GRADUATE SEMINAR REPORT

ON

CONSTRAINED FUNCTION MINIMIZATION

BY
JEMA HAJI
Advisor
Prof.Dr.rer. nat. habil.R.Deumlich

A.A.U.
JUNE 1998

CONTENTS

ACKNOWLEDGMENT

It is really difficult with out the day to day follow up of my respected advisor prof. Dr. rer. nat. habil. R. Deumlich that this seminar stood in its foot. His fatherly advice not only in the field of study but in all other social aspects is also unforgottable.

I am also grateful to many of my colleagues in the department of mathematics who shared their ideas and gave me some reference materials.

PREFACE

It is a reality that nature is full of constraints. This seminar describes various methods of function minimization with constraints on the variables. This we can do by following or developing certain algorithms.

Algorithms are inventions which very often appear to have little or nothing in common with one another. As a result it was held for a long time that a coherent theory of algorithms could not be constructed . The last few years have shown that this belief was incorrect and that most convergent algorithms share certain basic properties and hence that a unified approach to algorithms is impossible.

This paper presents a few number of optimizations which deals with algorithm convergence and implementation. The algorithms set forth below are iterative in character. This means we can construct finite or an infinite sequence of points x_k, k = 0, 1 - - - which is said to converge to the solving of minimization problem.

The sequence of points are related by the equation $x_{k+1} = x_k + \alpha_k p_k$, where p_k is the direction vector and α_k is a step along the direction p_k. There fore the description of any of the algorithm given below consists in estabileshing the method of choosing the vector pk and the length of the step α_k. It should be noted that the method of choosing the vector p_k determines the general rate of convergence of the process and the method of choosing α_k has an important influence on the amount of calculations at each iteration.

This seminar is a compilation of two seminars I have delivered for a qualification for M.Sc in mathematics. It is divided in to five chapters.

Chapter I . Preliminaries: A review of definitions of related terms and concepts.

Chapter II. Develops methods of solving problems of quadratic programming which is a subsidiary problems in many algorithms.

Chapter III, IV and V . Describes the algorithms (and their convergence) for solving problems of convex and non-convex programming.

CHAPTER ONE

PRELIMINARIES (REVIEW)

1.1. Convex Sets and Functions

Definitions 1. A set $K \subseteq R^n$ is said to be convex if and only if $\lambda x + (1-\lambda)y \in K$ for each $x,y \in k$ and $\lambda \in [0,1]$.

2. A function $f: R^n \to, R \cup \{\infty\}$ not identically ∞ is said to be convex if and only if for each $x, y \in R^n$ and $\lambda \in [o,1]$ there holds $f(\lambda x + (1-\lambda)y) \leq f(x) + (1-\lambda) f(y)$.
3. An (n,n) matrix is said to be positive definite if and only if $<Ax, x> \geq 0 \ \forall x \in R^n$

Corollary 1. The quadratic function $f(x) = ½ < Ax, x> + < b, x> + c$ is convex iff the matrix A is positive definite.

1.2. Necessary and Sufficient Condition for a Minimum

The problem of quadratic programming is formulated as:

$f(x) = ½ < Ax, x > + < b, x > \quad \to \quad \min, x \in s$

$S = \{ x : < a_i, x > - b_i \leq 0, i \in y^- , <a_i, x> - b_i = 0^i \in y^{0\}}$ where $x \in E^n$, $a_i \in E^n$, $i \in y^- \cup y^0$, $b \in E^n$, b_i are constants, A is an n x n symmetric and positive definite matrix and y^o and y^- are finite set of indices.

Theorem 1 . x^* is a minimum point of f(x) or a solution of (1) iff there are numbers U^{i*}, $i \in y^0 \cup y^-$

$$Ax^* + b + \sum_{i \in y^- \cup y^o} u^{i*} a_i = 0 \text{ if } < a_i, x^* > - b_i < 0 \ i \in y^- \cup y^o \ i \in y^- , u^{i*} 0, i \in y^-$$

1.3. Conjugate Gradient Method

Definition: A finite set of vectors $p_1, p_2, \dots, p_r$ are said to be conjugate with respect to H iff $p_i^T H^p{}_j = 0, \ \forall i \neq j, i, j = 1, \dots, r$.

Theorem 2. Let f(x) ,$x \in E^n$ be a quadratic form with positive definite matrix H as its hessian , and let $P_1, P_2, \dots, P_n$ be a set of non zero vectors conjugate with sespect to H.

Then starting from any point x, and with successive steps in the direction $P_1, P_2 \dots, P_n$ each appropriate length given by the algorithm $x_{k+1} = x_k + \lambda_k P_k, \lambda_k = \dfrac{p_k^T f^l{}_k}{p_k^T H P_k}$

the minimum point x^* is reached in atmost n steps, that is $x_{r+1} = x^*$, $r \leq n$

The conjugate gradient algorithm is as follows:

Choose x_1 ; $P_1 = -f'(x_1) =: f'$

$$x_{k+1} = x_k + \alpha_k p_k$$

$$p_{k+1} = -f'_{k+1} + \frac{\|f'_{k+1}\|^2}{\|f'^2_k\|} p_k$$

$$\alpha_{k+1} = \frac{-< f'_{k+1}, P_{k+1}>}{< Ap_{K+1}, p_{k+1}>}$$

1.4. **Taylor's Formula**

Taylor's exansion of second order: Let f(x) be a twice continuously differentiable function and let $x = x_k + \alpha p$. Then the expansion of the function in Taylor's series about x_k gives:

$$f(x) = f(x_k) + \alpha < f'(x_k), P > + \frac{\alpha^2}{2} < f''(x_{kc}) P, P >$$

Where $x_{kc} = x_k + \theta (x - x_k)$ and $\theta \varepsilon [0, 1]$

1.5. **Pates of Convergence**

Definitions 1) We say that a sequence (x_k) converges to a point x* at a linear rate orat a rate of geometric progression (with ratio q) if from a certain k the inequality $\| x_{k+1} - x^* \| \leq q \| x_k - x^* \|$ where $0 < q < 1$

2) If the inequality $\|x_{k+1} - x^*\| \leq q_k \|x_k - x^*\|$ is satisfed where $q_k \to 0$ as $k \to \infty$, then we say that rate of convergence of the sequence (x_k) is super linear, or faster than the rate of convergence of any geometric progression.

3) If $q_k \leq C \|x_k - x^*\| \to 0$, then $\|x_{k+1} - x^*\| \leq C \|x_k - x^*\|^2$ This estimate is characterstic of quadratic rate of convergence.

1.6. **Characterizing Theorem of Convex Optimization**

Theorem 3. Let f(x) be a continuously differentiable convex function in a convex set K. Then $x^* \in K$ is a minimum point of f(x) in K iff the inequality $< f'(x^*), x - x^* > \geq 0, x \in K$ is satisfied.

1.7. **Weierstrass Theorem**

Thearmem 4 In a compact set, a continuous function attains its minimum

1.8. **Sub gradient Inequality**

Lemma 1 . Let $f(x)$ be a continuously differentiable function whose continuous gradient is $f'(x)$.Then $f(x)$ is convex iff $f(x_1) - f(x_2) \geq < f'(x_1), x_2-x_1 >$ for all $x_1, x_2 \in E^n$.

1. 9 . **Strongly Convex Function**

Definition: A function $f(x)$ is said to be strongly convex iff for any $x_1, x_2 \in E^n$,

$$f\left(\frac{x_1 + x_2}{2}\right) \leq \frac{1}{2} [f(x_1) + f(x_2)] - \gamma \| x_2 - x \|^2$$ where $\gamma > 0$ is an arbitrary small constant.

Lemma 2 . If $f(X)$ is a twice continuously differentiable function, then the condition of strong convexiry is equivalent to the condition $m \| p \|^2 \leq < f(x) p,p > \leq M \| P \|^2, m > O$ for any x and $P \in E^n$.

1.10. **Lagranges Formula for Operators**

If $F(x)$ is a non linear differentiable operator then for $x, y \in E^n$ the following formula is valid:

$$< F(x + h) - F(x), y > + < F'(x + \theta h) h, y > , 0 \leq \theta \leq 1$$

CHAPTER TW0

CONSTRAINED QUADRATIC PROGRAMMING PROBLEM

2. 1 INTRODUCTION

Some times the problem of variables with in a region are not entirely free, but must satisfy certain bounds or functional relation ships. This auxilairy conditions are known as constraints.

Usually constraints represent either system limitations or the physical and economic laws that the variables must satisfy.

This chapter is aimed at finding a solution to a non linear (Quadratic programming problem) with constraints on the variables as linear inequalities and equations.

The different methods of finding a solution to optimization problems such as the steepest decent, Newtons, dual directions and conjugate gradients are all applied to optimization problems with out constraints. In order to apply these methods we must reduce the given quadratic programming problem with constraints to a quadratic programming problem with out constraints.

To do so we first consider a quadratic programming problem with constraints satisfied as equalities and reduce it to a quadratic programming problem with out constraints and then apply one of the methods to find a solution of a quadratic programming problem with out constraints. Now since these equality constraints determine a certain face of a polyhedral set defined by linear inequalities we find the minimum point of f(x) on this face by the method of conjugate gradient (we have choosen here this method because it has much faster convergence and is convenient computationally).

If the point obtained is the minimum point of f(x), then we are done. If not, we again apply the method on the other face and the process is repeated.

The process will converge in a finite number of steps since by the method of conjugate gradient the minimum point is obtained in a finite number of steps and the number of faces of a polyhedral face is limited.

2.2. Problem of Quadratic Programming

A quadratic programming problem is a non linear programming problem with a quadratic objective function and linear constraints.

Thus the problem of quadratic programming is the minimization of a quadratic function

$$f(x) = \tfrac{1}{2} < x, Cx> + < d, x > \text{ --- } (2.1)$$

with the following constraints

$$< a_i, x> - b_i \leq 0 \qquad i \in y^-$$

$$< a_i, x > - b_i = 0 \qquad i \in y^0 \text{ --- } (2.2)$$

where $x \in E^n$ $a_i \in E^n$, $i \in y^- U y^o$, $d \in E^n$, b_i are numbers C is an nxn symmetric, positive definite matrix and y^- and y^o are finite set of indices.

2.2. 1. Operators of Projection

Definition A linear mapping $P : X \to X$, X a vector space is said to be a projection in x iff $p^2 = P$.

Let $y = y^o \; U \; y^-$ and Y be a subset of the set of indices Y. We form a matrix A_Y whose rows are a_i, $i \in Y$, so that the matrix is mxn dimensional where m is the number of elements in set Y.

Lemma 1 . If the vectors a_i, $i \in Y$ are linearly independent then the matrix $A_Y A^T_Y$ is non singular.

Proof: Let $y \in E^m$ be non zero and $A_Y A_Y y = 0$

Then $y^T A_Y A^T_Y y = (A^T_Y y)^T A^T_Y y = < A^T_Y y, A^T_Y y > = \|A_Y^T y\|^2 \quad = \quad 0$

This implies $A_Y^T y = 0$ or $\sum_{i \in y} a_i y^i = 0$

But this is impossible as the vectors a_i are linearly independent and y is non zero vector

Thus the matrix $A_Y A_Y^T$ can be made zero only by a zero vector.

There fore $A_Y A_Y^T$ is non singular / /

Define an operator p by $P: = A_Y^T (A_Y A_Y^T)^{-1} A_Y$ --- (2. 3)

Then P has the following properties

i) P is a projection, ie $p^2 = P$ --- (2.4)

Proof: $P^2 = P.P = A_Y^T (A_Y A_Y^T)^{-1} A_Y A_Y^T (A_Y A_Y^T)^{-1} A_Y$

$= A_Y^T A_Y^{T-1} A_Y^{-1} A_Y A_Y^T (A_Y A_Y^T)^{-1} A_Y$

$= A_Y^T (A_Y A_Y^T)^{-1} A_Y = P$

ii) P is a symmetric matrix. ie $P^T = p$--- (2. 5)

Proof; $P^T = (A_Y^T (A_Y A_Y^T)^{-1} A_Y)^T$

$= Ay^T [(Ay Ay^T)^{-1} Ay$

$= A_Y^T ((A_Y A_Y^T)^{-1})^T (A_Y^T)^{-1}]^T (A_Y^T)^T$

$= A_Y (A_Y A_Y^T)^{-1} A_Y$

$= P$

iii) $P(I - P) = (I - P)P = 0$ --- (2.6)

Proof: $P(I - P) = PI - PP = P - P$ (= 0

iv) Px and (I - P) x are orthogonal resolution of vector x.

Proof: For any vector $x \in E^n$, we have $x = Px + (I - P) x$

and $< Px, (I - P) x > = < x, P^T (I - P)x >$.

$= < x, P(I - P)x >$ (By ii)

$= < x, 0 >$ (By iii)

$= 0$

Hence Px and (I - P) x are orthogonal resolutions of a vector x.

V) Px lies in the subspace spanned by the vectors a_i , $i \in y$

Proof : $Px = A_Y^T (A_Y A_Y^T)^{-1} A_Y x = A_Y^T U$ where $U = (A_Y A_Y^T)^{-1} A_Y^X$

$= \sum_{i \in y} a_i U^i$

This means Px is a linear combination of the vectors a_i .

Hence Px lies in the subspace spanned by the vectors a_i , $i \in y$.

Vi) $A (I - P) = 0$ --- (2.7)

Proof: $A (I - P) = A_Y - A_Y P = A_Y - A_Y (A_Y^T (A_Y A_Y^T)^{-1} A_Y)$

$= A_Y - (A_Y A_Y) (A_Y A_Y^T)^{-1} A_Y$ (But $A_Y A_Y) (A_Y A_Y^T)^{-1} = I$)

$= A_Y - IA_Y$

$= A_Y - A_Y = 0$

There for for any $x \in E^n$ the vector $y = (I - P)x$ satisfies the system of equations $A_Y y = 0$

2.2. 2. **Minimization of Quadratic Function in a Subspace**

The basis of numerical method of solving a quadratic programming problem (2. 1) and (2.2) is the method of conjugate gradients. The main idea of the application of this method is as follows:

Let x_o be a point which satisfies (2. 2). We pick out among the constraints which are satisfied as equalities. These constraints determine a certain face of a polyhedral set defined by linear inequalities (2. 2). We find the minimum of f(x) on this face using the method of conjugate gradients. The point obtained is a solution of our problem or indicates a transition to a new face and then the procedure is repeated. Since the method of conjugate gradients minimizes function f(x) after a finite number of steps, and the number of faces of a polyhedral set is limited an algorithm of this kind converges after a finite number steps.

