2$.

Proof : From condition (3.2) and the definition of infimum ,

$$\begin{aligned}\varphi(x_k)-\varphi(x_{k+1}) &\geq \lambda_k(\varphi(x_k)-\omega_k)\\ &=\lambda_k\left(\varphi(x_k)-\inf_{\beta\geq 0}\varphi(x_k-\beta s_k)\right)\\ &\geq\lambda_k(\varphi(x_k)-\varphi(x_k-\beta s_k)) \quad .\end{aligned}$$

By Lemma 1.2.3. and by definition of α_k ,

$$\begin{aligned}\varphi(x_k)-\varphi(x_{k+1}) &\geq \lambda_k(\varphi(x_k)-\varphi(x_k-\beta s_k))\\ &\geq\lambda_k\left(\beta\langle\varphi'(x_k),s_k\rangle-\frac{L}{2}\beta^2\|s_k\|^2\right)\\ &=\lambda_k\left(\beta\alpha_k\|\varphi'(x_k)\|\|s_k\|-\frac{L}{2}\beta^2\|s_k\|^2\right) .\end{aligned}$$

The inequality holds for all $\beta\in\Re_+$. Fortunately the right side expression is a quadratic form with respect to β with a leading coefficient $-\frac{L}{2}\|s_k\|^2$, also the inequality holds true for,

$\beta=\frac{\alpha_k}{L}\frac{\|\varphi'(x_k)\|}{\|s_k\|}$, where the quadratic expression attains its maximum value . There for the assertion of the theorem follows .

Lemma .3.4 :

If 1. $\varphi \in C^{1,1}(\Re^n)$ and $C_0 \le \varphi(x) \quad \forall x \in \Re^n$,

2. the sequence (x_k) is constructed by formula (3.1) and satisfies condition (3.4) ,

then $\varphi(x_k)-\varphi(x_{k+1}) \ge \frac{1}{L}\gamma_k \alpha_k{}^2 \|\varphi'(x_k)\|^2$, where $\gamma_k = 2q_k(1-q_k)$ and $q_k \in (0,1)$.

proof : Since β_k is the maximum number satisfying the inequality in (3.4) and by corollary.(3.1) , $\beta_k \ge \bar{\beta}_k = 2(1-q_k)\dfrac{\langle \varphi'(x_k), s_k \rangle}{L\|s_k\|^2}$.

$$\begin{aligned}
\varphi(x_k)-\varphi(x_{k+1}) &= \varphi(x_k)-\varphi(x_k-\beta_k s_k) \\
&\ge q_k \beta_k \langle \varphi'(x_k), s_k \rangle \\
&\ge q_k \bar{\beta}_k \langle \varphi'(x_k), s_k \rangle \\
&= \frac{1}{L}\gamma_k \alpha_k^2 \|\varphi'(x_k)\| \quad . \quad //
\end{aligned}$$

Let $\varphi : \Re^n \to \Re$ be continuously differentiable .

Defniton 3.1. The set of stationary points:

$X^* := \{x^* | \varphi'(x_k) = 0\}$ is called the set of stationary points of φ.

$X_0 := N_\alpha(\varphi)$, where $\alpha := \varphi(x_0)$ and $X_0^* := X^* \cap X_0$

Definition 3.2: Let $\varepsilon > 0$. $U_\varepsilon := \{x | d(x, X_0^*) \le \varepsilon\}$, where $d(x, X_0^*) = \inf\limits_{x^* \in X_0^*} \|x - x^*\|$.

From the point of view of Theorem1.2.1. and the assignment of $Y := X_0$, $X := X_0^*$ $f(x_k) := \|\varphi'(x_k)\|$, we observe for each $\varepsilon > 0$ there is $\delta = \delta(\varepsilon)$ such that $\|\varphi'(x_k)\| \ge \delta$ for all $x_k \in X_0^* \setminus U_\varepsilon$.

(3.5)

Theorem 3.1:

If 1. $\varphi \in C^{1,1}(\Re^n)$, $\varphi(x) \ge C \quad \forall x \in \Re^n$ and $X_0^* \ne \emptyset$,

2. the sequence (x_k) is constructed by formula (3.1) and satisfies either (3.2) for $0 < \lambda \le \lambda_k \le 1$, $k \in N_0$ or condition (3.4) for $0 < \varepsilon \le q_k \le 1-\varepsilon$, $\varepsilon < ½$ and $k \in N_0$,

3. $\alpha_k \ge \alpha > 0$, $k \in N_0$

4.conditon (3.5) is fulfilled,

then

$$\lim_{k \to \infty} d(x_k, X_0^*) = 0 .$$

proof : If (x_k) satisfies condition (3.2) , $\varphi(x_k - \beta_k s_k) \le (1-\lambda_k)\varphi(x_k) + \lambda_k \omega_k$ where $0 < \lambda \le \lambda_k \le 1$, k $\in N_0$. By Lemma (3.1)

$$\varphi(x_k) - \varphi(x_{k+1}) \ge \frac{1}{2L}\lambda_k \alpha_k \|\varphi'(x_k)\|^2$$

$$\ge \frac{1}{2L}\lambda\alpha^2 \|\varphi'(x_k)\|^2 . \qquad \text{(a)}$$

If (xk) satisfies condition (3.4) , by Lemma 3.2 :

$$\varphi(x_k) - \varphi(x_k - \beta_k s_k) \ge \frac{1}{L}\gamma_k \alpha_k{}^2 \left\|\varphi'(x_k)\right\|^2, \gamma_k = 2q_k(1-q_k), q_k \in (0,1)$$

$$\ge \gamma\alpha^2 \left\|\varphi'(x_k)\right\|^2 \qquad \text{(b)}$$

Since $0 < \varepsilon \le q_k \le 1-q$ and $\varepsilon < \frac{1}{2}$ it follows

$0 < \gamma < 2\varepsilon^2 \le \gamma_k \le \frac{1}{2}$, k=0,1,...

For condition (a) and (b) of the proof we obtain

$\varphi(x_k) - \varphi(x_{k+1}) \ge c \left\|\varphi'(x_k)\right\|^2$

Where c: $= \frac{\alpha^2}{L}\min\{\gamma, \frac{1}{2}\lambda\}$.

As $\{\varphi(x_k)\}$ is monotone decreasing and bounded,

$0 = \lim_{k\to\infty} (\varphi(x_k) - \varphi(x_{k+1})) \ge c \lim_{k\to\infty} \left\|\varphi'(x_k)\right\|^2 \ge 0$.

Hence

from $\|\varphi'(x_k)\|^2 \xrightarrow{k\to\infty} 0$ and supposition of the theorem the assertion follows. //

Remark 3.1: according to Theorem 3.1 , choice of the initial point x_o determines the stationary points that lie in X_0^* .Off course the choice is immaterial , if φ is convex.

Theorem 3.2:

Let

1. φ is convex and $\varphi \in C^{1,1}(R^n)$,
2. diam $X_o = \eta < \infty$,
3. the sequence (x_k) is constructed according to formula (3.1) ,
4. $\alpha_k > 0$, $k \in N_0$.

If condition (3.2) is fulfilled then

$$\varphi(x_m) - \varphi(x^*) \le \mu_o \left[1 + c\mu_o \sum_{k=0}^{m-1} \lambda_k \alpha_k{}^2 \right]^{-1},$$

where $0 < c \le \frac{1}{2L\eta^2}$ and $m \in N_0$.

If condition (3.4) holds, then

$$\varphi(x_m) - \varphi(x^*) \le \mu_o \left[1 + c\mu_o \sum_{k=0}^{m-1} \gamma_k \alpha_k{}^2 \right]^{-1},$$

where $0 < c \le \frac{1}{L\eta^2}$ and $m \in N$.

Proof: (I) if condition (3.2) is fulfilled, by Theorem 2. 3: and Lemma 3. 1: we obtain

$$\varphi(x_m) - \varphi^* \le \mu_o \left[1 + \frac{\mu_o}{\eta^2} \sum_{k=0}^{n-1} \frac{\varphi(x_k) - \varphi(x_{k+1})}{\left\|\varphi'(x_k)\right\|^2} \right]^{-1}, m \in N$$

$$\le \mu_o \left[1 + \mu_o \frac{1}{\eta^2} \sum_{k=0}^{m-1} \frac{1}{2L} \lambda_k \alpha_k{}^2 \right]^{-1}$$

$$= \mu_o \left[1 + \mu_o \left(\frac{1}{2L\eta^2}\right) \sum_{k=0}^{m-1} \lambda_k \alpha_k{}^2 \right], \ m \in N .$$

Let $c := \frac{1}{2L\eta^2}$ the inequality follows.

(II) if condition (3.2) is fulfilled ,by Theorem 3(chap.3) and Lemma 2 we have

$$\varphi(x_m) - \varphi(x^*) \le \mu_o \left[1 + \frac{\mu_o}{\eta^2} \sum_{k=0}^{n-1} \frac{\varphi(x_k) - \varphi(x_{k+1})}{\left\|\varphi'(x_k)\right\|^2} \right]^{-1}$$

$$\le \mu_o \left[1 + \mu_o \frac{1}{\eta^2} \sum_{k=0}^{m-1} \frac{1}{L} \gamma_k \alpha_k{}^2 \right]^{-1}$$

$$= \mu_o \left[1 + \mu_o \frac{1}{L\eta^2} \sum_{k=0}^{m-1} \gamma_k \alpha_k{}^2 \right]^{-1} \quad , m \in N$$

$c := \dfrac{1}{L\eta^2}$, hence the assertion follows. //

Theorem3.3:

Let

1. φ be strongly convex and $\varphi \in C^{1,1}(\Re^n)$,
2. the sequence $(x_k)$is constructed by formula (1) ,
3. $\alpha_k > 0$, k=0,1,2,….