Suppose now that we have to minimize a quadratic function $f(x) = \frac{1}{2} <x,Cx> + < d, x>$ with constraints $<a_i , x > -b_i = 0 \; i \in y$ - - - (2. 8)

Assume the vectors a_i, $i \in y$ are linearly independent and xo satisfies (2. 8)

Then $A_Y x_o - b_y = 0$ Where b_Y is a vector whose components are b_i .

We now introduce a new variable y defined as follows: $x = x_0 + (I - P)\, y$ - - - (2.9)

and consider the quadratic function $\varphi(y) = f(x_0 + (I - P)\, y)$

Thene $\varphi'(y) = f'(x_0 + (I - P)\,y)\,(I - P) = (I - P)\, f'(x)$ - - - (2.10)

Lemma 2 If y is a point of absolute minimum of $\varphi'(\overline{y})$ then the corresponding poi $x = x_0 + (I - P)\, y$ is the minimum point of f(x) with constraints(2. 8).

Proof: Since Y is the minimum point of $\varphi(y)$ we have $\varphi(\overline{y}) = 0$

$\Rightarrow (I - P)\, f'(x) = 0$

$\Rightarrow f'(x) - P\, f'(x) = 0$

$\Rightarrow f'(x) - A^T_Y (A_Y A^T_Y)^{-1} A_Y f'(x) = 0$

Letting $U := -(A_Y A^{T}_{Y})^{-1} A_Y f'(x)$ we have $f'(x) + A^T_Y U = 0$ - - - (2. 11)

More over $A_Y \overline{x} = A_Y (x_o + (I - P)\, y) = A_Y x_o + A_Y (I - P)y = 0$ by (2. 7)

$= A_Y X_0$

$= b_Y$

$\Rightarrow$ $\bar{x}$ satisfies (2. 8).

Hence x is a feasible point and at this point condition (2.11) is satisfied which is the necessary and sufficient condition for x to be the minimum point of f(x) with condition (2.8). / /

This shows that the problem under consideration can be reduced to the minimization of a quadratic function φ(y) with out constraints.

Now we apply the method of conjugate gradients for the minimization of a quadratic function y(y) with out constraints.

Start with $y_0 = 0$, $p_1 = \varphi'_0$

$$y_{K+1} = y_K + \alpha_{K+1} P_{K+1}$$

$$p_{K+1} = -\varphi'_k + \frac{\|\varphi'_K\|^2}{\|\varphi'_{k-1}\|^2} P_K$$

$$x_{K+1} = \frac{<\varphi'_k, P_{K+1}>}{<P_{K+1}, (I-P)\, C\, (I-P)\, P_{K+1}>}$$

Where (I - P) C(I-P) is a matrix which determine the quadratic term of the function φ(y). These formulas determine the process involving the additional variable y . But we can go back to the original variable x. To this end we first show that(I - P) $P_K = P_K$ for all k

This we can do by induction on k

If k = 1, then $(I-P)P_1 = -(I-P)\varphi'(o) = -(I-P)(I-P)f'(x_o)$

$$= -(I-P)f'(x)$$
$$= -(\varphi'(o))$$
$$= P_1$$

We assume it is true for k and prove that it is true for k + 1.

$$(I-P)P_{K+1} = -(I-P)\varphi'(y_k) + \frac{\|\varphi'(y_k)\|^2}{\|\varphi'(y_{k-1})\|^2}(I-P)P_k$$

$$= -(I-P)(I-P)f'(x_k) + \frac{\|\varphi'(y_k)\|^2}{\|\varphi'(y_{k-1})\|^2} p_k$$

$$= -(I-P)f'(x_k) + \frac{\|\varphi(y_k)\|^2}{\|\varphi(y_{k-1})\|^2}$$

$$= \varphi'(y_k) + \frac{\|\varphi'(y_k)\|^2}{\|\varphi'(y_{k-1})\|^2} P_k$$

$$= -p_{k+1}$$

Hence by induction principle we have $(I-P)p_k = p_k$ for all k.

Then $x_{k+1} = x_o + (I-P)\, y_{k+1}$

$$= x_o + (I - P)\, y_k - (I-p)y_k + (I - p)\, y_{k+1}$$

$$= x_k + (I - p)\,(y_{k+1} - y_k)$$

$$= x_k + \alpha_{k+1}\, p_{k+1}(I - p)$$

$$= x_k + \alpha_{k+1}\, p_{k+1}$$

$$p_{k+1} = (I - P)\, f'(x_k) + \frac{\| (I - P)\, f'(x_k) \|^2}{\| (I - P)\, f'(x_{k-1}) \|^2}$$

$$\text{and } \alpha_{k+1} = -\frac{< (I - P)\, f'(x_k),\, p_{k+1} >}{< (I- P)\, P_{k+1},\, C(I- P)\, p_{k+1} >}$$

$$= -\frac{(I - P) < f'(x_k),\, P_{k+1} >}{(I - P)^2 < p_{k+1},\, Cp_{k+1} >}$$

$$= -\frac{< f'(x_k),\, p_{k+1} >}{< p_{k+1},\, CP_{k+1} >}$$

Then we proved the following theorem

Theorem 1. The problem of minimization of a quadratic function f(x) with constraints (2. 2) given the initial point x_o which satisfies (2.2), is solved after a finite number of steps by the following process:

$$p_1 = (I - P)\, f'(x_0)$$

$$x_{k+1} = x_k + \alpha_{k+1}\, p_{k+1}$$

$$pk + 1 = -(I - P)\, f'(xk) + \frac{\| (I- P)\, f'(x_k) \|^2}{\| (I-P) f'(x_{k-1}) \|^2} P_k$$

$$\alpha_{k+1} = -\frac{< f'(x_k),\, p_{k+1} >}{< p_{k+1},\, CP_{k+1} >} \qquad k = 0, 1, ---$$

2.2. 3. Algorithim of General Problem of Quadratic Programming

Consider the general quadratic programming problem (2.1) and (2.2.).

For each point x satisfying (2.2), set $Y(x) = \{ i : < a_i , x > - b_i = 0, i \in y^- \cup y^0 \}$

Assume again that the vectors a_i , $i \in y(x)$ are linearly independent. Let x_0 be an arbitrary point satisfying (2.2) and is the first approximation.

Take a set of indices $Y_0 = Y(x_0)$ and construct the operator P_{Y0}:

$$py_{0} := A_{Y}{}^{T}{}_{0}(A_{Yo}\, A_{Y}{}^{T}{}_{0})^{-1}\, Ay_0$$

If we put $U_0 := -(A_{Y0}\, A_{Y}{}^{T}{}_{0})^{-1}\, A_{Y0}\, f'(x_o)$, then $(I - P_{Y0})\, f(x_0) = f'(x_0) - p_{Y0}\, f'(x_0)$

$$= f'(x_0) + A^T{}_{Y0}\, U_0$$

Now there are two possible cases

Case 1. $(I - P_{Y0}) f'(x_0) = 0$

$\Rightarrow \; f'(xo) + A^T_{Y0}\, U_0 = 0$ - - - (2.12)

$\Rightarrow \; x_0$ is the minimum point of f(x) on the face defined by the system of equations $< a_i , x > - b_i = 0 , i \in y^0$ (by theorm 1.3.1).

If i) $U_0^{\,i} \geq \; 0 \; \forall \in y^0 n y^-$

Then x_0 is a solution of (2.1) and 2.2) by theorem 1.3.1

ii) There is an index $j \in y^0 n y^-$ such that $U_0^{\,i} < 0$

In this case we construct a new set of indices Y'_0 by delting index j and apply the method of conjugate gradients to solving the problem of minimization of f(x) with constraints

$<ai, x> - b_i = 0, \; i \in Y'_0$ - - - (2. 13)

How ever in applying this method, the process must not go beyond the limits of (2. 2). There fore a every step of the algorith the following check should be made compute the quantity

$$\overline{\alpha_{k+1}} = \min \; \frac{b_i - < a_i , x_k >}{< a_i , p_{k+1} >} \quad \text{- - - (2. 14)}$$

where the minimum is taken over all i for which $< a_i \, , \, p_{\,k\,+\,1} > 0$, x_k - point just constructed by the algorithm and p_{k+1} is the conjugate direction. Let now α_{k+1} be the corresponding step length in the method of conjugate gradients.

Then there are to cases

case i. $\alpha_{k+1} < \overline{\alpha_{k+1}}$

$$\alpha_{k+1} < \overline{\alpha_{k+1}} = \min \; \frac{b_i - < a_i, x_k >}{< a_i \, , p_{k+1} >} \leq \frac{b_i - < a_i, x_k >}{< a_i, p_{k+1} >}$$

$\Rightarrow \; \alpha_{k+1} < a_i \, , p_{k+1} > \leq b_i - < a_i, x_k >$ (since $<a_i, p_{k+1} > >0$)

$\Rightarrow \; < a_i, \alpha_{k+1} \, p_{k+1} > + <a_i, x_k> - b_i \leq 0$

$\Rightarrow < a_i, x_k + \alpha_{k+1} \, p_{k+1} > - b_i \leq 0$

$\Rightarrow x_{k+1} = x_k + \alpha_{k+1} \, p_{k+1}$ satisfies condition (2.2) and the process goes on.

case ii. $\alpha_{\,k+1} \geq \overline{\alpha_{k+1}}$

In this case $x_{\,k+1} = xk + \alpha_{k+1} \, p_{k+1}$ satisfies condition(2.2) . This is because

$$<a_i, xk + \alpha_{k+1} \, p_{k+1} > - b_i = < a_i, x_k > + \min \frac{b_i - < a_i, x_k >}{< a_i, p_{k+1} >} P_{k+1} - b_i$$

$$\leq a_i, x_k > + \leq \frac{b_i - <a_i, x_k >}{< a_i \, , p_{k+1} >} \; p_{\,k+1} > - b_i$$

$$= < a_i \, , x_k > - b_i + \frac{b_i - < a_i, x_k >}{<a_i, P_{k+1} >} <a_i, p_{k+1} >$$

$\Rightarrow p_i = < a_i, x_k> - b_i + b_i - a_i, x_k>$

More over $\varphi'(=\ 0$

And the process stops in this case.

Thus either we find the minimum point of f(x) under conditions (2. 13) or will be truncated which $\alpha_{k+1} \geq \overline{\alpha}_{k+1}$. In both cases we take the point obtained to be the initial point and proceed using the new point as we did with intial one , x_o .

Case 2. $(I - p_{Y0})\, f'(xo) \neq 0$

In this case we apply the method of conjugate gradients to solving the problem of minimization of f(x) with constraints $< a_i, x > - b_i = 0 \quad i \in y_0$ - - - (2. 15)

Starting with point x_o and checking at every step whether the points obtained are feasible or not, ie we calculate α_{k+1} by formula (2.14) and apply the method of conjugate gradients until either we find the minimum point of f(x) with constraints (2.15) or the condition (2. 15) or the condition $x_{k+1} = x_k + \alpha_{k+1}\, p_{k+1}$ is obtained. In both cases we take the point obtained as the initial one and repeat at it the operations performed with x_o.

Convergence of the method

Since new points are obtained by the method of conjugate gradients and in this case the function decreases at each step it suffice to check only that $\alpha_{k+1} > 0$ always. ie constraint (2. 2) permit to take a non zero step in the direction choosen P_{k+1}.

This is because $\overline{\alpha}_{k+1} > 0 \Rightarrow \min \dfrac{b_i - <a_i, x_k>}{<a_i, p_{k+1}>} > 0$

$$\Rightarrow \frac{b_i - < a_i, x_k>}{< a_i, p_{k+1}>} > 0$$

$$\Rightarrow b_i - < a_i, x_k> > 0 \text{ (since } < a_i, p_{k+1} > < 0 \text{)}$$

$$\Rightarrow <a_i , x_k > - b_i < 0$$

More over in case(1) point x_0 is not the minimum point of f(x) with constraints (2. 13) for if it were so the method of conjugate gradients would not have moved the process from point x_o.

Lemma 3. Vector $p_1 = - (I - P_{Y0})\, f'(x_o)$ is the solution of minimizing the function

$$\varphi(p) = < f'(x_o), p > + ½ \|P\|^2 \text{ with constraints}$$

$$A_{Y0}\, p = 0 \text{ - - - (2. 16) and the minimum value is } - ½ \| (I - P_{Y0})\, f'(xo) \|^2$$

Proof: Since by (2. 7) $A_Y(I - P) = 0$, we have $A_{Y0}P_1 = - A_{Y0}(I - P_{Y0})f'(x_o) = 0$

$\Rightarrow p_1$ satisfies (2. 16)

More over $\varphi'(p) = f'(x_o) + p$

$$\Rightarrow \varphi'(p_1) = p_1 + f'(x_o)$$

$$= -(I - P_{Y0}) f'(x_o) f'(x_o)$$

$$= -f'(x_o) + p_{Y0} f'(x_o) + f'(x_o)$$

$$= p_{Y0} f'(x_o)$$

$$= A_Y o f'(x_o)$$

$$= -A^T_{Y0} U_0$$

where $U_0 = -(A_{Yo} A^T_{Y0})^{-1} A_{Yo} f'(x_o)$

$$\Rightarrow \quad \varphi'(p_1) + A^T_{Y0} U_0 = 0$$

Hence P_1 is the minimum point of $\varphi(p)$ with constraint (2.16)

To find the minimum value of $\varphi(p)$ we formulate a problem which is the dual of the problem of minimizing $\varphi(p)$ with constraint(2.16).

This is equivalent to finding the minimum of the function $\varphi(p) + U^T A_{Y0} p$.

Differentiating with respect to p and equating the derivative to zero we have

$$\varphi'(p) + A^T_{Y0} U = 0 \Rightarrow P + f'(x_o) + A^T_{Y0} U = 0$$

$$\Rightarrow p = -f'(x_o) + A^T_{Y0} U$$

Substituting this expression for p we have $\varphi(p) + U^T A_{Yo} p = < f'(x_o, p> + ½ \|p\|^2 + U^T A_{Y0} p$

$$= < f'(x_0), -f'(x_o) - A^T_Y o U > ½ < f'(x_o) + A_{Yo} U, f'(x_o) + A^T_{Yo} U > (-f'(x_o) - A^T_{Y0} U A^T_{Y0} U$$

$$= -\|f'(x_o)\|^2 < f'(x_o), A^T_Y o U > ½ \|f'(x_o)\|^2 + < f'(x_o), A^T_{Yo} U >$$

$$+ ½ < A^T_{Yo} U, A^T_Y U > - < f'(x_o), A^T_{Yo} U > - < A^T_{Yo} U, A^T_{Yo} U >$$

$$= -½ \|f'(x_o)\|^2 - < f'(x_o), A^T_{Yo} U > - ½ < A^T_{Yo} U, A^T_{Yo} U >$$

$$= -½ \|f'(x_o) + A^T_{Yo} U\|^2$$

Thus the dual problem consists in finding over all possible vectors U the maximum of the function

$$\varphi^* (U) = -½ \|f'(xo) + A_Y^T U\|^2$$

Now differentiating $\varphi^*(U)$ and equating the derivative to zero we have $\varphi^{*\prime}(U) = 0$

$$\Rightarrow \quad (-½ < f'(x_o) A^T_{Y0} U, f'(x_o) + A^T_{Y0} U >)^1 = 0$$

$\Rightarrow -(f'(x_o) + A^T_{Y0} U) A^T_{Y0} = 0 \Rightarrow f'(x_o) + A^T_{Yo} U = 0$

But $f'(x_o) + A_Y{}^T_o U o = f'(x_o) + A^T_{Yo}(-(A_{Yo} A^T_{Yo})^{-1} A_{Yo} f'(x_o))$

$= f'(x_o) + -A^T_{Yo} A^T_{Yo}{}^{-1} A_Y{}^{-1}_o A_Y o\, f'(x_o)$

$= f'(x_o) - f'(x_o) = 0$

$\Rightarrow U_0 = -(A_{Yo} A^T_{Yo})^{-1} A_{Yo} f'(x_o)$ is the solution of the dual problem. That means U_0 maximizes $\varphi^*(U)$.

Since U_0^i, $i \in Y_o$ are components of U_0, vector U_0 is a vector of lagrange multipliers in the problem of minimization of $\varphi(p)$ with constraints (2. 16).

Then we obtain the value of the minimum of $\varphi(p)$ with constraints (2. 16) to be equal to the maximum of $\varphi^*(U)$ over U is $-\frac{1}{2} \| f'(x_o) + A^T_{Yo} U_0 \|^2$ or $-\frac{1}{2} \|(I - pY_0) f'(x_0)\|^2$

Lemma 4. Let $A_Y{}'_o$ be the matrix formed from A_{Y0} by deleting the row with index i for which $U^i_o < 0$ and more over let $(I - p_{Yo}) f'(x_0) = 0$.

Then i) $\overline{P}_1 = -(I - p'_{Yo}) f'(x_o) \neq 0$

ii) $< a_i, \overline{p}_1 > < 0$

Proof i) We can express the vector $\overline{p}_1$ as $\overline{p}_1 = -(f'(x_o) + A^T_{Y0} V)$

where $V = -(A_Y{}'_o A^T_{Y'o})^{-1} A_Y{}'_o f'(x_o)$

Assume $\overline{p}_1 = 0$ Then $f'(xo) + A_Y{}'_o V = 0$ - - - (*)

But by assumption we have $(I - pY_0) f'(x_o) = 0$ Which implies $f'(x_o) + AY^T_o U_0 = 0$ - - - (**)

Subtracting (*) from (**) we obtain $A^T_{Y0} U_0 - A^T_{Yo} V = U_o{}^i a_j + \sum_{i \neq j} (U_o{}^i - V^i) a_i = 0$

But this is a contradiction to $U^i_o \neq 0$ (since $U^i o < 0$) and the vectors a_i, $i \,\varepsilon\, Y_o$ are linearly independent.

Hence $P_1 = -(f'(x_o) + A^T_{Yo} V) \neq 0$.

ii) To prove the second part of the lemma (**) can be written in component form as

$$f'(x_o) + \sum_{i \neq j} U_o{}^i a_i + (-U^j{}_o)(-a_j) = \text{- - -} \quad (2.17)$$

Since $U^i{}_o < 0$, we have $-U^i_o > 0$ Now consider the problem of minimizing

$\varphi(p) + < p, f'(x_o) > + \frac{1}{2} \|p\|^2$ with constraints $< a_i, p > \; = \; 0$, $i \in Y'_o$,

$-< a_j, P > \leq 0$ - - - (2.18)

Since $\varphi'(p) = f'(x_o) + p$, we have $\varphi'(o) = f'(x_o)$

There fore(2. 17) is the necessary and sufficient condition for the point $p = 0$ to be the solution of the problem of minimization of $\varphi(p)$ with constraints (2. 18).

On the other hand by lemma 2 $\overline{p_1}$ is the solution of the problem of minimization of $\varphi(p)$ with constraints $A_{Y_o} p = 0$ or in component form $< a_i, \overline{p}_1 > = 0$ - - - (2.19).

suppose that $< a_j, \overline{p_1} > \geq 0$ Since P_1 satisfies constraints (2. 19)

(as $A_{Y'0} (I - p_{Y'o}) f'(x_o) = o$) it satisfies constraint (2. 18) too.

But $< f'(x_o), p_1 > = - < f'(x_o) , (I - p_{Y'0}) f'(x_o) >$

$$= - < p_{Y'0} f'(x_o) + (I-p_{Y'0}f'(xo),(I-p_{Y'o}) f'(x_o) >$$

$$= - < (I - p_{Y'o}) f'(x_o), (I-p_{Y0}) f'(x_o) >$$

$$= - \|P_1\|^2$$

There fore $y(p1) = <f'(xo), p_1 > + ½\|\overline{p}_1\|^2$

$$= - ½\|\overline{p_1}\|^2 < 0$$

But this is a contradiction to the fact that the minimum value of $\varphi(p)$ with constraints (2.17) is attained with $p = 0$ and is equal to zero.

Hence $< a_j, \overline{p_1} > < 0$. / /

We now return to the algorithm constructed. Consider case(1) in the algorithm constructed and let point x_o be not the solution of the problem of quadratic programming. According to the algorithm, we should apply the method of conjugate gradients inorder to minimize function f(x) with constraints(2.13). In accordance with the formulas of the method, the first step is made in the direction of the vector. $\overline{p_1} = - (I - p_{Y'o}) f'(x_o)$

By the above lemma $\overline{p_1} \neq 0$. Then x_o is not the solution of the subsidiary problem minimization problem under consideration.

we now show that $\alpha_1 > 0$

since $< a_i, \overline{p_1} > = 0$ $i \in Y'_0$ and $< a_j, \overline{p_1} > < 0$. we have $< a_i, \overline{p_1} > \leq o$ $i \in {}_{Yo}$ - - - (2.20)

$i \notin y_o \Rightarrow \leq a_i, x_o > - b_i < 0$

There fore $\overline{\alpha}_1 = \min \frac{b_i - < a_i, x_o >}{< a_i , p_i >} > 0$ (since the minimum is taken only over all i for which $<a_i, \underline{p}_1 > > 0$ and consequently over a certain subset of indices i that does not interesect Y_0 according to (2.20) and $b_i - < a_i, x_o > > 0$ with such i.

To show . if $\overline{\alpha_1} > 0$, then $x_o + \alpha p_1$ satisfies condition (2.2) for $0 \leq \alpha \leq \overline{\alpha_1}$

If $i \in Y_0$, then $< a_i, x_o + \alpha p_1 > - b_i = < a_i, x_o > - b_i + \alpha < a_i, \bar{p}_1 >$

$= \alpha < a_i, \bar{p}_1 >$

$= 0$ if $i \in _o$

< 0 if $i = j$

Hence $x_o + \alpha p_1$ satisfies condition (2.2).

If $i \notin Y_0$, then we consider two cases

Case 1. $<a_i, \overline{p_1}> \leq 0$

Then $<a_i, x_o + \alpha p_1 > - b_i = <a_i, x_o > - b_i + \alpha < a_i, p_1 > < 0$

Case 2. $<a_i, \overline{p_1} > > 0$

Then $< a_i, x_o + \alpha p_1 > b_i = <a_i, x_o > - b_i + \alpha < a_i, \bar{p}_1 >$

$\leq < a_i, x_o > - b_i + \alpha_1 < a_i, \overline{p_1} >$

$= <a_i, x_o > - b_i + \min \dfrac{b_i - <a_i, x_o >}{<a_i, p_1 >} < a_i, \overline{P_1} >$

$\leq <a_i, x_o > - \dfrac{b_i + b_i - <a_i, x_o > <a_i, \bar{p}_1 >}{<a_i, p_1 >} = 0$

In both cases $x_o + \alpha p_1$ satisfies condition (2.2)

According to the algorithm, two cases are possible: $\alpha_1 < \alpha_1$ or $\alpha_1 \geq \alpha_1$

Case 1. $\alpha_1 < \overline{\alpha_1}$

Then we obtain a new point $x_1 = x_o + \alpha_1 \bar{p}_1$ that satisfies the relations $< a_i , x_1 > - b_i = o$, $i \in Y'_o$, $< a_i, x_1 > - b_i < 0$ $i \notin Y'_0$ - - - (2. 21)

Case 2. $\alpha_1 \geq \overline{\alpha_1}$

Then we obtain the point $x_1 = x_o + \overline{\alpha_1}\ \bar{p}_1$ and it is taken as a new initial point from which the algorithm begins to operate in checking case(1) or (2).

Let $i \in Y'_0$, then $< a_i, x_1 > - b_i = < a_i, x_o + \overline{\alpha_1}\ \overline{p_1} > - b_i$

$= <a_i, x_o> - b_i + \alpha_1 < a_i, p_1 >$

$= 0$

$\Rightarrow i \in Y(x_1)$ Or $Y'_o \subseteq Y(x_1)$

And more over $<a_i, x_1> - b_i = <a_i, x_o + \bar{\alpha}_1 p_1 > - b_i = <a_i, x_o > - b_i + \overline{\alpha x_1} < a_i, p_1 >$

$= 0$ if $\overline{\alpha_1} = \dfrac{b_i - <a_i, x_o}{<a_i, p_1>}$ $i \notin Y'_0$

Which means $<a_i, x_1> - b_i$ can be zero even for an index not in Y'_0

$\Rightarrow Y(x_1) \supset Y'_o$ the inclusion being strict.

Now we return to case 1. ($\alpha_1 < \overline{\alpha_1}$)

Here the application of the method of conjugate gradients goes on and so long as α_{k+1} holds, all the points, α_{k+1} continue to satisfy (2. 21) like x_1) , since

$A_{Y'o} p_k = A_{Y'o} (I - pY'_o) p_k = 0$ by 2.7 and $(I - p_{Y'_o}) p_k = p_k$)

That means $<a_i, p_k> = 0, i \in_{Y'o}$

$\Rightarrow <a_i, x_k> - b_i = <a_i, x_o> b_i + \alpha_1 < a_i, p_1 > \alpha_2 <a_i, p_2> + - - - + \alpha_k <a_i, p_k>$(since $\alpha_k =$ $x_o + \alpha_1 p_1 + \alpha_2 p_2 + - - - + \alpha_k p_k$), $i \in Y'_o$

$\Rightarrow <a_i, x_k> - bi = 0_i \in_{Y'o}$. More over $<a_i, x_k> - b_i < 0 \; i \notin Y'_o$ holds true.

Suppose not . Then $< a_i, x_k > - b_i \geq 0$

$\Rightarrow < a_i, \alpha_{k+1} - \overline{\alpha_{k+1}} p_{k+1} > b_i \geq 0$

$\Rightarrow < a_i, x_{k+1} > - \alpha_{k+1} < a_i, p_{k+1} > - b_i \geq 0$

$\Rightarrow - \alpha_{k+1} < a_i, p_{k+1} > \geq b_i < a_i, x_{k+1} >$

$$\Rightarrow \alpha_{k+1} \geq \frac{b_i - < a_i, x_{k+1} >}{<-a_i, p_{k+1}>} \quad (< -ai, p_{k+1} >> 0 \text{ since } <ai, p_{k+1} > < 0)$$

$$\Rightarrow \alpha_{k+1} \geq \min \frac{b_i - <a_i, x_{k+1}>}{< - a_i, p_{k+1} >}$$

$\Rightarrow \alpha_{k+1} \geq \overline{\alpha_{k+1}}$ which is a contradiction to our assumption $\alpha_{k+1} < \overline{\alpha_{k+1}}$.

Thus we have shown that in case (1) the iterative process constructs successively points $x_o, x_1, - - -, x_k$, $k \geq 1$ and the value of f(x) strictly decreases along this sequence because it is constructed by the method of conjugate directions. The last point x_k is either the minimum point of f(x) with constraints (2. 13) or $Y(x_k)$ contains strictly the set Y'_o .

In case(2) the direction of motion from point x_o coincides with the vector $p_1 = -(I - pY_o) f'(x_o) \neq 0$, $< a_i, p_1 > = 0$, $i \in Y_o$ and $< a_i, x_o > -b_i < 0$ for $i \notin Y_o$ then

$$\alpha_1 = \min \frac{b_i - < a_i, x_o >}{<a_i, p_1} > 0$$

since the minimum is taken over those i for which $< a_i, \overline{p_1} > > 0$ Then by the method of conjugate gradients it is possible to make at least one non-zero step to the new points at which the value of f(x) is strictly smaller.

In a similar way as shown above we obtain the sequence of points $x_o, x_1, - - - , x_k, k \geq 1$, and x_k is either the minimum point of f(x) with constraints (2. 15) or $Y (x_k) \supset Y_0$

Remark (1) In both cases if x_k is a minimum point of f(x) with constraints (2. 13), then x_k is the minimum point of f(x) on the face of a polyhedral set which is determined by the expressions $< a_i, x > - b_i = 0 , i \in Y (x_k)$.

Since by construction, $Y(x_k) \supset Y'_0$ in case (1) and $Y(x_k) \supset Y_o$ in case(2) and the minimum point on the broader set is the minimum point on the narrower one.

(2) From the foregoing we have that if the method of conjugate gradients does not result in finding the minimum point the set of i indices is extended and for them the next point attained satisfies the relation $<a_i, x_k> - b_i = 0$. But by assumption the vectors $a_i, i\in Y(x_k)$ are linearly independent. Then this extension must be truncated after a finite number steps not exceeding n where n is the dimension of x. Hence starting with a point x_o we shall inevitably come rothe point x_k after a finite number of steps which is iself a minimum point of f(x) with constraints(2. 27) .

<u>Example</u> solve the optimization problem

$$(p) : f(x) = x_1^2 + x_2^2 + 2x_1 + 2x_2 \rightarrow \min, x \in S$$

$$S = \{ x\in R^2 \mid 2x_1 + x_2 \le 2 , x_1 - x_2 = 1 \}$$

by the algorithm described above

Solution : $f(x) = ½ < Ax , x > + < b, x >$ where $A = \begin{pmatrix} 2 & 0 \\ (2 & 2 \end{pmatrix}$ and $b = (2, 2)^T$

Let $x \in S$, put $Y(x) = \{ i : <a_i, x > - b_i = 0, i \in y^- \cup y^o \}$.