If condition (3.2) holds, then

$$\varphi(x_m) - \varphi(x^*) \le \mu_o \exp\left[\frac{-\rho}{2L} \sum_{k=0}^{m-1} \lambda_k \alpha_k{}^2 \right], m \in N \quad \text{and}$$

$$\left\| x_m - x^* \right\|^2 \le \frac{2}{\rho} \mu_o \exp\left[\frac{-\rho}{2L} \sum_{k=0}^{m-1} \lambda_k \alpha_k{}^2 \right], m \in N .$$

If condition (3.4) holds, then

$$\varphi(x_m) - \varphi(x^*) \le \mu_o \exp\left[\frac{-\rho}{L} \sum_{k=0}^{m-1} \gamma_k \alpha_k{}^2 \right], m \in N \quad \text{and}$$

$$\left\| x_m - x^* \right\|^2 \le \frac{2}{\rho} \mu_o \exp\left[\frac{-\rho}{L} \sum_{k=0}^{m-1} \gamma_k \alpha_k{}^2 \right], \quad m \in N .$$

Proof: I. If condition (3.1) holds for the sequence (x_n), by Lemma 3.1 and Theorem 2.5

$$\varphi(x_m) - \varphi(x^*) \le \mu_o \exp\left[-\rho \sum_{k=0}^{n-1} \frac{\varphi(x_k) - \varphi(x_{k+1})}{\left\| \varphi'(x_k) \right\|^2} \right], m \in N$$

$$\le \mu_o \exp\left[-\rho \sum_{k=0}^{m-1} \frac{1}{2L} \lambda_k \alpha_k{}^2 \right], m \in N$$

$$= \mu_o \exp\left[\frac{-\rho}{2L} \sum_{k=0}^{m-1} \lambda_k \alpha_k{}^2 \right], m \in N .$$

and

$$\left\| x_m - x^* \right\|^2 \le \frac{2}{\rho} \mu_o \exp\left[-\rho \sum_{k=0}^{n-1} \frac{\varphi(x_k) - \varphi(x_{k+1})}{\left\| \varphi'(x_k) \right\|^2} \right], m \in N$$

$$\le \frac{2}{\rho}\mu_o \exp\left[-\rho\sum_{k=0}^{m-1}\frac{1}{2L}\lambda_k{\alpha_k}^2\right], m\in N$$

$$\le \frac{2}{\rho}\mu_o \exp\left[\frac{-\rho}{2L}\sum_{k=0}^{m-1}\lambda_k{\alpha_k}^2\right], m\in N .$$

II .If condition (3.4) holds for the sequence (x_k) , by Lemma 3. 2 and Theorem 2. 5

$$\varphi(x_m)-\varphi(x^*) \le \mu_o \exp\left[-\rho\sum_{k=0}^{n-1}\frac{\varphi(x_k)-\varphi(x_{k+1})}{\left\|\varphi'(x_k)\right\|^2}\right], m\in N$$

$$\le \mu_o \exp\left[-\rho\sum_{k=0}^{m-1}\frac{1}{L}\gamma_k{\alpha_k}^2\right], m\in N$$

$$= \mu_o\, exp\left[\frac{-\rho}{L}\sum_{k=0}^{m-1}\gamma_k{\alpha_k}^2\right], m\in N$$

and

$$\left\|x_m - x^*\right\|^2 \le \frac{2}{\rho}\mu_o \exp\left[-\rho\sum_{k=0}^{n-1}\frac{\varphi(x_k)-\varphi(x_{k+1})}{\left\|\varphi'(x_k)\right\|^2}\right], m\in N$$

$$\le \frac{2}{\rho}\mu_o \exp\left[-\rho\sum_{k=0}^{m-1}\frac{1}{L}\gamma_k{\alpha_k}^2\right], m\in N$$

$$= \frac{2}{\rho}\mu_o\, exp\left[\frac{-\rho}{L}\sum_{k=0}^{m-1}\gamma_k{\alpha_k}^2\right], m\in N .$$

Remark 3.2: If $\alpha_k \ge \alpha > 0$, $\gamma_k \ge \gamma > 0$, $\lambda_k \ge \lambda > 0$ for all $k\in N_0$,

Theorem 3.1 implies $\varphi(x_m)-\varphi^* \le C_1\frac{1}{m}$, $m\in N_0$,

Where $C_1 := \min\{\gamma, \frac{\lambda}{2}\}\left(\frac{\alpha^2}{2L\eta^2}\right)$. and

Theorem 3.2 implies

$\varphi(x_m)-\varphi^* \le \mu_0\, exp\{-C_2 m\}$, $m\in N_0$ and

$$\left\|x_m - x^*\right\|^2 \leq \frac{2}{\rho}\mu_0 \exp[-C_1 m] \quad , m \in N_0 \; .$$

Where $C_2 := \min\left\{\gamma, \frac{\lambda}{2}\right\}\left(\frac{\rho\alpha^2}{L}\right)$. //

3.2 Choice of the direction of descent.

As we have said, the choice of $- s_k$ and β_k in relaxation sequence is to satisfy the inequality $\varphi(x_k) - \varphi(x_{k+1}) \geq 0$, where $x_{k+1} = x_k - \beta_k s_k$, $k \in N_0$. Now we consider $\varphi \in C'(\Re^n)$,

$\langle \varphi'(x_k), s_k \rangle = \varphi'_+(x_k, s_k) = \lim_{t \downarrow 0} \frac{\varphi(x_k - ts_k) - \varphi(x_k)}{t}$ so the decrease in the direction $-s_k$ at x_k is guaranteed if $\langle \varphi'(x_k) - s_k \rangle > 0$.

3.3 Classification of descent methods

The classification is mostly based on the choice of the direction of descent. We will consider three major classes of descent methods.

1. Zero – order method
2. First – order method
3. Second – order method

3.3.1 Zero – order methods

The direction of descent in all of the methods in this class is the direction along the coordinate axis.

I. Coordinate – wise decent: - As the name imply, the direction of descent $- s_k$ at x_k is either $- e_j$ or ej ($e_j = (\delta_{ji})$, $i, j \in \{0, \ldots, n\}$).

So the general trend of the sequence (x_k) is as follows :

$$x_{k+1} = x_k \pm \beta_k e_j(k), \tag{3.6}$$

where $j(k) \equiv k \pmod{n} + 1$, $k \in N_0$,

Depending on the methods of calculating the step we consider three way's of constructing the sequence (x_k).

A. **Coordinate-wise descent** with duplication of step (the first Version).

The parameter β_k in (3.6) calculated at each iteration using the duplication method and x_0 is chosen arbitrarily.

B. **Coordinate wise descent with duplication of step** (second version).

Still calculated in duplication method with a slit modification, as follows.

Let l be the l^{th} cycle and each cycle consists n-iterations:

$x_k, x_{k+1}, \ldots, x_{k+(n-1)}$ is taken as the iterations in the l^{th} cycle,

where $k+i = n(l-1)+i$, $i \in \{0, \ldots, n-1\}$, $1 \in N$.

So in the formula (3.6) the index k can be replaced by $n(l-1)+i$.

The duplication technique in this case takes place in terms of the cycle l. i.e the parameter β remains constant for each iteration within the cycle.

i) We take $\beta_{k-1} = \beta_{k+i}$, $i \in \{0, \ldots, n-1\}$

where $k = n(l-1)$

In the l^{th} cycle the construction of x_k is as follows

ii) If $\varphi(x_{kti} - \beta e_{j(k+1)}) < \varphi(x_{k+i})$,

$x_{k+i+1} := x_{k+i} - \beta e_{j(k+i)}$ and pass to the next iteration.

iii) if $\varphi(x_{k+i} - \beta e_{j(k+i)}) \geq \varphi(x_{kti})$, then we chenge the direction of descent to $-e_{j(k+i)}$

iv) If $\varphi(x_{k+i} + \beta e_{j(k+i)}) < \varphi(x_{k+l})$, then

$x_{k+i+1} := x_{k+i} + \beta e_{j(k+i)}$

V) If in case $\varphi(x_{k+i+1} + \beta e_{j(k+i)}) \geq \varphi(x_{k+i})$, then

$$x_{k+i+1} := x_{k+i},$$

When the inequality in (ii) and (V) holds for each iterations of the cycle, the coming cycle proceeds with step length $\beta = \frac{1}{2}\beta_{k-1}$, and so the process continues.

Therorem 3.4: Let $\varphi \in C^{1,1}(\Re^n)$ and $X_0 := N_\alpha(\varphi)$ be bounded , then

1. $\beta_k \xrightarrow{k \to \infty} 0$,
2. $\lim\limits_{k \to \infty} d(x_u, X_0^*) = 0$.