$2 > 0$ and $| A| = 4 > 0 \Rightarrow$ A is a positive definite matrix .

Then f(x) is a strictly convex function.

Let $x_o = (0,0)^T$ be the first approximation . Then $Y_0 = Y(x_o) = \{1, 2\}$

$Ay_0 = 2 \;\; 1$ $f'(x_0) = A x_o + b = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$

$$PY_0 = A^T_{Y0} (A_{Y0} A^T_{Y0})^{-1} A_{Y0} = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}^{-1}\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}$$

$$= \frac{1}{9}\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Then $(I - p_{Y_0}) f'(x_o) = 0$

$\Rightarrow x_o = (1,0)$ is a minimum point on the face define by the system of equations $\langle a_i, x\rangle - b_i = 0 \ i \in Y_0$

$$U_0 = -(A_{Y_0} A^T_{Y_0})^{-1} A_{Y_0} f'(x_o) = -\frac{1}{9}\begin{pmatrix}2 & -1\\ -1 & 5\end{pmatrix}\begin{pmatrix}2 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}3\\ 1\end{pmatrix}$$

$$= -\frac{1}{9}\begin{pmatrix}3 & 3\\ 3 & 3\end{pmatrix}\begin{pmatrix}3\\ 1\end{pmatrix} = \frac{1}{9}\begin{pmatrix}12\\ 3\end{pmatrix} = \begin{pmatrix}-4/3\\ -1/3\end{pmatrix}$$

$\Rightarrow U^i_0 < 0$ for $i \in Y_0$ ny-

Now we delete this index from Y_0 and consider an index set $Y'_0 = \{2\}$ and then solve

$(p) : f(x) \longrightarrow \min, x \in S, S = \{x \in R^2 \mid x_1 - x_2 = 1\}$ by the method of conjugate gradients.

According to the method the 1st step is made in the direction $p_1 = -(I - p_{Y'_0}) f'(x_o)$

$A^T_{Y_0} = (1 \ -1)$

$$p_{Y'0} = A^T_{Y'0}(A_{Y'0} A^T_{Y'0})^{-1} A_{Y'0} = \begin{pmatrix}1\\ -1\end{pmatrix}(1 -1)\begin{pmatrix}1\\ -1\end{pmatrix}^{-1}(1 - 1)$$

$$= \begin{pmatrix}1\\ -1\end{pmatrix}(2)^{-1}(1 -1)$$

$$= \tfrac{1}{2}\begin{pmatrix}1\\ -1\end{pmatrix}(1 -1)$$

$$= \tfrac{1}{2}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}$$

$$\Rightarrow -(I - P_{Y_0}) f'(x_o) = -\left(\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \begin{pmatrix}\tfrac{1}{2} & -\tfrac{1}{2}\\ -\tfrac{1}{2} & \tfrac{1}{2}\end{pmatrix}\right)\begin{pmatrix}3\\ 1\end{pmatrix} = \begin{pmatrix}-2\\ 2\end{pmatrix} \neq \begin{pmatrix}0\\ 0\end{pmatrix}$$

$\Rightarrow x_o$ is not a solution of (p).

$$\alpha_1 = -\frac{\langle f'(x_o), p_1\rangle}{\langle p_1, Ap_1\rangle} = \tfrac{1}{2},\quad x_1 = x_o + \alpha_1 p_1 = \begin{pmatrix}1\\ 0\end{pmatrix} + \tfrac{1}{2}\begin{pmatrix}-2\\ -2\end{pmatrix} = \begin{pmatrix}0\\ -1\end{pmatrix}$$

$$f(x_1) = -1, \quad p_2 = -(I - p_{Y_0}) f'(x_1) + \frac{\| (I - p_{Y'0}) f'(x_1) \|^2}{\| (I - p_{Y'0}) f'(x_o) \|^2} p_1$$

But $f'(x_1) = Ax_1 + b = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$

$\Rightarrow -(I - P_{Y0}')\, f'(x_1) = -\frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 3 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \neq 0 \Rightarrow x_1$ is not a solution of (P).

Then $p_2 = \begin{pmatrix} -1 \\ -1 \end{pmatrix} + \begin{pmatrix} -2 \\ -2 \end{pmatrix} = \begin{pmatrix} -3 \\ -3 \end{pmatrix}$, $\quad \alpha_2 = - \dfrac{\langle (3, -1), (-3, -3) \rangle}{\left\langle (-3, -3), \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} -3 \\ 3 \end{pmatrix} \right\rangle} = 1/6$

$x_2 = x_1 + \alpha_2 P_2$

$= \begin{pmatrix} 0 \\ -1 \end{pmatrix} + 1/6 \begin{pmatrix} -3 \\ 3 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ -3/2 \end{pmatrix}$

$f(x_2) = -3/2 \Rightarrow f(x_2) < f(x_1)$

$P_3 = -(I - p_{Y'0})\, f'(x_2) + \dfrac{\| (I - p_{y'0})\, f'(x_2) \|^2}{\| (I - p_{y'0})\, f'(x_1) \|^2}\, p_2$

$f'(x_2) = Ax_2 + b = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} -\frac{1}{2} \\ -3/2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -3 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$-(I - p_{Y'0})\, f'(x_2) = -\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \quad p_3 = 0$

Hence $x_2 = (-\frac{1}{2}, 3/2)$ is a minimum point of $f(x)$ with constraint $x_1 - x_2 = 1, 2x_1 + x_2 \leq 2$ and thus a solution of (p).

CHAPTER 3

METHODS OF FEASIBLE DIRECTIONS

Suppose it is required to solve the following opimization problem

$$(p) : f_o(x) \longrightarrow \min, \; x \in D$$

$$D = \{ x \mid f_i(x) \leq 0, \; i = 1, \text{---}, m, \; Ax - b = 0 \} \text{ --- } (3.1)$$

Where $x \in E^n$ $f_i(x)$ are convex continuously differentiable functions, A is an l x m matrix and b is an l dimensional vector.

Moreover, suppose that the gradients of the functions $f_i(x)$ i = 0, 1, ---, m i) satisfies the lipschitz condition : ie $\|f'_i(x_1) - f'i(x_2)\| \leq C \|x_1 - x_2\|\|$ --- (3.2)

ii) are bounded : ie $\|f_i'(x)\| \leq k \; \forall x \in D$.

In what follows we assume that the set D is compact. The basic idea of the method of feasible direction is to choose a starting point satisfying all the constraints and to move a better point according to the iterative scheme.

$$x_{k+1} = x_k + \alpha_k p_k$$

where x_k is the initial point for the k^{th} iteration, p_k is the direction of movement, α_k is the step length and x_{k+1} is the final point obtained at the end of each iteration. The new point x_{k+1} is taken as an initial point and the whole procedure is repeated.

Def: Let $x_o \in D$. A direction $p \in E^n$ with the property that with small α, $x_o + \alpha p \in D$ and $f_o(x_o + \alpha p) < f_o(x_o)$ is called a feasible direction.

Now the problem consists of working an effective method of finding a feasible direction and choosing a step length α so as to provide for convergence to a minimum point.

Below we assume that always the condition of non- degeneracy is fulfilled. That is there is a point x such that $Ax - b = 0$, $f_i(x) < 0$, $i = 1,2, \text{---}, m$.

3.1. Method of Choosing a Feasible Direction

Consider the optimization problem (p) and assume that we have located a feasible point x_o where the constraints are binding or nearly so . We wish to take a step in that direction which will maximize, decrease β, in our objective function $f_o(x)$, and yet will not violate any of the constraints. If we assume that over sufficiently small interval we can linearize our constraints and our objective function, then the problem can be approximated as a linear programming problem.

Now for each $X \in D$, define $y_\delta^-(x) := \{ i \mid f_i(x) \geq -\delta,\ i = 1, ---, m \}$. Let $\xi_i > 0$ be arbitrary numbers.

Then we consider the following optimization problems

$$(p) : \min \eta$$

subject to $< f'i(x), p > \leq \xi_i \eta_i \quad i \in y_\delta^-(x) \cup \{0\}$

$$\|p^i\| \leq 1 \qquad (3.3)$$

$$Ap = 0$$

Where $\eta = -\beta$

The restriction $< f'_o(x), p > \leq \xi_o \eta$ requires that β be not greater than the change in $f_o(x)$. Hence by minimizing η, we maximize the decrease in $f_o(x)$. In the second case $< f_i(x), p > \leq \xi_i\ \eta$, $i \in y_{\delta(x)}$, we add the term $- \xi_i\eta$, to force movement away from the constraint boundaries and into the feasible region. In the second case $< f_i'(x), p> \leq \xi_i\eta$, $i \in y^-\alpha$ and ξ_i is given the value of zero if f_i's are linear and the value 1 other wise. After the determination of p_1 we normalize this vector so that it has a unit length.

Theorem 3. 1 . Let $p(_\delta(x)$ and $\eta_\delta(x)$ be a solution of (p). Then $P_\delta(x)$ is a feasible direction if $\eta_\delta(x) < 0$.

Proof: Let $x \in D$ and $\alpha > 0$. To show $p_\delta(x)$ is a feasible direction requires

i) $f_o(x + \alpha p_\delta(x)) < f_o(x)$

ii) $f_i(x + \alpha p_\delta(x)) \leq 0,\ i = 1, ---, m$

ii) $A(x + \alpha p_\delta(x)) - b = 0$

We consider three cases

case 1. $i = 0$. By Taylor's formula, we have

$f_o(x + \alpha p_\delta(x)) = f_o(x) + \alpha < f'_o(x), p_\delta(x) >$ (where $\theta_o = x + \xi_o\ \alpha p_\delta(x), 0 \leq \xi_0 \leq 1$)

$= f_o(x) + \alpha < f'_o)x, p_\delta\ (x) > + \alpha < (f'(\theta_o) - f'_o(x)), p_\delta(x) >$

$\leq f_o(x) + \alpha < (f'_o(x), p_\delta(x) > + \alpha\|f'_o(\theta_o) - f'_o(x)\|\ \|p_\delta(x)\|$ (cauchy schwartz inequality)

$\leq f_o(x) + \alpha < f'_o(x), p_\delta(x) > + \alpha C \|\theta_0 - x\| \|p_\delta(x)\|$

(Lipschitz' condition. Since $\theta o = x + \xi_o \alpha p_\delta(x)$ & $0 \leq \xi_o \leq 1$

We have that $\|\theta_o - x\| = \|\xi_o \alpha p_\delta(x)\| = \xi_o \alpha \|p_\delta(x)\| \leq \alpha\|p_\delta(x)\|$

Then $(*) \leq f_o(x) + x < (f'_o(x), p_\delta(x) > + \alpha^2 C \|p_\delta(x)\|^2$

$\leq f_o(x) + x \xi_o \eta_\delta(x) + \alpha^2 C \|p_\delta(x)\|^2$ (constraint condition)

$$= f_o(x) + x \xi_o \eta\delta(x) \left[1 + \frac{\alpha C\|p_\delta(x)\|^2}{\xi_o \eta_\delta(x)} \right] \text{ - - (3. 4)}$$

Case 2. $i \in y_\delta(x)$. In a similar way as in case 1, we get $f_i(x + \alpha p_\delta(x)) \leq f_i(x) + \alpha \xi i \eta_\delta(x) + \alpha^2 C \|p_\delta(x)\|^2$ - - - (3. 5)

Case 3. $i \notin y_\delta(x)$. A gain using Taylor's formula, we have

$f_i(x + \alpha p_\delta(x)) = f_i(x) + \alpha < f'_i (\theta i), p\delta(x))$

$\leq f_i(x) + \alpha \|f_i(\theta_i\| \|p\delta(x)\|$ (cauchy schwart'z inequality)

But since the gradients are bounded, we have that $\|f_i(\theta i)\| \leq k$ and thus $f_i(x + \alpha p_\delta(x))$

$\leq f_i(x) + \alpha k \|p_\delta(x)\| \leq -\delta + \alpha k\|p_\delta(x)\|$ - - - (3. 6) (Since $f_i(x) < -\delta$, $i \notin y'_0$

We now choose $\alpha \geq 0$ satisfying the inequalities

$f_o(x + \alpha p_\delta(x)) \leq f_o(x) + ½ \alpha \xi_o \eta_\delta(x)$ (Since from this inequality follows $f_o(x+\alpha p_\delta(x) \leq 0$

$i \in y_\delta(x)$ as $½ \alpha \xi_o \eta_\delta(x) < 0$)

$f_i(x + \alpha p_\delta(x) \leq 0 \; i \in y_\delta(x)$

$f_i(x + \alpha p_\delta(x)) \leq 0 \; i \notin y_\delta\text{-}(x)$ - - - (3. 7)

These inequalities are satisfied provided that α satisfies the inequalities:

$$1 + \frac{\alpha C \|p_\delta(x)\|^2}{\xi o \eta_\delta (x)} \leq ½ , \; \xi_i \eta_\delta(x) + x C \|p_\delta(x)\|^2 \leq 0 \; i \in y_\delta^-(x) \text{ and}$$

$-\delta + \alpha k\|p_\delta(x)\| \leq 0 \; i \notin y_\delta^-(x)$ - - -(3. 8)

$$\text{or } x \leq -½ \frac{\xi o \eta_\delta(x)}{C \|p_\delta(x)\|^2}, \; \alpha \leq \frac{-\xi i \eta_\delta(x)}{C\|p_\delta(x)\|^2} \quad i \in y_\delta^-(x) \text{ and } \alpha \leq \frac{\delta}{K\|p_\delta (x)\|} \quad i \notin y_\delta^-(x) \text{ - - - (3. 9)}$$

Since $\eta_\delta(x) < 0$,$\xi_i > 0$,$C> 0$ and $k > 0$this inequalites will not violate the choice that $\alpha > 0$

More over since $p_\delta(x)$ is a solution of (p) we have that $Ap_\delta(x) = 0$

There fore $A(x + \alpha p_\delta(x)) - b = Ax - b + \alpha Ap_\delta(x) = 0$

Thus choosing α in the above way, we have that $p_\delta(x)$ is a feasible direction. ||

Lemma 3.1.Let $x \in D$ be not a solution of (p).Then $\eta_\delta(x) < 0$ for any sufficiently small δ.

Proof: From the condition of non-degeneracy, we have that there is a point $\overline{x}$ such that

$A\overline{x} - b = 0$, $f_i(\overline{x}) \leq \sigma, i = 1, - - -, m, \sigma < 0$ - - (3.10)

Letx_* be as solution of (p).

Put $y^-_0(x) := \{ i \mid f_i(x) = 0\ i = 1 - - - ,m\}$ and $-\delta^o(x) := \max\limits_{i \notin y_0(x)} f_i(x)$

We now show that, if $\delta < \delta^o$, then $y_\delta(x) = y^-_0(x)$

Let $i \notin y_0(x)$. Then $f_i(x) \leq -\delta^o < -\delta$ This implies that $i \notin y_\delta(x)$

Hence $y_\delta(x) \subseteq y^-_0(x)$ - - - (1)

To show the other inclusion, let $i \in y^-_o(x)$. Then $f_i(x) = 0$

which implies that $f_i(x) \geq -\delta\ (\delta > 0)$. That means $i \in y^-_\delta(x)$

There fore $y^-_0(x) \subseteq y_\delta(x)$ - - - (2)

From (1) and (2) we obtain $y_\delta^-(x) = y^-_0(x)$. //

Now suppose $\delta < \delta^o$ so that $y^-_\delta(x) = y_0(x)$

Set $x_\rho := \rho\overline{x} + (1-\rho) x_*, 0 < \rho \leq 1$

Since $f_i(x), i = 0,1,2, - - -, m$ are convex and $f_i(x_*) \leq 0 i = 1, - - -, m$, we have

$$
\begin{aligned}
f_i(x_\rho) &= f_i(\rho\overline{x} + (1-\rho)x_*) \\
&\leq \rho f_i(\overline{x}) + (1-\rho) f_i(x_*) \\
&\leq \rho\sigma
\end{aligned}
$$

Further, for $i \in y_\delta(x)$, $f_i(x) = 0$ (Since $y^-_\delta(x) = y^-_0(x)$ and $f_i(x) = 0$ for $i \in y^-_0(x)$) and there fore for $0 < \lambda < 1$ we have

$$
\begin{aligned}
\lambda\rho\sigma \geq \lambda f_i(x\rho) &= \lambda f_i(x_\rho) + (1-\lambda)f_i(x) \\
&\geq f_i(\lambda x_\rho + (1-\lambda)x \text{ (convexity of } f_i \text{)} \\
&= f_i(x + \lambda(x_\rho - x)) - f_i(x) \\
&\geq \langle f_i(x), \lambda(x_\rho - x) \rangle \text{ (Sub gradient inequality)} \\
&= \lambda \langle f_i'(x), x_\rho - x \rangle
\end{aligned}
$$

Thus $\rho\sigma \geq \langle f_i'(x), x_\rho - x \rangle, i \in y^-_\delta(x)$ - - - (3.11)

More over since x is not a minimum point of $f_o(x)$ in D we have $0 > \gamma = f_o(x_*) - f(x) \geq$

$\langle f'_0(x), x_* - x \rangle$ (sub gradient inequality)

Hence $\langle f'_0(x), x_\rho - x \rangle = \langle f'_o(x), \rho\overline{x} + (1-\rho)x_* - x \rangle$

$$= \langle f'_o(x), \rho\overline{x} - \rho x + \rho x + (1-\rho)x_* - x \rangle$$

$= <f'_o(x), \rho (\bar{x}-x) + (1 - \rho) (x_*-x) >$

$= < f'_o(x) \rho(\bar{x} - x) > + < f'_o(x), (1- \rho) (x_*-x) >$

$= \rho<f'_o(x), \bar{x}-x > + (1-\rho) < f'_o(x), x_* - x >$

$\leq \rho < f'_o(x),\bar{x}-x > + (1-\rho) \gamma$ - - - (3. 12)

(Since $x \in D$ is not a solution of (p) by the converse of the characterizing theorm of convex optimization we have that $< f'_o(x), \bar{x}- x> \leq 0$. Moreover $(1-\rho) \rho\gamma < 0$.

From (2.11) and (2.12) with sufficiently small $\rho > 0$, the following inequalities holds:

$<f'_o(x) , \bar{P}_\rho > < 0, < f'i(x), \bar{p}_\rho > < 0 \quad i \in y_\delta(x)$ - - - (3.13) where $\bar{p}_\rho = x_\rho - x$

Take $p_\rho : - p_\rho$ if $\| p_\rho\| \leq 1$ and $p_\rho : = \dfrac{p^-_\rho}{\|p_\delta\|}$ if $\| \bar{p_\rho} \| \geq 1$

and so $\| p_\delta\| \leq 1$. Now if we put $\eta_\delta : = \max\limits_{i\in y_\delta(x) U\{\theta\}} < \dfrac{f_i'(x), p_\delta}{\xi i}>$

then by (3. 13), $\eta_\rho < 0$ and the inequality $< f'_i(x), P_\rho > \leq \xi_i \eta_\delta \; i \in y^-_\delta(x) U \{ 0 \}$ holds true - - - (3. 14).

More over $Ax_\rho - b = A(\rho\bar{x} + (1 - \rho) x_*) - b$

$= \rho A\bar{x} + (1 - \rho) Ax_* - b$

$= \rho (A\bar{x} - b) + (1 - \rho) (Ax_* - b)$ (Since $Ax - b = 0$ & $Ax_* -b = 0$)

$= 0$

From the definition of p_ρ, we have $p_\rho = \bar{\alpha} p_\rho = \alpha(x_\delta-x)$ where $0 < \alpha \leq 1$.

Then $Ap_\delta = Ap_\rho + (Ax-b) = \alpha [Ax_\delta - b] + (1 - \alpha) [Ax - b] = 0$ - - - (3.15)

Hence p_ρ and η_ρ satisfies condition (3. 3).

Since $\eta_\rho < 0$, so much the more $\eta_\delta(x)< 0$, for $\eta_\delta(x) \leq \eta_\rho$ by definition. ||

3.2. <u>Algorith of Method of Feasible Direction</u>

Let $x_0 \in D$ be an arbitrary first approximation and $\delta_o > 0$.

The general step of the algorithm is described as follows:

Let the point $x_k \in D$ be obtained at the k^{th} step and $\delta_k > 0$. Point x_k and the number $\delta_k > 0$ have been computed .

1) Solve the problem of linear programming

$\min \eta$

$< f'_i(x_k), p > \leq \xi_i\eta, \; i \in y^-_\delta(x_k) U \{ 0 \}$

$= \langle f'_o(x), \rho(\bar{x}-x) + (1-\rho)(x_*-x) \rangle$

$= \langle f'_o(x)\, \rho(\bar{x}-x) \rangle + \langle f'_o(x), (1-\rho)(x_*-x) \rangle$

$= \rho \langle f'_o(x), \bar{x}-x \rangle + (1-\rho) \langle f'_o(x), x_* - x \rangle$

$\leq \rho \langle f'_o(x), \bar{x}-x \rangle + (1-\rho)\gamma$ - - - (3. 12)

(Since $x \in D$ is not a solution of (p) by the converse of the characterizing theorm of convex optimization we have that $\langle f'_o(x), \bar{x}- x \rangle \leq 0$. Moreover $(1-\rho)\rho\gamma < 0$.

From (2.11) and (2.12) with sufficiently small $\rho > 0$, the following inequalities holds:

$\langle f'_o(x), \overline{P_\rho} \rangle < 0, \langle f'i(x), \overline{p_\rho} \rangle < 0 \quad i \in y_\delta(x)$ - - - (3.13) where $\overline{p_\rho} = x_\rho - x$

Take $p_\rho := p_\rho$ if $\| p_\rho \| \leq 1$ and $p_\rho := \dfrac{p^-_\rho}{\|p_\delta\|}$ if $\| \overline{p_\rho} \| \geq 1$

and so $\| p_\delta \| \leq 1$. Now if we put $\eta_\delta := \max\limits_{i \in Y_\delta(x) \cup \{0\}} \dfrac{\langle f_i'(x), p_\delta \rangle}{\xi i}$

then by (3. 13), $\eta_\rho < 0$ and the inequality $\langle f'_i(x), P_\rho \rangle \leq \xi_i \eta_\delta \; i \in y^-_\delta(x) \cup \{0\}$ holds true - -- (3. 14).

More over $Ax_\rho - b = A(\rho\bar{x} + (1-\rho)x_*) - b$

$= \rho A\bar{x} + (1-\rho)Ax_* - b$

$= \rho(A\bar{x} - b) + (1-\rho)(Ax_* - b)$ (Since $Ax - b = 0$ & $Ax_* - b = 0$)

$= 0$

From the definition of p_ρ, we have $p_\rho = \alpha\bar{p}_\rho = \alpha(x_\delta - x)$ where $0 < \alpha \leq 1$.

Then $Ap_\delta = Ap_\rho + (Ax-b) = \alpha[Ax_\delta - b] + (1-\alpha)[Ax - b] = 0$ - - - (3.15)

Hence p_ρ and η_ρ satisfies condition (3. 3).

Since $\eta_\rho < 0$, so much the more $\eta_\delta(x) < 0$, for $\eta_\delta(x) \leq \eta_\rho$ by definition. ||

3.2. <u>Algorith of Method of Feasible Direction</u>

Let $x_0 \in D$ be an arbitrary first approximation and $\delta_o > 0$.

The general step of the algorithm is described as follows:

Let the point $x_k \in D$ be obtained at the k^{th} step and $\delta_k > 0$. Point x_k and the number $\delta_k > 0$ have been computed .

1) Solve the problem of linear programming

$$\min \eta$$

$$\langle f'_i(x_k), p \rangle \leq \xi_i\eta, \; i \in y^-_\delta(x_k) \cup \{0\}$$

$Ap = 0,\ -1 \le P^i \le 1, j = 1, - - - , m$

by one of the standard methods

Let the solution be η_k, p_k.

2) If $\eta_k < \delta_k$, then $x_{k+1} = x_k + \alpha_k p_k$, $\delta_{k+1} = \delta_k$ where $\alpha_k = \frac{1}{2^{q_o}}$, and q_o is the first integer of q = 0, 1 - - - for which of the following inequality hods:

$$f_o(x_k + \frac{1}{2^q} p_k) \le f_o(x_k) + \frac{1}{2}\frac{1}{2^q}\eta_k$$

$$f_i(x_k + \frac{1}{2^q} p_k) \le 0,\ i = 1, - - - , m \quad - - - (3.16)$$

3) If $\eta_k \ge -\delta_k$, then $x_{k+1} = x_k$, $\delta_{k+1} = \frac{1}{2}\delta_k$

4) Return to (1)

Condition of halt (stop) of the algorithm

If at a certain step, $\delta_k < \delta^o(x_k)$, where $\delta^o(x_k) = -\max_{i \notin y_o^{\cdot}(x_k)} f_i(x_k)$ and $\eta_k = 0$, then x_k is a solution of(p). That is x_k is the minimum point of $f_o(x)$ on D.

3.3. Subtantiation of the convergence of the method

Theorem 3.2. Let (x_k) be a sequence of points constructed. Then (x_k) is finite and its last element, say x_k is a solution of (p) or else(x_k) is infinite and every accumulation point x* of(x_k) is minimum point of $f_o(x)$.

Proof: Let (x_k) be a sequence constructed by the method of feasible directions.

Case 1 . The sequence (x_k) is truncated at a certain step k because the conditions of halt of the algorithm has been fulfilled.

That means $\eta_k = \eta_\delta(x_k) = 0$ and $\delta_k < \delta^o(x_k) = -\max_{i \notin y_o(x_k)} f_i(x_k)$ - - - (3. 17)

In the proof of lemma 3.1. we have that if condition (3.17) has been fulfilled, then $\eta_\delta(x_k) < 0$ provided that x_k is not the solution of (p) . But by assumption we have that $\eta_k = 0$ and hence x_k is a solution of (p).

Case2. Let the iterative process be now continued with out limit so that we have an infinite sequence(x_k) k = 0, 1, 2 - - - .

More over let x_k be a point for which $\eta_k < -\delta_k$. Since $\|p_k\| \le 1$, then by (3.9) we have if $\alpha \le \frac{1}{2}\frac{\xi o \eta k}{C}$

$x \leq -\frac{\xi_i \eta k}{C}$ $i \in y_k$ and $\alpha \leq \frac{\delta_k}{k}$ $i \notin y^{-}_k$ - - - (3. 18), then $f_o(x_k + \alpha p) \leq f_o(x_k) + ½ \alpha\xi o\eta_k$, $f_i (x_k + \alpha p_k) \leq 0$ $i = 1, - - -, m$ holds.

According to the algorithm, α_k coincides with the first of the quatities $\frac{1}{2^i}$, $i = 1, - - -, m$ satisfying (3.16). Since such an α exists , these inequalities will be satisfied after a finite number of trials.

Let i_o be the 1[st] indices satisfying the inequalities (3. 16) and so $\alpha = \frac{1}{2^{io}}$

Then $\frac{\alpha - 1}{2^{io-1}}$ did not satisfy (3.16) and hence (3.18) .

That means $\frac{1}{2^{io-1}} > ½ \min \{ -½ \frac{\xi o\eta_k}{C}, \frac{\delta_k}{k}, \min_{i\in y^{-}_k} -\frac{\xi_i\eta_k}{C} \}$

or $\alpha_k = \frac{1}{2^{io}} > ½ \min \{ - ½ \frac{\xi_o\eta_k}{C}, \frac{\delta_k}{k}, \min_{i\in y^{-}_k} \frac{-\xi_i\eta_k}{C}\}$ - - - (3.19)

Now from $- \eta_k > \delta_k$ it follows that $\alpha k > ½ \min \{\frac{\xi_o\delta_k}{2}, \frac{\delta_k}{k}, \min \frac{\xi_i \delta_k}{C}\}$

Hence $\alpha_k > \frac{\delta_k}{2} \in_o$ where $\in o = \min \{ \frac{\xi o}{2C}, \frac{\xi i}{C}, - - - \frac{\xi m}{C}, /k\}$ (3.20)

By (3.6), (3.20) and $\eta_k < - \delta_k < 0$, we have that $f_o(x_{k+1}) \leq f_o(x_k) + ½\alpha_k\xi_o\eta_k$

$\leq f_o(x_k) - ½\alpha_k\xi_o\delta_k$

or $f(x_{k+1}) \leq f_o(x_k) - \frac{\xi_o \in_o\delta_k^2}{4}$ - - (3.21)

We now show that from (3.21) it follows that $\delta_k \rightarrow 0$ as $k \rightarrow \infty$.