Proof:

1. From the construction of β , (β_k) is a decreasing and bounded sequence. Suppose

$\beta_k \xrightarrow{k\to\infty} 0$, hence there is $k_o \in N$ such that

$\beta_k = \beta \; \forall k \geq k_0$ with $\varphi(x_{k+1}) < \varphi(x_k)$ for all $k \geq k_0$. Hence we have

$\| x_{k+1} - x_k \| = \beta \; \forall_k \geq k_0$ as a result there is $k_1 > k_0$ such that x_k □ X_0 for all

$k \geq k_1$ which contradicts the definition of X_0.

2. Let us consider a cycle for which the inequality (ii) & (V) hold simultaneously for all $i \in \{0, \ldots, n-1\}$. Let re index these cycle by m , keeping there original order.

Hence we have $x_m := x_{kti}$ for all

$i \in \{0, \ldots, n-1\}$ in the m^{th} cycle.

Since $\varphi(x_m + \beta_m e_{j+1}) \geq \varphi(x_m)$, $i \in \{0, \ldots, n-1\}$, $m \in N$,

And $\varphi(x_m - \beta_m e_{i+1}) \geq \varphi(x_m)$, $i \in \{0, \ldots n-1\}$, $m \in N$,

implies

$$\varphi(x_m + \beta_m e_{j+1}) - \varphi(x_m) = \beta_m \langle \varphi(x_m + \theta_1 \beta_m e_{i+1}), e_{j+1} \rangle \geq 0$$

and $\varphi(x_m - \beta_m e_{j+1}) - \varphi(x_m) = \beta_m \langle \varphi' (x_m + \theta_2 \beta_m e_{i+1}), - e_{j+1} \rangle \geq 0$

Hence we have $\langle \varphi' (x_m + \theta_1 \beta_m e_{i+1}), e_{i+1} \rangle \geq 0$ (a)

$$i \in \{0,1 \ldots, n-1\},$$

and

$$\langle \varphi' (x_m - \theta_2 \beta_m e_{i+1}), - ei+1 \rangle \geq 0 \quad . \tag{b}$$

$$i \in \{0,1 \ldots, n-1\}$$

Since $(x_m) \subseteq X_0$ (a bounded set) so (x_m) posses a convergent subsequence (x_{m_j}) (Bolzano-Wearistrass Theorem).

$\bar{x} := \lim_{j\to\infty} x_{m_j}$, and applying property 1. of the Theorem

$$0 \leq \lim_{j\to\infty} \langle \varphi'(x_{mj} + \theta_1 \beta_{mj} e_{i+1}), e_{i+1} \rangle = \langle \varphi'(x), e_{+1} \rangle , i \in \{ 0,1, \ldots, n-1\}$$

and

$$0 \leq \lim_{j\to\infty} \langle \varphi'(x_{mj} - \theta_2 \beta_{mj} e_{i+1}), e_{i+1} \rangle = \langle \varphi'(x), -e_{i+1} \rangle , i \in \{0,1, \ldots, n-1\}$$

implies $\varphi'(\overline{x}) = 0$, Consequently $\overline{x} \in X_0^*$

As $\lim_{k\to\infty} \beta_k = 0$, we have $\|x_k - x_{k+1}\| \xrightarrow{k\to\infty} 0$

Hence $\lim_{k\to\infty} d(x_k, X_o^*) = 0$.

Example 3.1: $\varphi(x) = \eta_1^2 - 2\eta_1\eta_2 + \eta_2^2 - 2\eta_1 - 1$, $x := (\eta_1, \eta_2) \in \Re^2$.

Now $x_0 := (1,1)$,

implies $\varphi(x_0) = -1$.

Let $\beta_0 = ½$,

hence $(1,1) - ½\,(1,0) = (½, 1)$.

But $\varphi(½, 1) = -3/4 > -1$, undesired result, so we change the direction of descent to (-1, 0), hence

$(1,1) + 1/2(1,0) = (3/2, 1)$

$\Rightarrow$ $\varphi(3/2, 1) = -7/4 < -1$

so set $x_1 = (3/2, 1)$.

Let $\beta = ½$

$\Rightarrow$ $(3/2, 1) - 1/2\,(0,1) = (3/2, ½)$.

$\Rightarrow$ $\varphi(3/2, ½) = -5/2 < -7/4$

Again we set $x_2 := (3/2, ½)$

Then we change the direction of descent to either (1, 0) or (-1,0)

Now, $(3/2, ½) - ½\,(1,0) = (1, ½)$

$\Rightarrow$ $\varphi(1, 1/2) > 2$ undesired.

Let $(3/2, ½) + 1/2\,(1, 0) = (2, ½)$

$\Rightarrow$ $\varphi\,(2, ½) = -9/4 > -8/2$ undesired.

So we set $x_3 := x_2$.

Now changing the direction of descent to either (0,1) or (0,-1) for 4^{th} iteration.

Let $(3/2, ½) - ½\,(0, 1) = (3/2, 0)$

$\varphi(3/2, 0) = -7/4 > -5/2$ undesired.

Again we let (3/2, ½)+ ½ (0,1) = (3/2, 1), so

$\varphi(3/2, 1) = -5/4 > -5/2$ undesired.

Hence $x_4 = x_3 = x_2$

for x_5 we let $\beta = 1/4$, but still $x_5 = x_2$

and also this is true for all $k \geq 2$, i.e $x_k = x_2 \ \forall k \geq 2$.

So $x_k \to x_2 = (3/2, 1/2)^T$.

Therefore $(3/2, 1/2)^T$ is a minimal point of φ.

II. Cyclic coordinate wise descent. The only difference of this method from the methods I is in the selection of β. In this method one- dimensional minimization technique or the technique in section 4.1.3 is used for selecting β.

Theorem 3.5:

Let 1. $\varphi \in C^{1,1}(\Re^n)$ and $\varphi(x) \geq c_0 \ \forall x \in X_0$,

2. $X_0^* \neq \varnothing$

3. (x_k) satisfies condition (3.6)

4. (X_k) stifles either condition (3.2) for $1 \geq \lambda_k \geq \lambda > 0$ or condition (3.4) for $1/2 \geq q_k \geq q > 0$

5. $\| x_{k+1} - x_k \| \to 0$,

6. $\forall \varepsilon > 0 \ \exists \delta > 0: \| \varphi'(x_k) \| \geq \delta \ \ \forall x \in X_0^* \setminus U_\varepsilon$.

Then $\lim\limits_{k \to \infty} d\left(x_k, X_0^*\right) = 0$.

Proof: since (x_k) satisfies assumption 3. and 4. of the Theorem. By Lemma 3.1 and Lemma 3.2 we have,

$$\varphi(x_k) - \varphi(x_{k+1}) \geq c\alpha_k^2 \| \varphi'(x_k) \|^2$$

Where $C := \min\left\{\frac{\lambda}{2}, \gamma\right\}$, $\alpha_k = \dfrac{\left\langle \varphi'(x_k), e_{j(k)} \right\rangle}{\|\varphi'(x_k)\|} > 0$

implies $\varphi(x_k) - \varphi(x_{k+1}) \geq C \langle \varphi'(x_k), e_{jk} \rangle^2 \geq 0$

Since φ is bounded on X_0 and $(\varphi(x_k))$ is monotone decreasing sequence.

$$0 = \lim_{k \to \infty} [\varphi(x_k) - \varphi(x_{k+})] \geq C \lim_{k \to \infty} \left\langle \varphi'(x_k) e_{j(k)} \right\rangle^2 \geq 0.$$

Hence. $\lim_{k\to\infty}\langle \varphi'(x_k), e_{j(k)}\rangle = 0$

Consequently given $\varepsilon > 0$ there is $k_1 \in \mathbf{N}$ such that

$\langle \varphi'(x_k), e_{j(k)}\rangle| \leq L\,\varepsilon$, $\forall\ k \geq k_1$ and also there is $k_2 > o$ such that

$\| x_{k+1} - x_k \| \leq \varepsilon\ \forall k \geq k_2.$

Now $\quad \| x_{k+1} - x_k \| = \| x_{k+1} - x_{k+i-1} + x_{k+i-1} \ldots + x_{k+1} - x_k \|$

$$\leq \| x_{k+i} - x_{k+i} - 1 \| + \ldots + \| x_{k+1} - x_k \|$$

$$\leq \frac{\varepsilon}{n}\, i \leq \varepsilon \text{ for } i \leq n \text{ and } k \geq k_0$$

Where $k_0 = \max\,(k_1, k_2\}$.