Assume the contrary. That is $\delta_k \rightarrow 0$. As the sequence (δ_k), $k = 0, 1, - - -, m$ decreases monotonically and if $\delta_{k+1} < \delta_k$, then $\delta_{k+1} = ½\delta_k$, and the fact that $\delta_k \rightarrow 0$ can mean only $\delta_k = \delta > 0$ for sufficiently large k. But if δ_k remains constant, then the condition $\eta_k < - \delta_k$ is fulfilled and thus the inequality (3.21) holds.

Thus for sufficiently large k $(k \geq ko)$, $\delta_k = \delta$ and the inequality .

$f_o(x_{k+1}) \leq f_o(xk_o) - \frac{\xi_o \in_o \delta^2}{4}$ is fulfilled.

There fore $f_o(x_N) \leq f_o((x_{ko})-(N-k) - \frac{\xi_o \in_o \delta^2}{4}$

This is because $f_o(x_N) \leq f_o(x_{N-1}) - \frac{\xi_o \in_o \delta^2}{4}$

$$\leq f_o(x_{N-2}) - 2\,\frac{\xi_o \in_o \delta^2}{4}$$

$$\leq f_o(x_{ko}) - (N-k_o)\,\frac{\xi o \in o\, \delta^2}{4}$$

Then $f_o(x_N) \xrightarrow{N\to\infty} -\infty$ (since $f_o(x_{ko})$ is a constant, $\frac{\xi_o \in_o \delta^2}{4} > 0$ because $\xi_o \geq 0$, $\in_o$ is positive since both C & k are positive and N - ko $\xrightarrow{N\to\infty} \infty$

There fore $f_o(x_n) \xrightarrow{N\to\infty} \infty$. But this is a contradiction to the fact that a continuous function $f_o(x)$ in a compact region D is bounded. Hence $\delta_k \xrightarrow{k\to\infty} 0$, which means that the initial f_o is successively halved an infinite number of times. Or case (3) takes place an infinite number of times: $\eta_k \geq -\delta_k$

Let $y := \{ k : \eta_k \geq -\delta_k \}$ Then $\eta_k \xrightarrow{k\to\infty} 0$, $k \in y$ (Because for $k \in y$, we have $-\delta_k \leq \eta_k \leq 0$ and $\delta_k \to 0$)

Consider the sequence of points(x_k), $x_k \in y$. Since D is a compact we can assume w.l.o.g that $x_k \to x^*$. We now show that x^* is a minimum point of $f_o(x)$ in D.

Assume the contrary. That is let x^* be not the minimum point of $f_o(x)$ in D. Then by Lemma 3.1 with all $\delta < \delta^o(x^*)$ where $\delta^o(x^*) = -\overline{\max}\, f_i(x^*)$, we have that $y_\delta(x^*) = y_o^-(x^*)$ $i \in y^-_o(x^*)$ and $\eta_\delta(x^*) < 0$.

More over, from $y^-_\delta(x^*) = y^-_o(x^*)$ it follows that $\eta_\delta(x^*) = \eta_o(x^*)$. Further, $y_\delta(x_k) \subseteq y^-_0(x^*)$ with sufficiently large $k \in y$. To show this suppose that $i \notin y^-_0(x^*)$. We show that $i \notin y^-_{\delta k(xk)}$.

Now $i \notin y\text{-}_o(x^*)$ implies $fi(x^*) < 0$ (By definition of $y^-_{o(x^*)}$ and $f_i(x^*) \leq 0$).

Since $\delta_k \xrightarrow{k\to\infty} 0$ we have $f_i(x^*) < -\delta_k$ and since $x_k \xrightarrow{k\,-\to\infty} x^*$ we have by continuity of f_i that $f_i(x_k) < -\delta_k$. This implies that $i \notin y^-_{\delta k}(x_k)$. there fore $y^-_{\delta k}(x_k) \subseteq y^-_o(x^*)$.

Since by assumption x^* is not the minimum point of $f_o(x)$ on D there is a vector $p(x^*)$ such that $Ap(x^*) = 0$, $\| px^* \| \leq 1$, $< f'_i(x^*) P(x^*) > \leq \xi_i \eta_o(x^*)$ $i \in y^-_o(x^*) \cup \{0\}$ This is because by lemma 3.1. $\eta_o(x) < 0$ and hence by theorem1 $p(x^*)$ is feasible direction) .

Then by the continuity of f'_i and the scalar product we have from $x_k \to x^*$ follows $f_i(x_k) \to f_i(x^*)$ and $< f_{i'}(x_k), p(x^*) > \to < f'_i(x^*), p(x^*) >$.

That means $< f'_i(x_k), p(x^*) > \leq \xi_i \eta_o(x^*)$, $i \in y^-_k \cup \{ 0 \}$

$$Ap(x^*) = 0 \ , \| p(x^*) \| \leq 1$$

How ever, the last relations means that $\eta_k \equiv \eta_{\delta k}(x_k) \le \eta_o(x^*) < 0$ with sufficiently great k and this contradicts the fact that $\eta_k \xrightarrow{k \to \infty} 0$, $k \in y$.

Hence x* is a minimum point of fo(x) in D. ||

<u>Theorem 3.3.</u> Let (x_k) be a sequence of points constructed by the method of feasible directions. Then $f_o(x_k)$ with out increasing monotonically converges to $f_o(x^*)$ where x* is the minimum point of $f_o(x)$ in a region D.

<u>Proof</u>: Let (x_k) be sequence constructed by the method of feasible directions. Then $f(x_{k+1}) \le f(x_k)$. That means this sequence of numbers does not increase monotonically Since it has a lower bound (A continuous function on a compact set has a lower bound) it converges to a certain limit f_o. However it was shown above that there is a subsequence(x_{kj}), $j \in y$ such that $x_{kj} \to x^*$.

There fore by the continuity of f_o we have $f_o(x_{kj}) \to f_o(x^*)$

As the whole sequence and its subsequence converges to the same limit it follows that $f_o(x_k) \to f_o(x^*)$. ||

CHAPTER 4

MTHOD OF CONDITIONAL GRADIENT

This method is used for solving the problem of minimization of non linear function in a region in which the problem of minimization of a linear function can be solved with out great difficulties. That means in this method the function being minimized is approximated at each iteration by a linear form.

Now consider solving the following optimization problem: (p) : $f(x) \longrightarrow \min, x \in \Omega$

Where $f'(x)$, $x \in E^n$ is continuously differentiable function, Ω is a compact convex region and the gradient $f'(x)$ of the function $f'(x)$ in Ω satisfies the lipschitz' condition: ie

$\|f'(x_1) - f'(x_2)\| \le L\|x_1 - x_2\|$ for all $x_1, x_2 \in \Omega$ - - - (4.1)

The method of conditional gradient is described as follows

Let x_k, the approximation at the kth step of the iterative process be already constructed.

Calculate $f'(x_k)$ and solve the following optimization problem

$(p) : < f'(x_k), z > \longrightarrow \min, z \in \Omega$

Let $z(x_k)$ be solution of (p). Now take $p_k := z(x_k) - x_k$ and $x_{k+1} = x_k + \alpha_k p_k$ where $\alpha k > 0$ is the step length in the direction p_k. Then x_{k+1} is taken as an initial one and the process is repeated.

We now show that with definite rule of choosing α_k the process converges and the bounds on the rate of convergence will be estabilished.

4.1. Rule of choosing the step length

Let $x \in \Omega$ be arbitrary and let $z(x)$ be the minimum point of $< f'(x), z >$ in Ω.

That is $< f'(x), z(x) > \le < f'(x), z >, \forall z \in \Omega$ - - - (4.2)

Take $p(x) := z(x) - x$ and $\eta(x) := \min < f'(x), z - x >, z \in \Omega$

Property 1. i) $\eta(x) = < f'(x), p(x) >$

ii) $\eta(x) \le 0$

Proof: i) $\eta(x) = \min_{z \in \Omega} < f'(x), z - x >$ (By definition).

$= \min_{z \in \Omega} [< f'(x), z > - < f'(x), x >]$

$= \min_{z\in\Omega} < f'(x),z> - < f'(x), x> = < f'(x), z(x) > - < f'(x), x >$ (Since z(x) is a solution of (p)

$= < f'(x), z(x)-x > = < f'(x), p(x) >$

ii) By(4. 2) we have that $< f'(x), z(x) > \le < f'(x), z > \forall z\in \Omega$.

That means $< f'(x), z(x) - x > \le 0 \ \forall \in \Omega$. Now since $x\in \Omega$. we have that $< f'(x), z(x) - x> \le 0$.

There fore $\eta(x) \le 0$.

Using Taylor's formula and (4.1) we have

$f(x + \alpha p_\delta(x) = f(x) + \alpha < f'(\theta), p(x) >$ where $\theta = x + \xi\alpha p(x)$, $0 \le \xi \le 1$

$= f(x) + \alpha < f'(x) , p(x) > + < f'(\theta)-f'(x), p(x) >$

$\le f(x) + \alpha\eta(x) + \alpha\|f'(\theta) - f'(x)\| \, \|p(x)\|$

$\le f(x) + \alpha\eta(x) \alpha L \|\theta-x\| \, \|p(x)\|$

$\le f(x) + \alpha\eta(x) \alpha L\|p(x)\|^2$

Thus $f(x+\alpha p(x)) \le f(x) + \alpha [\eta(x) + \alpha L \|p(x)\|^2]$ - - - (4.3)

If we choose $\alpha \le -\frac{1}{2}\frac{\eta(x)}{L\|p(x)\|^2}$ - - - (4. 4), then $\eta(x) + \alpha L \|p(x)\|^2 \le \frac{1}{2}\eta(x)$ and hence

$f(x + \alpha p(x) \le f(x) + \alpha\frac{\eta(x)}{2}$ - - - (4. 5).

4.2. Discription of the Algorithm

Let $x_o \in \Omega$ be an arbitrary point. Then the general step of the algorithm is described as follows:

Let the point x_k be already constructed , $k \ge 0$.

1. Solve the optimization problem $< f'(x_k), z > \longrightarrow \min,\ z \in\Omega$ and then calculate $z(x_k)$, $p(x_k)$ and $\eta(x_k)$.
2. Construct $x_{k+1} = x_k + \alpha_k p(x_k)$ where $\alpha_k = 2^{-io}$ and io is the first indices satisfying the inequality $f(x_k + 2^{-i} p(x_k)) \le f(x_k) \frac{2^{-i}\eta(x_k)}{2}$ - - - (4. 6)
3. The condition of halt : The process stops if $\eta(x_k) = 0$.

4.3. Substantiation of convergence of the method and Estimation of the rate of convergence

From the rule of choosing the step length (4.5)we have $f(x_{k+1}) \le f(x_k)+\alpha_k\frac{\eta(x_k)}{2}$ - - - (4. 7)

In order to substantiate the convergence of the algorithm it is necessary to show that inequalities (4. 6) and (4.7) are always satisfied.

By (4.4) and (4. 5) inequality (4.6) will be satisfied if the inequality $2^{-i} \leq -\frac{1}{2}\frac{\eta(x_k)}{L\|p(x_k)\|^2}$,

is satisfied and since i_o is the first index satisfying (4. 6), we have

$$2\alpha_k = 2^{-io-1} > -\frac{\frac{1}{2}\eta(x_k)}{L\|p(x_k)\|^2}, \quad \text{which means } \alpha_k > 1 - \frac{1}{4L}\frac{\eta(x_k)}{\|p(x_k)\|^2} \quad \text{--- (4. 8)}$$

Hence if $\eta(x_k) < 0$, then inequality (4.6) will be satisfied after a finite number of trials and then x_k choosen will satisfy inequality(4.8) .

Lemma 4.1. If (x_k) , k = 0,1 , - - - is a sequence of points obtained in implementing the algorithm of the method of conditional gradient, then

a) $x_k \in \Omega$

b) $f(x_k)$ decreases monotonically

c) $\eta(x_k) \longrightarrow o$ as $k \longrightarrow \infty$.

Proof: Let $x_k \in \Omega$ for $k \leq m$ .we now show that $x_{k+1} \in \Omega$. That is we prove it by induction on k .

Since $0 \leq \alpha_k \leq 1$ ($\alpha_k = 2^{-k}$) and $z(x_k) \in \Omega$, we have by convexity of Ω, that

$$x_{m+1} = x_m + \alpha_m p(x_m) = x_m + x_m(z(x_m) - x_m)$$

$$= (1-\alpha_m)\, x_m + x_m\, z(x_m) \in \Omega$$

Hence $x_k \in \Omega\ \forall k = 0, 1,$ - - - (a) is proved .

Since $p(x_k) = z(x_k) - x_k$, $z(x_k) \in \Omega$, $x_k \in \Omega$ and Ω is a compact set, $\|p(x_k)\|$ has a limit say C in Ω .

$$\text{Now } f(x_{k+1}) - f(x_k) \leq \alpha_k \eta(x_k) \leq \frac{1}{2} - \frac{1}{4L}\frac{\eta(x_k)}{\|p(x_k)\|^2} \frac{\eta(x_k)}{2} \quad \text{(by 4. 8)}$$

$$= -\frac{1}{8LC^2}\eta^2(x_k)$$

Hence $f(x_{k+1}) - f(x_k) \leq \frac{1}{8LC^2}\eta^2(x_k)$ --- (4. 9).

Adding (4.9) $\forall k = 0,1,$ - - -, m - 1, we have $f(x_m) - f(xo) \leq -\frac{1}{8LC^2}\sum_{k=0}^{m-1}\eta^2(x_k)$

Since Ω is compact and f(x) is continuous we have that f attains its minimum and then $f(x_m) \geq f^*$ where f^* is the minimum value of f(x) in Ω.

There fore $\sum_{k=0}^{m-1} \eta^2(x_k) \leq 8LC^2 [f(x_o) - f(x_m)] \leq 8LC^2 [f(x_o) - f^*]$

Hence $\sum_{k=0}^{\infty} \eta^2(x_k)$ converges or $\sum_{k=0}^{\infty} \eta(x_k)$ converges

Thus $\eta(x_k) \longrightarrow 0$ as $k \longrightarrow \infty$. We proved(C).

From this it follows that $f(x_{k+1}) - f(x_k) \leq 0$ (By 4. 9)

That means $f(x_{k+1}) \leq f(x_k)$ and hence we proved (b) .||

Now there are two possibilities

Case 1. The algorithm stops after a finite number steps and the condition $\eta(x_k) = 0$ is fulfilled

In this case $\eta(x_k) = 0$ implies that $< f'(x_k), p(x_k) > = 0$ or $< f'(x_k), z(x_k) - x_k > = 0$

or $< f'(x_k) z(x_k) > - < f'(x_k), x_k > = 0$

That means $< f'(x_k), z(x_k) > = < f'(x_k), z(x_k) > \leq < f'(x_k), z > z \in \Omega$ which is a necessary condition for f(x) to attain its minimum point x_k.

Case 2. Montonically decreasing sequence of values of the function f(x) (ie $(f(x_k))$) is obtained. This case is subject to the following lemma.

Lemma 4.2. At any limit point of the sequence(x_k) k = 0,1 - - - the necessary condition for a minimum of f(x) in the set Ω is fulfilled.

Proof: Let x* be the limit point of the sequence (x_k). Then there is a subsequence (xk_j) such that $x_{kj} \xrightarrow{j \to \infty} x^*$. Since $xk_j \in \Omega$, we have that $\eta(xk_j) = < f'(xk_j) - xk_j >$ and $< f'(xk_j), z(xk_j) > \leq < f'(xk_j), z > z \in \Omega$ (*).

W.l.o g. assume that $z(xk_j) \longrightarrow z^*$. Since $\eta(x_k) \xrightarrow{k \to \infty} 0$, we have that $\eta(xk_j) \xrightarrow{j \to \infty} 0$.

Then by the continuity of f' and the scalar product we get $< f'(x^*), z^* - x^* > = 0$

Which means $< f'(x^*), z^* > = < f'(x^*), x^* >$ But from(*) we get $< f'(x^*), z^* > \leq < f'(x^*), z > z \in \Omega$. Then $< f'(x^*), x^* > \leq f'(x^*) z >, z \in \Omega$.

Or $< f'(x^*)\ z-x^* > \geq 0\ \forall z \in \Omega$. Which is the necessary condition for x^* to be the minimum point of $f(x)$ in Ω.

Therem 4. 1 Let the function $f(x)$ be convex. Then

a) $\lim_{k\to\infty} f(x_k) = f^*$ where $f^* = \min_{x\in\Omega} f(x)$

b) the estimate $f(x_k) - f^* \leq C/k$ (C is positive constant) holds.

Proof : a) Since $f(x)$ is convex, the subgradient inequality will hold true. That means
$f^*-f(x) = f(x^*) - f(x) \geq < f'(x), x^*-x > \geq \min_{z\in\Omega} < f'(x), z-x > = \eta(x)$

Thus $0 \leq f(x) - f^* \leq -\eta(x)$. There fore for all k, $0 \leq f(x_k) - f^* \leq -\eta(x_k)$ - - - (4.10)

But by lemma 4.1, we have that $\eta(x_k) \xrightarrow{k\to\infty} 0$.

Hence $f(x_k)) \xrightarrow{k\to\infty} f^*$.

b. From (4.9) and (4.10) follows $f(x_{k+1}) - f(x_k) = (f(x_{k+1}) - f^*) - [f(x_k) - f^*]$

$$\leq \frac{-1}{8LC^2}\eta(x_k)^2 \quad \text{(by 4.9)}$$

$$\leq \frac{-1}{8LC^2}[f(x_k) - f^*]^2$$

Setting $\varphi_k := f(x_k)-f^*$, we have

$\varphi_{k+1} - \varphi_k \leq \frac{-1}{8LC^2}(\varphi_k)^2$ or $\varphi_{k+1} \leq \varphi_k \left[1-\frac{\varphi_k}{8LC^2}\right]$

or $\varphi_{k+1} \leq \varphi_k\ [\ 1 - \mu\varphi_k\ ]$ putting $\mu := \frac{1}{8LC^2}$

Taking $\varphi_k := \frac{\gamma_k}{k}$ we have $\frac{\gamma_{k+1}}{k+1} \leq \frac{\gamma_k}{k}\left[1 - \frac{\mu\gamma_k}{k}\right]$

or $\frac{\gamma_{k+1}}{\gamma_k} \leq \frac{k+1}{k}\left[1 - \frac{\mu\gamma_k}{k}\right]$ - - - (4. 11)

or $\frac{\gamma_{k+1}}{\gamma_k} \leq 1 + \frac{1}{k} - \mu\,(k+1)\frac{\gamma_k}{k^2}$

Now there are two cases

i) $\frac{\gamma_{k+1}}{\gamma_k} \leq 1$. That is $\gamma_{k+1} \leq \gamma_k$

That means the sequence (γ_k) is a monotone decreasing sequence. Since it is a non-negative sequence there exists ko such that $\gamma_k \leq 1\ \forall k \geq ko$. This implies that $\varphi_k \leq \frac{1}{k}$ $\forall\ k \geq ko$

Hence $f(x_k) - f^* \leq \frac{1}{k}$ $\forall k \geq ko$

ii) $\frac{\gamma_{k+1}}{\gamma_k} > 1$

Then $1 < 1 + \frac{1}{k} - \frac{\gamma_{k+1}}{k^2} \rho_k$ or $\frac{1}{k} - \frac{\gamma_{k+1}\gamma_k}{k^2} > 0$

Or $\frac{\mu (_{k+1}) \gamma_k}{k^2} < 1/k$ or $\gamma_k < \frac{1}{\mu} \frac{k}{k+1} < \frac{1}{\mu}$

More over from (4.11) we have $\frac{\mu_{k+1}}{\gamma_k} \leq \frac{k+1-\mu}{k}{}_{k+1} \frac{\gamma_k}{k^2} \leq \frac{k+1}{k} \frac{-1}{k} \leq \frac{k+1}{k} \leq 2$ for $k \geq 1$

Now only two situations are possible

1. There is only a finite number of indices k for which $\gamma_k < \frac{1}{\mu}$ Since $\gamma_{k+1} \leq 2\gamma_k$ for all great k, the sequence (γ_k) doesnot increase monotonically.

That means it remains bounded.

Hence $\gamma_k \leq C$ for some constant C . Or $\frac{\varphi_k}{k} \leq \frac{C}{k}$

There fore $f(x_k) - f^* \leq \frac{C}{k}$.

2. There is an infinite number of indices for which $\gamma_k < \frac{1}{}$

We shall denote the set of such indices k by y so that $\gamma_k < 1$ for $k \in y$

Let $j \notin y$. Then there will be indices $k_1,k_2, \in y$ such that $k_1 < j < k_2$ and $k \notin y$ $\forall k_1 < k < k_2$

Then $\delta_{k+1} \leq 2$, $\delta_{k1} \leq \frac{2}{\mu}$ and $\gamma_{i+1} \leq \gamma_i$ $\forall i = k_{i+1} - - - k_2 - 1$

There fore $\gamma_j \leq \frac{2}{\gamma}$, $j \notin y$.

That means, in this case as well , the sequence has a limit, a certain constant C.

Hence $\varphi_k = \frac{\gamma_k}{k} \leq \frac{C}{k}$ or $f(x_k) - f^* \leq \frac{C}{k}$

There fore in all cases we have that $f(x_k) - f^* \leq \frac{C}{k}$ ||

4. 4. Estimation of the convergence for strongly convex Region

Let the region Ω be strongly convex , ie there is a number $\delta > 0$ such that for any $x,y \in \Omega$ points $\frac{x+y}{2} + w \in \Omega$ for all w such that $\|w\| \leq \delta \|x-y\|^2$. Then

$$\eta(x) = \min_{z \in \Omega} < f'(x)\, z - x > \leq \min_{\|w\| \delta \|z(x) - x\|^2} < f'(x)\, \frac{z(x) + x}{2} + w - x > \text{(Taking } y = z(x)$$

$$= \min_{\|w\| \le \delta \| z(x) - x \|^2} < f'(x), \frac{z(x) - x + w}{2} >$$

$$= \tfrac{1}{2} < f'(x), z(x) - x > + \min_{\|w\| \le \delta \|z(x) - x\|^2} < f'(x), w > \quad *$$

But min < f '(x),w >= min < f '(x), -w> (Since the minimum is taken over ||w||and ||w|| = ||-w ||

$$\min_{\|w\| \le \delta \|2(x) - x\|^2} = \min_{\|w\| \le \delta \| z(x) - x \|} [-< f'(x), w >] = -\max < f'(x), w > = -\max < f'(x), \lambda f'(x) > = -\| f'(x) \|^2 \max \lambda$$

$$\|w\| \le \delta \|z(x) - x\|^2 \quad \|w\| \le \delta \|z(x)^{-x}\|^2 \quad \lambda \le \delta \frac{\|z(x) - x\|^2}{\|f'(x)\|} \qquad \lambda \le \delta \frac{\|z(x) - x\|^2}{\|f'(x)\|}$$

$$= - \|f'(x)\| \, z \, \delta \frac{\| z(x) - x\|^2}{\|f'(x)\|} = -\delta \|f'(x)\| \, \| z(x) - x\|^2$$ Then from (*) we have that

$$\eta(x) \le \tfrac{1}{2} < f'(x), z(x) - x > - \delta \|f'(x)\| \, \|z(x) - x\|^2$$

or $\eta(x) \le \tfrac{1}{2}\eta(x) - \delta \|f'(x)\| \|z(x) - x\|^2$

or $\tfrac{1}{2}\eta(x) \le \delta \|f'(x)\| \, \|z(x) - x\|^2$

Or $\dfrac{-\tfrac{1}{2}\eta(x)}{\|p(x)\|^2} \ge \delta \|f'(x)\|$ --- (4. 12)

Theorem 4.2. If f(x) is a convex function and the region Ω is strongly convex and if $\|f'(x)\| \ge \varepsilon_o > 0$ for all $x \in \Omega$, then the method of conditional gradient converges at the rate of geometric progression. ie $\|x_k - x^*\| \le c q_o^k$, $q_o < 1$.

proof: By (4.7) and (4.8), we get

$$\varphi_k - \varphi_{k+1} = f(x_k) - f(x_{k+1}) \ge \frac{1}{8L} \frac{\eta^2(x_k)}{\|p(x_k)\|^2}$$

(Since $\varphi_k - \varphi_{k+1} = f(x_k) - f(x_{k+1}) \ge -\dfrac{\alpha_k \eta(x_k)}{2}$ (by 4.7)

$$\ge \frac{1}{8L} \frac{\eta^2(x_k)}{\|p(x_k)\|^2} \quad \text{(By 4. 8)}$$

Using (4.12) and (4. 10) , we obtain

$$\varphi k - \varphi_{k+1} \ge \frac{1}{8L} \frac{\eta^2(x_k)}{\|p(x_k)\|^2} = \frac{1}{4L} \left(-\eta(x_k) \frac{(-\eta(x_k)}{2\|p(x_k)\|^2}\right)$$

$$\ge \frac{1}{4L} \delta \|f'(x_k)\| (-\eta(x_k)) \quad \text{(by 4.12)}$$

$$\geq \frac{\delta \in_o \varphi_k}{4L} \quad \text{(By assumption and 4.10)}$$

This implies that $\varphi_{k+1} \leq \left(1 - \frac{\delta \in_o}{4L}\right) \varphi_k$

There fore $\varphi_m \leq q^m \varphi$, where $q = 1 - \frac{\delta\varepsilon_o}{4L} < 1$

Because of the necessary and sufficient condition for a minimum,we have $< f'(x^*), x-x^* > \geq 0$

Then for all w with $\|w\| \leq \| x-x^* \|^2$, we get $< f'(x^*), \frac{x+x^*}{2} + w - x^* > \geq 0$

or $\frac{1}{2} < f'(x^*), x-x^* > + < f'(x^*), w > \geq 0$

or $\frac{1}{2} < f'(x^*), x-x^* > - \delta \|x - x^*\|^2 \|f'(x^*)\| \geq 0$

or $\frac{1}{2} < f'(x^*), x-x^* > \geq \delta \|x-^*\|^2 \|f'(x^*)\|$

But $\| f'(x^*) \| \geq \varepsilon_o$ and since f(x) is convex, we have

$$f(x) - f(x^*) \geq < f'(x^*), x - x^* > \text{(subgradient inequality)}$$

$$\geq 2\delta \|x-x^*\|^2 \|f'(x^*)\| \geq 2\delta\varepsilon_o \|x - x^*\|^2 \quad \text{--- (4.13)}$$

Hence $\|x_k - x^*\| \leq \sqrt{\frac{1}{2\delta\varepsilon_o\varphi_k}} \leq \sqrt{\frac{\varphi o}{2\delta\varepsilon_o}} \sqrt{q}^k$

Putting $C := \sqrt{\frac{\varphi_o}{2\delta\varepsilon}}$ $\quad q_o = \sqrt{\left(1 - \frac{\delta\varepsilon_o}{4L}\right)}$ we obtain $\|x_k - x^*\| \leq C q o^k$. ||

CHAPTER 5

NEWTON'S METHOD

5.1. Newton's Method with Step Adjustment

Consider the following optimization problem

(p): $f(x) \longrightarrow \min, x \in \Omega$. where f(x) is a convex smooth function and Ω is a convex compact set.

For solving this problem the iterative process

$$x_{k+1} = x_k + \alpha_k p_k, \alpha_k > 0 \text{ - - - } (5.1)$$

can be used in which the direction of motion $p_k = \bar{x}_k - x_k$ where x_k is the solution of the problem of minimization in a set Ω of a quadratic function

$\psi_k(x) = \langle f'(x_k), x - x_k \rangle + \frac{1}{2} \langle f'(x_k)(x - x_k), x - x_k \rangle$

and as α_k we take the maximum value of the parameter α, obtained by reducing the initialx = 1 until the parameter satisfies the inequality.

$$f(x_k + \alpha p_k) - f(x_k) \leq \varepsilon \alpha \psi_k(\bar{x}_k) \quad o < \varepsilon < 1 \text{ - - - } (5.2)$$

Theorem 5.1. If for them inimization of a convex twice continuously differentiable function f(x) in a convex compact set Ω, we used method (5.1) in which α_k and p_k are determined as described above, then (what ever the initial approximation x_o is choosen):

1. f(x) decreases monotonically
2. $\lim_{k \to \infty} f(x_k) = f(x^*) = \min_{x \in \Omega} f(x)$

Proof: Since $\psi_k(x)$ is continuous on a compact set Ω by weiers trass theorem $\psi_k(x)$ has a minimum point x_k.

More over, by convexity of Ω, from $x_k \in \Omega$ it follows that $x_{k+1} \in \Omega$.