As $\varphi \in C^{1,1}((\Re^n)$,

$\| \varphi'(x_{k+i}) - \varphi'(x_k) \| \leq L \| x_{k+i} - x_k \| \leq L\varepsilon$, $i \in \{1, \ldots n-1\}$, $\forall k \geq k_0.$

Also

$$|\langle \varphi'(x_k), e_{j(k)} \rangle| = |\langle \varphi'(x_k) - \varphi'(x_{k+i}), e_{j(k+i)}\rangle + \langle \varphi'(x_{k+i})\rangle|$$

$$\leq \left|\langle \varphi'(x_k) - \varphi'(x_{k+i}), e_{j(k+i)}\rangle\right| + \left|\langle \varphi'(x_{k+i}), e_{j(k+i)}\rangle\right|$$

$$\leq \|\varphi'(x_k) - \varphi'(x_{k+i})\| + \left|\langle \varphi'(x_{k+i}), e_{j(k+i)}\rangle\right|$$

$$\leq 2L\varepsilon,\ i \in \{0,\ldots, n-1\} \quad \text{and for all } k \geq k_0\ .$$

implies $\quad \max_{i\in\{1,\ldots,n\}} \left|\langle \varphi'(x_k), e_{j(k)}\rangle\right| \leq 2L\varepsilon.$

$$\text{Now} \quad \| \varphi'(x_k) \| = \sqrt{\sum_{i=1}^{n} \langle \varphi'(x_k), e_{j(k)}\rangle^2}$$

$$\leq \sqrt{n \max_{i\in\{1,..,n\}} \langle \varphi'(x_k), e_{j(k+i)}\rangle^2}$$

$$\leq \sqrt{n} \max_{i\in\{1,..,n\}} \left|\langle \varphi'(x_k), e_{j(k+i)}\rangle\right| \leq 2\sqrt{n}L\varepsilon\ .$$

Hence $\lim_{k\to\infty} \varphi'(x_k) = 0$. Then the assertion of the Theorem follows from Theorem1.2.5. //

III. **The classical scheme:** even though the direction of descent of φ at a point is either e_j or-e_j ($i \in \{1, \ldots n\}$) the choice of j^{th} coordinate axis is dependant of continuously partially differentiability of φ at that

$$x_{k+1:} = x_k \pm \beta_k e_j, \tag{3.7}$$

where

e_j is selected if $\max_{i\in\{1,..n\}} |\langle \varphi'(x_k), e_i \rangle| = |\varphi'(x_k), e_j|$

we assume $|\varphi'(x_k), e_j| > 0$ otherwise $\varphi'(x_k) = 0$.

And the process of minimization stops.

$$\alpha_k^2 = \frac{\langle \varphi'(x_k), s_k \rangle^2}{\|\varphi'(x_k)\|^2 \|s_k\|^2}$$

$$= \frac{\langle \varphi'(x_k), e_j \rangle^2}{\sum_{i=1}^{n} \langle \varphi'(x_k), e_i \rangle^2} \geq \frac{\langle \varphi'(x_k), e_j \rangle^2}{n \langle \varphi'(x_k), e_j \rangle^n} = \frac{1}{n}.$$

Theorem 3.6:

Let 1. $\varphi \in C^{1,1}$ ($\Re^n$), and $\varphi(x) \geq c_0$ $\forall x \in X_0$,

2. $X^* \neq \emptyset$,
3. (x_k) Satisfies condition (3.7) above,
4. (x_k) satisfies either condition (3.2) or condition (3.4).

Then $d(x_k, X^*) \to 0$.

Proof: $\alpha_k^2 \geq 1/n > 0$ $\forall k \in \mathbf{N}$, where n is the dimension of $\Re^n$.

Hence the assertion follows immediately from Theorem 3.2. //

Example 3.2.

Let $\varphi(x) = \|x\| + e^{-\|x\|}$, $x \in \Re^2$

By example 1.3.2. $\varphi \in C^{1,1}(\Re^2)$

More over $\varphi'(x_0) = \frac{(1-e^{-\|x\|})}{\|x\|}x$

Choose $x_{0=(1,1)}$

$$\Rightarrow \varphi'(x_0) = \left(\frac{1-e^{-\sqrt{2}}}{\sqrt{2}}\right)(1,1)^T$$

$$\Rightarrow |\langle \varphi'(x_0), e_1\rangle| = |\langle \varphi'(x_0), e_2\rangle| = \frac{1-e^{-\sqrt{2}}}{\sqrt{2}}$$

where $e_1=(1,0)$ and $e_2=(0,1)$

So we can choose $s_0 = \pm e_1$ or $s_0 = \pm e_2$

Let $s_0 = e_1$

$$\psi(\beta) := \varphi(x_0 - \beta e_1)$$

$$\Rightarrow \psi'(\beta) = \langle \varphi'(x_0 - \beta e_1), -e_1\rangle$$

$$= \frac{1-e^{-\|x_0-\beta e_1\|}}{\|x_0-\beta e_1\|}\langle x_0 - \beta e_1, -e_1\rangle$$

$$\varphi'(\beta) = 0 \Rightarrow \langle x_0 - \beta e_1, e_1\rangle = 0$$

$$\Rightarrow \beta = \langle x_0, e_1\rangle$$

$$= \langle (1,1)^T, (1,0)^T\rangle$$

$$= 1$$

$\beta_0 := 1$

$x_1 := x_0 - \beta_0$

$:=(1,1)^T-(1,0)^T=(0,1)^T.$

So $\varphi'(x_1) = (1-e^{-1})(0,1)^T$

$$|\langle \varphi'(x_1), e_1\rangle| = |\langle (1-e^{-1})(0,1)^T, (1,0)^T\rangle| = 0$$

and

$$|\langle \varphi'(x_1), e_2\rangle| = |\langle (1-e^{-1})(0,1)^T, (0,1)^T\rangle| = 1-e^{-1}.$$

Here we choose $s_1 = \pm e_2$

$$\varphi(x_1 - \beta e_2) = \varphi((0,1) - \beta(0,1)), \qquad \beta \in [0,\infty)$$

$$= \varphi((1-\beta)(0,1))$$

$$= |1-\beta| + e^{-|1-\beta|}$$

$$\psi_2(\beta) = |1-\beta| + e^{-|1-\beta|}$$

then $\psi_2'(\beta) = \frac{1-\beta}{|1-\beta|}(1-e^{-|1-\beta|})$

Since ψ_2' is continuous over $[0,\infty)$,

$\psi'_2(\beta) = 0$ if $\beta = 1$

so $\beta_1 := 1$

$x_2 := x_1 - \beta_1 e_2$

$= (0,1) - 1(0,1)$

$= (0,0)$

and $\varphi'(x_2) = 0$, but $\langle \varphi'(x_2 - \beta e_i), e_i \rangle = 0$, for $i \in \{1,2\}$ and $\beta \in [0, \infty)$

$\Rightarrow \beta = 0$

$\Rightarrow x_3 = x_2 \pm \beta_2 e_i = x_2$

$\Rightarrow x_k = x_2 \ \forall k \geq 2$,

$\therefore x_2 = (0,0)$ is a minimal point of φ over $\Re^2$.

3.3.2 First order methods. Descent methods under this class involve either $-\varphi(x_k)$ or $-A_k\varphi'(x_k)$ as a direction of descent. Where

A_k is a symmetrical positive definite matrix.

I.Gradient descent methods:

Let $s_k := \varphi'(x_k)$, $k \in \mathbf{N}_0$

$x_{k+i} := x_k - \beta_k \ \varphi'(x_k)$

Usually β_k is calculated using one dimensional minimization method (sce. 3. 2. 1).

Theorem 3.7:

Let: 1. $\varphi \in C^{1,1}(\Re^n)$ and convex,

2. $X^* \neq \varnothing$,

3. (x_k) satisfies condition (3.4) with $q_k \in (1/2, 1)$,

4. $s_k = \varphi'(x_k)$

Then

$$\lim_{k \to \infty} d(x_k, X^*) = 0$$

Proof: Let $y \in X^* \neq \varnothing$. Hence $\varphi(y) \leq \varphi(x) \ \forall x \in \Re^n$.

By sub gradient inequality and supposition 3. of the Theorem,

$$\begin{aligned}\langle \varphi'(x_k), x_k - y \rangle \geq \varphi(x_k) - \varphi(y) &\geq \varphi(x_k) - \varphi(x_{k+1}) \\ &\geq q_k \beta_k \langle \varphi'(x_k), s_k \rangle \\ &= q_k \beta_k \| \varphi'(x_k) \|^2\end{aligned}$$

implies $\langle \varphi'(x_k), x_k\text{-}y\rangle \ \geq \ q_k \beta_k \ \| \varphi'(x_k) \|^2$. (a)

Multiplying the expression in (a) by - β_k and adding

$\beta^2_k \ \| \varphi'(xk) \ \|^2$ to both sides of the inequality we obtain,

$- \beta_k \langle \varphi'(x_k), x_k\text{-}y \rangle + \ \beta^2_k \| (\varphi'(x_k) \|^2 \ \leq \ \beta^2_k(1\text{-}2q_k) \| \ \varphi' (x_k) \|^2$. (b)

For $q_k \in (½ , 1)$, $(1\text{-}2q_k) < 0$

From this and condition (b)

$$
\begin{aligned}
\| x_{k+1}\text{-}y \|^2 \ &= \ \| x_k\text{-} \beta_k \ \varphi'(x_k) - y \|^2 \\
&= \ \| (x_k\text{-}y)\text{-} \beta_k \ \varphi'(x_k) \|^2 \\
&= \ \langle (x_k\text{-}y) - \beta_k \ \varphi'(x_k), (x_k\text{-}y)\text{-} \beta_k \ \varphi'(x_k) \rangle \\
&= \ \langle x_k\text{-}y, x_k\text{-}y \rangle \ \text{-}2\beta_k \langle \varphi'(x_k), x_k\text{-}y \rangle + \ \beta^2_k \langle \varphi'(x_k), \varphi'(x_k) \rangle \\
&\leq \ \| x_k\text{-}y \|^2 + \beta^2_k (1\text{-}q_k) \ \| \ \varphi'(x_k) \|^2 \\
&< \ \| x_k - y \|^2.
\end{aligned}
$$

Implies the sequence ($\| x_k - y \|$) is a monotone decreasing. By

Bolzano-werestrasses Theorem (x_k) possesses a convergent subsequence (x_{kj}).

$$x^* := \lim_{j\to\infty} x_{k_j}.$$

As φ is convex, $X^* \neq \varnothing$, and (x_k) a relation sequence, ($\varphi(x_k)$) is a mouton decreasing and bounded.

$$\lim_{k\to\infty} \varphi(x_k) = \varphi^* = \min_{x\in\Re^n} \varphi(x) \ .$$

Since $\varphi(x^*) = \varphi\left(\lim_{j\to\infty} x_{k_j}\right) = \lim_{j\to\infty} \varphi(x_{k_j}) = \lim_{k\to\infty} \varphi(x_k)$, $x^* \in X^*$ and $x_k \xrightarrow{k\to\infty} x^*$.

Therefore $d(x_k, X^*) \xrightarrow{k\to\infty} 0$. //

Example 3.2

$$\varphi(x)=\|x\|+e^{-\|x\|} \quad , x\in \Re^2$$

in Example 1.3.2 we have seen $\varphi \in C^{1,1}(\Re^2)$ and

$$\varphi'(x)=\frac{1-e^{-\|x\|}}{\|x\|}x \ .$$

Let $x_0 := (1,1)^T$

$$\Rightarrow \quad \varphi'(1,1)=\frac{1}{\sqrt{2}}\left(1-e^{-\sqrt{2}}\right)(1,1)^T$$

$$\psi_1(\beta):=\varphi(x_0-\beta\varphi'(x_0)) \ , \ \beta\in[0,\infty)$$

$$\Rightarrow \quad \psi_1'(\beta)=\langle \varphi'(x_0-\beta\varphi'(x_0)),-\varphi'(x_0)\rangle$$

$$=\frac{1-e^{-\|x_0-\beta\varphi'(x_0)\|}}{\|x_0-\beta\varphi'(x_0)\|}\langle x_0-\beta\varphi'(x_0),-\varphi'(x_0)\rangle$$

$$\psi_1'(\beta)=0 \Rightarrow \quad \beta\langle \varphi'(x_0),\varphi'(x_0)\rangle=\langle x_0,\varphi'(x_0)\rangle$$

$$\Rightarrow \beta=\frac{\sqrt{2}}{1-e^{-\sqrt{2}}}$$

$\beta_0 \quad := \dfrac{\sqrt{2}}{1-e^{-\sqrt{2}}}$, hence $x_1 := x_0-\beta_0\varphi'(x_0)$

$$=(1,1)^T-\left(\frac{\sqrt{2}}{1-e^{-\sqrt{2}}}\right)\left(\frac{1}{\sqrt{2}}\left(1-e^{-\sqrt{2}}\right)(1,1)^T\right)$$

$$=0$$

$$\Rightarrow \quad \varphi'(x_1)=\varphi'(0)=0$$

Since $\quad x_{k+1}=x_k-\beta_k\varphi'(x_k) \ \forall k\in N_0 \ , x_k = x_1 = 0$

$$x_k = x_1 = 0$$

Example3.3 : $\varphi(x) = \eta_1^2+\eta_2^2+\eta_1+\eta_2 \quad , x=(\eta_1 , \eta_2) \in \Re^2$.

$\varphi'(x) = (2\eta_1+1, 2\eta_2+1)$.

$\Rightarrow \quad \|\varphi'(x) - \varphi'(y)\| = 2\|x-y\| \ , x,y\in \Re^2$.

$\Rightarrow \quad \varphi \in C^{1,1}(\mathbf{R}^2)$

Choose $x_0 := (1,1)$

$\varphi'(x_0) = (3, 3)$

$\psi_1(\beta) := \varphi(x_0-\beta(x_0)) \qquad , \beta \in (0, \infty)$

$= \varphi(1-3\beta, 1-3\beta)$

$= 2(1-3\beta)^2 + 2(1-3\beta)$

$\Rightarrow \qquad \psi_1'(\beta) = -12(1-3\beta) + 2(1-3\beta)$

$\Rightarrow \qquad \psi_1'(\beta) = 0$ iff $\beta = ½$

$\beta_1 := ½$

and $x_1 := (1,1) - ½\,(3,3) = (-1/2, -1/2)^T$

similarly $\varphi'(x_1) = (0, 0)^T$

So, for any $\beta \geq 0$

$x_2 := x_1 - \beta_1\, \varphi'(x_1) = x_1$

Hence $x_k = x_1$, $\forall k \geq 2$

There for $x_k \to x = (-1/2, -1/2)^T$

$\Rightarrow$ $(-1/2, -1/2)$ is the minima of φ over $\Re^2$.

II. Generalized gradient descent method

This method is a little bit modified from gradient descent method $-s_k := A_k\, \varphi(x_k)$, $k \in \mathbf{N}_0$, where A_k is symmetrical positive definite matrix.

Theorem 3.8:

1. $\varphi \in C^{1,1}(\Re^n)$ and $\varphi(x) \geq c\ \forall x \in X_0$,
2. $X^* \neq \varnothing$,
3. $x_{k+1} = x_k - \beta_k A_k \varphi'(x_k)$, $k \in \mathbf{N}_0$,
4. β_k satisfies either condition (3.2) or (3.4),
5. $\forall\, \varepsilon > 0\ \exists \delta > 0$: $\|\varphi'(x_k)\| > \delta\ \forall x_k \in U\varepsilon$,
6. $\|\lambda_k\| \leq \xi$, $\qquad \forall k \in \mathbf{N}_0$,
7. There is $\nu > 0$ lower bound for all eigen values of s_k ($k \in \mathbf{N}_0$)

Then $\lim_{k\to\infty} d(x_k, X^*) = 0$.

Proof:

Since A_k is symmetrical positive definite, all it's eigen values are positive and

$\langle A_k x, x \rangle \geq \nu_k \langle x,x\rangle \geq \nu \langle x,x\rangle \;\; \forall x \in \Re^n, k \in \mathbf{N}_0,$

Where ν_k is the minimum eigen value of A_k.

$$\alpha_k = \frac{\langle \varphi'(x_k), A_k\varphi'(x_k)\rangle}{\|\varphi'(x_k)\|\|A_k\varphi'(x_k)\|} \geq \frac{\nu\|\varphi'(x_k)\|^2}{\|\varphi'(x_k)\|^2\|A_k\|} \geq \frac{\nu}{\xi} > 0$$

Hence by Theorems 3.1 the assertion follows.

Note : if $A_k = I_n$ (I_n = nxn identify matrix) for all $k \in \mathbf{N}_0$, $-s_k = -A_k\, \varphi'(x_k) = -\varphi'(x_k)$.

3.3.Second order methods.

I. Newton's methods:

Let $\varphi: \Re^n \to \Re$ be strongly and sufficiently smooth (has derivative of all order) in a neighbor hood U of the point $x^* \in X^*$. For each $x_k \in U$ we define $\psi: U \to \Re$ by

$\psi(x) = \varphi(x_k) + \langle \varphi'(x_k), x - x_k\rangle + ½ \langle \varphi''(x_k), (x-x_k)\rangle$.Obliviously ψ approximants φ on U hence, ψ is strongly convex and sufficiently smooth on U

$x \in M(\psi, U)$ iff $\psi'(x) = 0$

implies $x = x_k - (\varphi''(x_k))^{-1}\, \varphi'(x_k)$.

Now

$$x_{k+1} := x_k - \varphi''(x_k)^{-1}\, \varphi'(x_k)$$

$$\begin{aligned} x_{k+1} - x^* &= x_k - x^* - \varphi''(x_k)^{-1} \qquad \varphi'(x_k) \\ &= \varphi''(x_k)^{-1}\,[\varphi''(x_k)(x_k - x^*) - \varphi'(x_k)] \\ &= \varphi''(x_k)^{-1}\,[\varphi''(x_k)\,(x_k - x^*) - (\varphi'(x_k) - \varphi'(x^*))] \end{aligned}$$

Were $x^* \in X^*$ implies $\varphi'(x_k) = 0$.

Since φ is sufficiently smooth (i.e has a derivative of all order) and hence it is quadratic from on U.

$$\varphi''(x_k) = \varphi''(x^*) + \varphi'''(x^*)\,(x_k - x^*) + \omega(x_k + x^*),$$

where $\|\omega(x_k, x^*)\| \leq c_1 \|x_k - x^*\|$.

But $\varphi'''(x) = 0\; \forall x \in U$ implies

$$\varphi''(x_k) = \varphi''(x^*) + \omega(x_k, x^*); \|\omega(x_k, x^*)\| \le c_1 \|x_k - x^*\|$$

Similarly $\quad \varphi'(x_k) - \varphi'(x^*) = \varphi''(x^*)(x_k - x^*) + \xi(x_k, x^*)$

where $\|\xi(x_k, x^*)\| \le c_2 \|x_k - x^*\|^2$.

Therefore

$$\begin{aligned}
\| \varphi'(x_k) - \varphi'(x^0) - \varphi''(x_k)(x_k - x^*)\| &= \| \varphi''(x^*)(x_k - x^*) + \xi(x_k, x^*) - \varphi''(x^*)(x_k - x^*) \\
&\quad - \omega(x_k, x^*)(x_k - X^*)\| \\
&\le \|\xi(x_k, x^*)\| + \|\omega(x_k, x^*)\| \|x_k - x^*\| \\
&\le \|\varepsilon(x_k, x^*)\| + \|\omega(x_k, x^*)\| \|x_k - x^*\| \\
&\le c_2 \|x_k - x^*\|^2 + c_1 \|x_k - x^*\| \|x_k - x^*\| \\
&\le c \|x_k - x^*\|^2
\end{aligned}$$

where $c := c_2 + c_1$

implies $\quad \|x_{k+1} - x^*\| \le c \|x_k - x^*\|^2 \; k \in \mathbf{N}_0$.

W.O.L.G. take $\|x_k - x^*\| < 1$.As x^* is a stationary point and φ is strongly convex on U,

$$x_k \xrightarrow{k \to \infty} x^*.$$

We can put the above argumentation as a theorem.

Theorem 3.9

Let 1. $\varphi: \Re^n \to \Re$. be strongly convex and sufficiently Smooth on a neighbor hood U of the point $x^* \in X^*$,

2. $-s_k := \varphi''(x_k)^{-1} \varphi'(x_k)$, where $-s_k$ is a direction of descent of at the point x_k .

Then

$$\lim_{k \to \infty} d(x_k, X^*) = 0.$$

Example 3.4: let $\varphi(x) = 4\eta_1^2 + 3\eta_1\eta_2 + 2\eta_2^2,\ x := (\eta_1, \eta_2) \in \Re^2$

implies $\varphi'(x) = (8\eta_1 + 3\eta_2,\ 4\eta_2 + 3\eta_1)^T$ and

$$\varphi''(x) = \begin{pmatrix} 8 & 3 \\ 3 & 4 \end{pmatrix}$$

Since the hesse matrix is positive definite, φ is strongly convex over $\Re^2$.

$$(\varphi''(x))^{-1} = \frac{1}{23}\begin{pmatrix} 4 & -3 \\ -3 & 8 \end{pmatrix}.$$

Let $x_0 := (1,0)^T$.

$$\Rightarrow \quad x_1 := x_0 - \left(\varphi''(x_0)^{-1}\varphi'(x_0)\right)$$

$$= (1,0)^T - \frac{1}{23}\begin{pmatrix} 4 & -3 \\ -3 & 8 \end{pmatrix}\begin{pmatrix} 8 \\ 3 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

$$\Rightarrow \quad x_k = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \forall k \geq 2.$$

Hence $(0,0)^T$ is a global minimal of φ.

In practice it is not easy to calculate $((\varphi''(x))^{-1}$ at each iteration, so it is advisable to take $(\varphi''(x))^{-1}$ constant for a fixed umber of iterations say $m \geq 1$. For each 1^{th} cycle with length m we construct a recurrence relation as follows:

$x_{lm+i+1} = x_{lm+i} - \beta_{lm+i}\ (\varphi''(x_{lm}))^{-1}\ \varphi'(x_{lm+i})$, then pass to the next $(1+1)^{th}$ cycle. The above method is usually called modified Newton methods.

II Method of conjugate direction

The scheme of the method is as follows.

i) $x_{k+1} = x_k - \beta_k s_k,\ k \in N_0$,

ii) $s_0 = \varphi'(x_0),\ s_k = \varphi'(x_k) - \xi_k\, s_{k-1},\ k \in N$

where ξ_k is a certain parameter to be determined .

iii) $\beta_k \in M(\varphi(x_k - \beta s_k), \beta \geq 0)\ ,\ k \in N_0$.

The various choices of the parameter ξ_k is a base for the classification of the method. we will see two methods for calculating ξ_k . Before that we prove two important Lemmas.

Lemma 3.3: Let φ be differentiable function, and (x_k) be a sequence satisfies i., ii., and iii. above.

Then

a) $\langle\varphi'(x_{k+1}), s_k\rangle = 0$, $k \in N_0$.

b) $\langle\varphi'(x_h), s_k\rangle = \|\varphi'(x_k)\|^2$, $k \in N_{0.}$

Proof: a) since $q_k \in M\ (\varphi(x_k - \beta s_k), \beta \geq 0)$ and φ is differentiable,

If $\beta_k > 0$, then $\frac{d}{d\beta}\varphi(x_k - \beta s_k)\ |_{\beta=h} = 0$ and

If $\beta_{k=0}$, then $\frac{d}{d\beta}\varphi(x_k - \beta s_k)|_{\beta=\beta k} \geq 0$;

i.e. $\frac{d}{d\beta}\varphi(x_k - \beta_k)|_{\beta=\beta k} = \lim\limits_{\beta\downarrow 0}\frac{\varphi(x_k - \beta s_k)}{\beta} \geq 0$ (since $\beta_k = 0$ is the minimum of $\varphi(x_k - \beta s_k)$ on $(0, \infty)$.

Now, If $\beta_k > 0$, then $\quad 0 = \frac{d}{d\beta}\varphi(x_k - \beta S_k)|_{\beta=\beta k} = \langle\ \varphi'(x_k - \beta_k s_k), -s_k\ \rangle$

$$= \langle\varphi'(x_k\text{-}\beta_k s_k), -s_k\rangle.$$

For $\beta_k = 0$, we prove by induction on k

Let $\beta_0 = 0$, then $x_1 = x_0$ and $s_0 = \varphi'(x_0) = \varphi'(x_1)$.

$$0 \leq \frac{d}{d\beta}\varphi(x_0 - \beta s_0)|_{\beta=0} = \langle\varphi'(x_1), -s_0\ \rangle = \langle\varphi'(x_0), -\ s_0\rangle$$

$$= -\langle\varphi'(x_0), \varphi'(x_0)\rangle \leq 0.$$

Implies $\langle\varphi'(x_1), s_0\rangle = 0$.

We assume the relation $\langle\varphi'(x_k), s_{k-1}\rangle = 0$ holds for $k > 0$.

If $\beta_{k-1} = 0$,

$$0 \leq \frac{d}{d\beta}\varphi'(x_k - \beta s_k)\ |_{\beta=\beta k} = \langle\varphi'(x_{k+1}), -s_k\rangle$$

$$= - \langle \varphi'(x_{k+1}), \varphi'(x_k) - \xi_k s_{k-1} \rangle$$

$$= - \langle \varphi'(x_{k+1}), \varphi'(x_k) \rangle - \xi_k \langle \varphi'(x_{k+1}), s_{k-1} \rangle$$

$$= - \|\varphi'(x_k)\|^2 \le 0 .$$

Implies $\langle \varphi'(x_{k+1}\rangle, s_k\rangle = 0$.

The three equations from the last are followed from ii. and $x_{k+1} = x_k$.

In all cases $\langle \varphi'(x_{k+1}), S_k \rangle = 0$, $k \in N_0$. //

b) The proof is easily follows from ii. and a) of he Lemma.

$$\langle \varphi'(x_k), s_k \rangle = \langle \varphi'(x_k), \varphi'(x_k) - \xi_k s_{k-1} \rangle$$

$$= \langle \varphi'(x_h), \varphi'(x_k) - \xi_k \langle \varphi'(x_k), s_{k-1} \rangle$$

$$= \|\varphi'(x_k)\|^2 , k \in N_0.$$

Lemma 3.4: If $|\xi_k| \le c \dfrac{\|\varphi'(x_k)\|}{\|s_{k-1}\|}$, $k \in N$ for some $c > 0$,

then $\alpha_k = \dfrac{\langle \varphi'(x_k), s_k \rangle}{\|\varphi'(x_k)\|\|s_k\|} \ge \dfrac{1}{1+c}$

Proof: from Lemma 3.3 and ii. we have the following

$$s_k = \varphi'(x_k) - \xi_k s_{k-1}, k \in No,$$

so, a) $\|s_k\|^2 = \langle \varphi'(x_k), s_k \rangle - \xi_k \langle s_{k-1}, s_k \rangle$

$$= \|\varphi'(x_k), s_k\|^2 - \xi_k \langle s_{k-1}, s_k \rangle$$

b) $\langle s_k, s_{k-1} \rangle = \langle \varphi'(x_k), s_{k-1} \rangle - \xi_k \langle s_{k-1}, s_{k-1} \rangle$

$$= -\xi_k \| s_{k-1}\|^2.$$

From a), b) and the definition of α_k,

$$\alpha_k = \frac{\langle \varphi'(x_h), s_k \rangle}{\|\varphi'(x_k)\|\|s_k\|} = \frac{\|\varphi'(x_k)\|}{\|s_k\|} \tag{3.8}$$

and $\| s_k\|^2 = \|\varphi'(x_k)\|^2 + \xi_k^2 \| s_{k-1}\|^2$, $k \in N_0$.

From the supposition of the Lemma we have

$$\| s_k\|^2 \le \|\varphi'(x_k)\|^2 + \left(C\frac{\|\varphi'(x_k)\|}{\|s_{k-1}\|}\right)^2 . \ \| s_{k-1}\|^2$$

$$= \|\varphi'(x_k)\|^2 (1 + c^2)$$

$$\le \|\varphi'(x_k)\|^2 (1 + c)^2.$$

implies $\| s_k\| \le \|\varphi'(x_k)\| (1 + c)$

Substituting in (3.8), we obtain

$$\alpha_k = \frac{\langle \varphi'(x_k), s_k \rangle}{\|\varphi'(x_k)\|\|s_k\|} . //$$

Let us consider two methods for selecting the parameter ξ_k.

A . First Method

$$\text{Let } \xi_k := \frac{1}{\|\varphi'(x_k)\|^2} \langle \varphi'(x_k), \varphi(x_k) - \varphi'(x_{k-1}) \rangle \ k \in N_0 \tag{3.9}$$

Theorem 3.10:

Let 1. $\varphi \in C^{1,1}(\Re^n)$ and strongly convex.

2. the sequence (x_k) is constructed as in i., ii, and iii.

3. ξ_k is calculated by condition (3.9).

Then

$\varphi(x_m) - \varphi(x^*) \le \mu_0 \exp [c_1 m]$ and.

$\|x_m - x^*\| \le \frac{2}{\rho} \ \mu \exp [-c_1 m]$,

where $0 < c_1 \le \frac{\rho^3}{L(\rho + L)^2}$ and ρ a parameter of strong convexity .

Proof: from assumption of strong convexity Lemma 1.3.2

$$\langle \varphi'(x_k) - \varphi'(x_{k-1}), x_k - x_{k-n}\rangle \ge \rho \|x_k = x_{k+1}\|^2.$$

Since $x_k - x_{k-1} = \beta_{k-1} s_{k-1}$, and by Lemma 3.3

$$\rho \beta_k^2 \|s_{k-1}\|^2 \le -\beta_{k-1}\langle \varphi'(x_k) - \varphi'(x_{k-1}), s_{k-1} \rangle$$

$$= -\beta_{k-1}\langle \varphi'(x_k) - \varphi'(x_{k-1}), s_{k-1}\rangle$$

$$= \beta_{k-1}\langle \varphi'(x_k), s_{k-1}\rangle + \beta_{k-1} \langle \varphi'(x_{k-1}), \xi_{k-1}\rangle$$

$= \beta_{k-1}\ \|\varphi'(x_{k-1})\|^2.$

From the definition of ξ_k we always assume $\varphi'(x_k) \neq 0$, $k \in N$. Hence $\beta_k > 0$ otherwise by Lemma 3.3 we have

$$x_k = x_{k-1} \text{ and}$$

$$\|\varphi'(x_k)\|^2 = \langle \varphi'(x_k), s_k\rangle = \langle \varphi'(x_{k-1})\ s_k\rangle = 0.$$

This implies $\rho\beta_{k-1}\ \|s_{k-1}\|^2 \leq \|\varphi'(x_{k-1})\|^2$.

As $\varphi \in C^{1,1}(\Re^n)$.

$$\|\varphi'(x_k)\|^2 - \varphi'(x_{k-1})\| \leq L\|x_k - x_{k-1}\|.$$

$$|\xi_k| = \frac{1}{\|\varphi'(x_{k-1})\|^2}\ |\langle\ \varphi'(x_k), \varphi(x_k) - \varphi'(x_{k-1})\ \rangle|$$

$$\leq \frac{1}{\|\varphi'(x_{k-1})\|^2}\ \|\varphi'(x_k)\|\ \ \|\varphi'(x_k) - \varphi'(x_{k-1})\|$$

$$\leq \frac{1}{\rho\beta_{k-1}\|s_{k-1}\|^2}\ \|\varphi'(x_k)\|\ \ \|x_k - x_{k-1}\|L$$

$$= \frac{1}{\rho}\frac{\|\varphi'(x_k)\|}{\|s_{k-1}\|}.$$

$C: = \dfrac{1}{\rho}$, by Lemma 3.4: $\alpha_k \geq \dfrac{1}{L+\rho}$, $k \in N$.

The conclusion follows from Remark 3. 2:. //

Example3.5: Let $\quad \varphi(x) = 2\eta_1^2 + 2\eta_1\eta_2 + \eta_2^2 + \eta_1 + \eta_2,\ x := (\eta_1,\eta_2) \in \Re^2$

implies $\quad \varphi'(x) = (4\eta_1 + 2\eta_2 + 1,\ 4\eta_2 + 2\eta_1 + 1)$ and $\varphi''(x) = \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}$

(positive definite) hence φ is strongly convex.

$x:=(\eta_1,\eta_2) \quad y:=(\nu_1,\nu_2)$

$$\|\varphi'(x) - \varphi'(y)\| = \|(\eta_1 + 2\eta_2 - 4\nu_1 - 2\nu_2,\ 4\eta_2 + 2\eta_1 - 4\nu_2 - 2\nu_1)\|$$

$$= \|4(\eta_1 - \nu_1,\ \eta_2 - \nu_2) + 2(\eta_2 - \nu_2,\ \eta_1 - \nu_1)\|$$

$$\leq 4\ \|x - y\| + 2\ \|x - y\| = 6\|x - y\|.$$

Hence $\varphi \in C^{1,1}\ (\Re^n)$.

Let $x_0: = (0, 0)$

$$\Rightarrow\ s_0 = \varphi'(x_0) = (1, 1)$$

$$\Rightarrow x_1 = x_0 - \beta_0 s_0 = (\beta_0, \beta_0)$$

where $\beta_0 \in M\,(\varphi(x_0 - \beta s_0), \beta \geq 0)$

$$\psi_1(\beta) := \varphi(x_0 - \beta s_0), \quad \beta \in (0, \infty)$$

$$= 5\beta^2 - 2\beta$$

implies $\psi_1^1(\beta) = 10\beta - 2$

$$\psi_1^1(\beta) = 0 \text{ iff } \beta = \frac{1}{5}$$

$$\beta_0 := \frac{1}{5}.$$

So $\quad x_1 = \left(\frac{-1}{5}, \frac{-1}{5}\right)^T$

implies $\qquad \varphi'(x_1) = \left(\frac{-1}{5}, \frac{1}{5}\right)^T.$

$$\xi_1 = \frac{-1}{\|\varphi'(x_0)\|^2}\langle \varphi'(x_1), \varphi'(x_1) - \varphi'(x_0)\rangle$$

$$= \frac{-1}{2}\left\langle \frac{1}{5}\begin{pmatrix}-1\\1\end{pmatrix}, \frac{1}{5}\begin{pmatrix}-1\\1\end{pmatrix} - \begin{pmatrix}1\\1\end{pmatrix}\right\rangle$$

$$= \frac{-1}{25},$$

hence $\quad s_1 = \varphi'(x_1) - \xi_1 s_0$

$$= \frac{1}{5}\begin{pmatrix}-1\\1\end{pmatrix} + \frac{1}{25}\begin{pmatrix}1\\1\end{pmatrix} = \frac{1}{25}\begin{pmatrix}-4\\6\end{pmatrix}.$$

$$x_2 := x_1 - \beta_1 s_1$$

where $\beta_1 \in M\,(\varphi(x_1 - \beta s_1), \beta \geq 0)$.

$$\psi_1(\beta) := \varphi(x_1 - \beta s_1), \beta \in (0, \infty)$$

$$= \varphi\left(\frac{1}{5}\begin{pmatrix}-1\\-1\end{pmatrix} - \frac{1}{25}\begin{pmatrix}-4\\6\end{pmatrix}\right)$$

$$= \varphi\left(\frac{1}{5}\begin{pmatrix}-1+\beta 4\\-5-6\beta\end{pmatrix}\right)$$

$$\psi_1'(\beta) = 0 \text{ iff } \beta = \frac{5}{4}.$$

$\beta_1 := \frac{5}{4}$

There for $\quad x_2 = x_1 - \beta_1 s_1$

$$= \frac{1}{5}\begin{pmatrix}-1\\-1\end{pmatrix} - \frac{5}{4}\begin{pmatrix}1\\1\end{pmatrix}\left(\frac{1}{25}\begin{pmatrix}-4\\6\end{pmatrix}\right)$$

$$= \frac{1}{5}\begin{pmatrix}0\\ \frac{-1}{2}\end{pmatrix}$$

$$= \begin{pmatrix}0\\ \frac{-1}{2}\end{pmatrix}.$$

$s_2 = \varphi'(x_2) - \xi_2 s_1$

hence $\varphi'(x_2) = \varphi'\begin{pmatrix}0\\ \frac{-1}{2}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$

and also $\xi_2 = \dfrac{-1}{\|\varphi'(x_1)\|^2}\langle\varphi'(x_2), \varphi'(x_2) - \varphi'(x_1)\rangle = 0$,hence $s_2 = 0$.

Implies $\quad x_3 = x_2 - \beta_2 s_2 = x_2 = \begin{pmatrix}0\\ \frac{-1}{2}\end{pmatrix}.$

therefore $x_k = x_2\ \forall k \geq 2$

Hence $\quad x_k \xrightarrow{k\to\infty} \begin{pmatrix}0\\ \frac{-1}{2}\end{pmatrix}.$

Therefore $\begin{pmatrix}0\\ \frac{-1}{2}\end{pmatrix}$ is the minimal point of φ.

Example 3.6: Let $\Phi(x) = (\|x\| - 1)[1 - \ln(1 - \|x\|)]$, $x \in U := \{y \in \Re^2 | \, \|y\| \leq c_0 < 1\}$, $0 < c_0$.

By example 1.2.3: Φ is strongly convex.

$$\Phi'(x) = -\frac{\ln(1 - \|x\|)}{\|x\|}x.$$

For all $x \in U \backslash \{0\}$, $\dfrac{\ln(1 - c_0)}{c_0} \leq \dfrac{\ln(1 - \|x\|)}{\|x\|}$

$$\Rightarrow\ 0 < -\frac{\ln(1-\|x\|)}{\|x\|} \le -\frac{\ln(1-c_0)}{c_0}$$

$$\Rightarrow\ \|\Phi'(x)-\Phi'(y)\| = \left\|\frac{\ln(1-\|x\|)}{\|x\|}x - \frac{\ln(1-\|y\|)}{\|y\|}y\right\| \le -\frac{\ln(1-c_0)}{c_0}\|x-y\|\ ,$$

$\forall (x,y) \in U \times U.$

$\Rightarrow \Phi \in C^{1,1}(\Re^2)$

Let $x_0 = \frac{1}{2}(1,1)^T$.

$$S_0 := \Phi'(x_0) = -\frac{\sqrt{2}}{2}\ln\left(1-\frac{\sqrt{2}}{2}\right)\begin{pmatrix}1\\1\end{pmatrix}.$$

$\psi_1(\beta) := \Phi(x_0 - \beta s_0), \beta \in [0.\infty)$

$\Rightarrow\ \psi_1'(\beta) = \langle \Phi'(x_0-\beta s_0), s_0\rangle.$

$$\psi_1'(\beta) = 0 \Rightarrow \frac{\ln(1-\|x_0-\beta s_0\|)}{\|x_0-\beta s_0\|}\langle x_0 - \beta s_0\rangle = 0$$

$$\Rightarrow\ \beta = \frac{\langle x_0, s_0\rangle}{\langle s_0, s_0\rangle} = \frac{-1}{\sqrt{2}\ln\left(1-\frac{\sqrt{2}}{2}\right)} > 0\,.$$

$$\beta_0 := \frac{-1}{\sqrt{2}\ln\left(1-\frac{1}{\sqrt{2}}\right)}.$$

$x_1 := x_0 - \beta_0 s_0 = 0$

$\Rightarrow\quad \Phi'(x_1) = 0\,.$

Hence $\xi_1 := \frac{-1}{\|\Phi'(x_0)\|^2}\langle \Phi'(x_1), \Phi'(x_1)-\Phi'(x_0)\rangle = 0$ and $s_1 := \Phi'(x_1) - \xi_1\Phi'(x_0) = 0\,.$

Implies $x_2 := x_1 - \beta_1 s_1 = x_1$ .And also $x_k = x_1, \forall k \ge 1.$

So $x_1 = (0,0)^T$ is the minimal of Φ.

B.Second Method

Still the method is serving only for strongly convex functions.

$$\text{Let } \xi_k := \frac{\langle \varphi''(x_k)s_{k-1}, \varphi'(x_h)\rangle}{\langle \varphi''(x_k)s_{k-1}, s_{k-1}\rangle} \tag{3.10}$$

Theorem 3

Let 1. $\varphi.\in C^2(\Re^n)$ and strongly convex,

2. $\|\varphi''(x)\| \le v < \infty$, $\forall x \in N_\alpha(\varphi)$,

3. $\varphi. \in C^{1,1}(\Re^n)$,

4. ξ_k be calculated using formula (3.10)

5. the sequence (x_k) is constructed by i., ii. and iii.,

Then

$$\varphi(x_m) - \varphi(x^*) \le \mu \exp\{-c_2 m\}$$

$$\|x_m - x^*\| \le \frac{2}{\rho}\mu \exp\{-c_2 m\},$$

where $0 < c_2 \le \dfrac{\rho}{2 < (\rho + v)^2}$,$m\in N$.

Proof: Let $\psi(t) := \varphi'(x + \varepsilon t y)$ $\quad 0 \le t \le 1$

by the mean value theorem for differentiable function ψ

$$\psi(1) - \psi(0) = \psi'(t_0), \text{ for some to } \in (0, 1).$$

implies $\quad \varphi'(x + \xi y) - \varphi'(x) = \varphi''(x + \varepsilon t_0 y)(\varepsilon y),$

hence $\quad \langle \varphi''(x + \varepsilon to\, y)(y), y\rangle = \langle \varphi'(x + \varepsilon y) - \varphi'(x), y\rangle$

$$\ge \rho\|y\|^2.$$

The inequality follows form strong convexity of φ , Lemma 1.3.2, moreover it holds for all $\varepsilon \in [0, \infty)$.

Letting $\varepsilon \to 0$ we obtain

$$\langle \varphi''(x)y, y\rangle \ge \rho\|y\|^2, \text{ for all } y \in \Re^n.$$

$$|\xi_k| = \frac{|\langle \varphi''(x_k)s_{k-1}, \varphi(x_k)\rangle|}{|\langle \varphi''(x_k)s_{k-1}, s_{k-1}\rangle|}$$

$$\leq \frac{\|\varphi'(x_k)\| \, \|s_{k-1}\| \|\varphi'(x_k)\|}{\rho \|s_{k-1}\|^2} = \frac{\nu}{\rho} \frac{\|\varphi'(x_k)\|}{\|s_{k-1}\|}.$$

By Lemma 3.4, we have $\alpha_k \geq \dfrac{\rho}{\rho + \nu}$ $\forall k \geq 0$.

The conclusion of the proof follows from Remark 3.2. //

Example3.7: $\varphi(x) = \sum_{j=1}^{n} \|x - x_i\|^2$, where $x \in \Re^n$ and $x_i \in \Re^n$ $i \in \{1,\ldots,n\}$ are fixed.

Obviously $\varphi(x) = n \langle x, x\rangle - 2 \langle x, \sum_{j=1}^{n} x_i \rangle + \sum_{i=1}^{n} \langle x_i, x_i\rangle$, implies φ quadratic form with

$\varphi''(x) = In$ (In is $n \times n$ identity matrix). Hence φ is strongly convex.

$$\varphi'(x) = 2nx - 2\sum_{j=1}^{n} x_i$$

Implies $\|\varphi'(x) - \varphi'(y)\| = \|2n(x - 2)\| = 2n \|x - y\|$, $\forall (x, y) \in \Re^n$.

Let n=3 and $x_0 = (1,1,1)$

Let $x_1 = \left(\frac{1}{3},0,0\right)^T$, $x_2 = \left(0,\frac{1}{3},0\right)^T$ $x_3 = \left(0,0,\frac{1}{3},\right)^T$

There for $\varphi(x) = 3 \langle x, x \rangle - \frac{2}{3}\langle x, (1, 1, 1)^T\rangle + \frac{1}{3}$

and also $\varphi'(x) = 6x - \frac{2}{3}(1, 1, 1)^T$, $\varphi''(x) = 6\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

$$s_0 := \varphi'(x_0) = 6(1, 1, 1)^T - \frac{2}{3}(1, 1, 1)^T$$

$$= \frac{16}{3}(1, 1, 1)^T.$$

$x_1 := x_0 - \beta_0 s_0$ where $\beta_0 \in M(\varphi(x_0 - \beta s_0), \beta \geq 0)$

$$\psi_1(\beta) = \varphi(x_0 - \beta s_0),\ x_0 - \beta s_0 = \left(\frac{3 - 16\beta}{3}\right)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

$$\Rightarrow \psi_1(\beta) = (3 - 16\beta)^2 - \frac{2}{3}(3 - 16\beta) + \frac{1}{3}$$

$$\Rightarrow \quad \psi_1'(\beta) = (2)(-16)\,(3 - 16\beta) - \frac{2}{3}(-16)\,.$$

$$\psi_1'(\beta) = 0 \text{ iff } \beta = \frac{1}{6}$$

$$\beta_0 := \frac{1}{6}$$

$$\Rightarrow \quad x_1 = \begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{1}{6}\left(\frac{16}{3}\right)\begin{pmatrix}1\\1\\1\end{pmatrix} = \frac{1}{9}(1, 1, 1)^T.$$

$$\Rightarrow \quad \varphi'(x_1) = 6 \times \frac{1}{9}(1, 1, 1)^T - \frac{2}{3}(1, 1, 1)^T = 0$$

$$\Rightarrow \qquad \xi_1 = \frac{\langle \varphi''(x_1)s_0, \varphi'(x_1)\rangle}{\langle \varphi''(x)s_0, s_0\rangle} = 0$$

There fore $\quad s_1 = \varphi'(x_1) - \xi_1 s_1 = 0.$

Implies $\quad x_2 = x_1 - \beta_s s_1 = x_1$

hence we have $x_k = x_1$ for all $k \in N$.

Therefore $\frac{1}{9}\begin{pmatrix}1\\1\\1\end{pmatrix}$ is a minima of φ.

References

1. V.G. Karmanov:
   Mathematical Programming,
   1989 Revised Edition

2. Deumlich, R:
   Optimization and Theory of Approximation,
   Textbook for lecture,
   Addis Ababa, 1997

3. Deumlich,R.:
   Functional Analysis I,
   Textbook for the lecture,
   Addis Ababa 1997

4. Jean-Baptiste and Hiriart-Urrutr:
   Convex Analysis and minimizations Algorisms I