This is because $x_{k+1} = x_k + \alpha_k p_k = x_k + \alpha_k(\bar{x}_k - x_k)$

$= \alpha_k \bar{x}_k + (1 - \alpha_k) x_k \in \Omega$ (Since $x_k \in [0,1]$)

By convexity of $\psi_k(x)$, we have $\psi_k(x_{k+1}) = \psi_k(\bar{x}_k \alpha_k + (1-\alpha_k)x_k) \leq \alpha_k \psi_k(\bar{x}_k) + (1-\alpha_k)$ $\psi_k(x_k)$ $(-\eta(x_k))$

But $\psi_k(x_k) = 0$. Hence $\psi_k(x_{k+1}) \leq \alpha_k\psi_k(\bar{x}_k)$ - - - 5.3

Now by Taylor's formula and (5.3), we get $f(xk+1) = f(x_k + \alpha_k p_k) = f(x_k) + \alpha x_k < f'(x_k), p_k > \frac{\alpha_k^2}{2} < f'(\theta)p_k, p_k >$

This implies that $f(x_{k+1} - f(x_k) = \alpha_k < f'(x_k), p_k > + \frac{\alpha_k^2}{2} < f'(\theta)p_k, p_k >$

$$\text{But } \psi_k(x_{k+1}) = < f'(x_k), x_{k+1} - x_k = \tfrac{1}{2} < f'(x_k)(x_{k+1} - x_k)x_{k+1} - x_k >$$

$$= < f'(x_k), x_{k+1} - x_k > + \tfrac{1}{2} < f'(x_k)(x_{k+1} - x_k, \alpha_k p_k >$$

$$= \alpha_k < f'(x_k), p_k > \alpha_k^2 < f'(x_k p_k, p_k >$$

$$\text{Hence } f(x_{k+1}) - f(x_k) = \psi k(x_{k+1}) \frac{\alpha_k^2}{2} < f'_k p_k, p_k >$$

Where $Fk = f'(x_{kc}) - f'(x_k)$, $x_{kc} = x_k + \theta(x_{k+1} - x_k)$, $\theta \in [0,1]$

Using 5.3) we have $f(x_{k+1}) - f(x_k \leq \theta_k\psi_k(x_k) + \frac{\alpha_k^2}{2} < F_k p_k, p_k >$

$$\leq \alpha_k\psi_k(x_k) \left[1 + \frac{\alpha_k}{2} \frac{\|F_k\| \, \| pk \|^2 }{\psi_k(x_k)} \right] \quad \text{- - - (5.4)}$$

Since x_k is a minimum point of $\psi_k(x)$, if $\psi_k(x_k) \neq 0$, then $\psi_k(x_k) \leq \psi_k(x_k) = 0$

Then with a certain x_k(choosing $\alpha_k \leq 2\,\frac{(1 - \varepsilon)\,\psi_k(\bar{x}_k)}{\|F_k\| \, \|p_k\|^2}$

the inequality $1 + \frac{\alpha_k}{2} \frac{\|F_k\| \, \|p_k\|^2}{\psi_k(x_k)} \geq \varepsilon$ holds - - - (5. 5)

Hence inequality (5. 2) holds true and there fore the described method of choosing α_k can be applied. It follows from (5.2) that $f(x_{k+1}) \leq f(x_k)$. Hence 1 is true.

To prove the second part we first show that $\psi_k(x_k) \xrightarrow{k \to \infty} 0$.

In a closed bounded set Ω continuous function $f'(x)$ has an upper bound. That means

$\|f'(x)\| \leq M$. Then $\|F_k\| = \|f'(x_{kc}) - f'(x_k)\| \leq \|f'(x_{kc}\| + \|f'(x_k) \leq 2M$

Moreover the vector pk has an upper bound too . Then $\|p_k\| \leq \max_{x,y\,\varepsilon\Omega} \|x - y\| =: d$

Now assume the contrary. Suppose $\psi_k(x_k) \leq -\beta < 0$

$$\text{Then } 1 + \frac{\alpha_k}{2} \frac{\|F_k\| \, \|p_k\|^2}{\psi_k(x_k)} \geq 1 - \alpha_k \frac{\|F_k\| \, \|pk\|^2}{2\beta} \geq 1 - \frac{\alpha_k 2Md^2}{2\beta}$$

$$= 1 - \frac{\alpha_k Md^2}{\beta}$$

Hence inequality (5.5) and hence (5.2) is always satisfied even with $\alpha_k = \frac{(1 - \varepsilon)\beta}{Md^2}$ $=: C > 0$. But from (5.2) follows $f(x_{k+1}) - f(x_k) \leq \varepsilon\,\psi_k(x_k) \leq -\varepsilon\beta\alpha = -\varepsilon c\beta$.

This means f(x) is not bounded below. But this is a contradiction to the fact that a continous function on a compact set Ω has a lower bound.

Hence $\psi_k(x_k) \leq \beta < 0$ is not true. There fore $\psi_k(x_k) \xrightarrow{k \to \infty} 0$

That means at any limit point of the sequence (5.1) the necessary and sufficient condition for a minimum of the function f(x) in a set Ω is fulfilled.

Taking this in to account, the last assumption of the theorem can be proved as in theorem 4.1.||

5.2. Rates of Convergence for Strongly Convex Functions

Theorem 5.2. If in addition to conditions of theorem 5.1, function f(x) is strongly convex, ie $m\|y\|^2 \leq \langle f''(x)y, y\rangle \leq M\|y\|^2$, $m > 0$, $x \in \Omega$, $y \in E^n$, then sequnce (5.1) converges to the (5.6) solution at a super linear rate.

Proof: Since f(x) is strongly convex the solution exists and in unique. More over since $\bar{x}_k$ is a minimum point of $\psi_k(x)$ and $\psi_k(x)$ is convex, by the characterizing theorem of convex optimization $\langle \psi'_k(\bar{x}_k), x_k - \bar{x}_k\rangle \geq 0$ Or $\langle \psi'_k(\bar{x}_k), \bar{x}_k - x_k\rangle \leq 0$

As $\psi'_k(x) = f'(x_k) + f'(x_k)(x - x_k)$, we have

$\langle f'(x_k), \bar{x}_k - x_k\rangle + \langle f'(x_k)(\bar{x}_k - x_k), \bar{x}_k - x_k\rangle \leq 0$

Or $\langle f'(x_k), p_k\rangle \leq -\langle f'(x_k)p_k, p_k\rangle$ - - (5.7)

By (5.6) and (5.7) , we obtain

$$\begin{aligned}
\psi_k(x_k) &= \langle f'(x_k), \bar{x}_k - x_k\rangle + \tfrac{1}{2}\langle f'(x_k)\bar{x}_k - x_k, \bar{x}_k - x_k\rangle \\
&= \langle f'(x_k), p_k\rangle + \tfrac{1}{2}\langle f'(x_k)p_k, p_k\rangle \\
&\leq -\langle f'(x_k)p_k, p_k\rangle + \tfrac{1}{2}\langle f'(x_k)p_k, p_k\rangle \\
&= -\tfrac{1}{2}\langle f''(x_k)p_k, p_k\rangle \\
&\leq -\frac{m^2}{2}\|p_k\|^2 \quad \text{- - - (5.8)}
\end{aligned}$$

From (5.4) we have

$$\begin{aligned}
f(x_{k+1}) - f(x_k) &\leq \alpha_k\psi_k(x_k) + \frac{\alpha_k^2}{2}\|F_k\|\,\|p_k\|^2 \\
&\leq -\frac{m\alpha_k}{2}\|p_k\|^2 + \frac{\alpha_k^2}{2}\|F_k\|\,\|p_k\|^2
\end{aligned}$$

$$= -\frac{m}{2}\alpha_k \|p_k\|^2 \left[1 - \frac{\alpha_k \|F_k\| \|p_k\|^2}{m \|p_k\|^2}\right]$$

$$= -\frac{m\alpha_k}{2}\| p_k\|^2 \left[1 - \frac{\alpha_k \|F_k\|}{m}\right] \quad \text{--- (5.9)}$$

since $\psi_k(x_k) \xrightarrow{k\to\infty} 0$ (By theorem 5. 1),we have by 5.8 that $\|p_k\| \xrightarrow{k\to\infty} 0$
More over since f " is uniformly continuous $\| F_k \| \xrightarrow{k\to\infty} 0$

But from (5. 9) it follows that $\forall k \geq N_1(\varepsilon)$, the inequality (5.2) will be satisfied with $\alpha_k = 1$. That is method (5.1) is changed in to the usual Newton' s method with a unity.
For $\alpha > N_1(\varepsilon)$, we have that $\alpha_k = 1$ and hence

$\psi_k(x_k) = \psi_k(x_{k+1}) = < f'(x_k), x_{k+1} > + ½ < f'(x_k)(x_{k+1}) - x_k), x_{k+1} - x_k >$

$\geq < f'(x_k), x_{k+1}) - x_k >$

$= < f'(x_{k-1}), x_{k+1}) - x_k > + < f'(x_k) - f'(x_{k-1}), x_{k+1} - x_k >$

By lagranges formula for operators $< F(x+h) - F(x), y >$
$= < F'(x+\theta h) h, y >, 0 \leq \theta \leq 1$) we have that

$< f'(x_k) - f'(x_{k-1}), x_{k+1} - x_k > = < f'(x_{k-1}) + \alpha_{k-1} p_{k-1}) - f'(x_{k-1}), x_{k+1} - x_k >$

$= < f'(x_{k-1} + \theta(x_k - x_{k-1}\, x_k - x_{k-1}, x_{k+1}) - x_k >$

Hence $\psi_k(x_k) \geq < f'(x_{k-1}), x_{k+1} - x_k > + < f''(x_{k-1} + \theta(x_k - x_{k-1})(x_k - x_k), x_{k+1} - x_k >$

$= < f'(x_{k-1}) + f'(x_{k-1})(x_k - x_{k-1}), x_{k+1} - x_k > + < [f'(x_{k-1} + \theta(x_k - x_{k-1}))](x_k - x_{k-1}, x_{k+1} - x_k >$ --- (5.10)

Denoting $\Phi_k =: < f'(x_{k-1} + \theta(x_k - x_{k-1})(x_k - x_{k-1}), x_{k+1} - x_k >$

From $\psi_{k-1}(x_k) = f'(x_{k-1}), x_k - x_{k-1} > ½ < f''(x_{k-1})(x_k - x_{k-1}), x_k - x_{k-1} >$

We have $\psi'_{k-1}(x_k) = f'(x_{k-1}) + f''(x_{k-1})(x_k - x_{k-1})$

Then $< f'(x_{k-1}) + f''(x_{k-1})(x_k - x_{k-1}), x_{k+1} - x_k > = < \psi'_{k-1}(x_k), x_{k+1} - x_k >$
There fore $\psi_k(x_k) \geq < \psi'_{k-1}(x_k), x_{k+1} - x_k > + < \Phi_k(x_k - x_{k-1}), x_{k+1} - x_k >$
Since $\psi_{k-1}(x_k) = < f'(x_{k-1}), x_k - x_{k-1} > + ½ < f''(x_{k-1})(x_k - x_{k-1}), x_k - x_{k-1} >$

$= < f'(x_{k-1}), p_{k-1} > + ½ < f''(x_{k-1}) p_{k-1}, p_{k-1} >$

$= < f'(x_{k-1}), x_{k-1} - x_{k-1} > + ½ < f''(\overline{x}_{k-1})(x_{k-1} - x_{k-1}), \overline{x}_{k-1} - x_{k-1} >$

$= \psi_{k-1}(\overline{x_{k-1}})$

$= \min_{x \in \Omega} \psi_{k-1}(x)$

We have for any $x \in \Omega$ that $< \psi'_{k-1}(x_k), x - x_k > \geq 0$ (characterizing theorem of convex optimization) consequently $< \psi'_{k-1}(x_k), x_{k+1} - x_k > \geq 0$ holds .
There fore $\psi_k(\overline{x_k}) \geq < \Phi_k(x_k - x_{k-1}), x_{k+1} - x_k >$

Or $-\psi_k(x_k) \leq < \Phi_k(x_k - x_{k-1}), x_k - x_{k+1} >$

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

$= \|\Phi_k\| \|p_{k-1}\| \|p_k\|$ --- (5.11)

But from (5.8) we have $\|p_k\|2 \leq -2\,\psi_k(x_k) \leq \frac{2}{m} \frac{\|\Phi_k\|\,\|pk-1\|\,\|pk\|}{m}$

Or $\|p_k\| \leq \frac{2\|\Phi_k\|}{m}\|p_{k-1}\|$ Or $\|x_{k+1} - x_k\| \leq 2\|\Phi_k\|\,\|x_k - x_{k-1}\|$ - - - (5.12)

Since f "(x) is uniformly continuous on the compact set Ω We have that $\|\Phi k\| \xrightarrow{k\to\infty} 0$. Hence there is anumber N (ε) such that with k ≥ N(ε) we find $\lambda_k = \frac{2}{m}\|\Phi\| < 1$. If we take $\|x_N - x_{N-1}\| = : C$ and $1 - \lambda_N = : \gamma > 0$

Then $\|x_i - x_{N+1}\| = \|x_{N+l+1} - x_{N+l} + x_{N+l+2} - x_{N+l+1} + - - - + x_i - x_{i-1}\|$

$\leq \|x_{N+l+1} - x_{N+1}\| + \|x_{N+l+2} - x_{N+l+1}\| + - - - + \|x_i - x_{i-1}\|$

$= \sum_{k=N+1}^{i-1} \|x_{k+1} - x_k\|$

But $\|x_{N+l+1} - x_{N+1}\| \leq \frac{2\|\Phi_{N+1}\|}{m}\|x_{+1} - x_{N+1-1}\|$

$\leq \frac{2\|\Phi_{N+1}\|}{m} \cdot \frac{2\|\Phi_{N+1-1}\|}{m}\|x_{N+l-1} - x_{N+1-2}\|$

$\leq \frac{2\|\Phi_{N+1}\|}{m} \cdot \frac{2\|\Phi_{N+1-1}\|}{m} \cdot - - - \cdot \frac{2\|\Phi_N\|}{m}\|x_N - x_{N-1}\|$

Denoting $\frac{}{m} =: \lambda_{N+l} =: \frac{2\|\Phi_{N+1}\|}{m}$

$\lambda_{N+1-1} =: \frac{2\|\Phi_{N+1-1}\|}{m}$

$\lambda N =: \frac{2\|\Phi_N\|}{m}$

$C =: \|x_N - x_{N-1}\|$

$= C\,\lambda_N\lambda_{N+1} - - - \lambda_{N+l}-1\lambda_{N+1}$

and $\|x_{N+l+2} - x_{N+l+1}\| \leq C\lambda_N\lambda_{N+1} - - - \lambda_N+l\lambda_{N+l+1}$

Hence $\sum^{i-1} \|x_{k+1} - x_k\| \leq C\lambda_N\lambda_{N+1} - - - \lambda_{N+l}\,[\,1 + \lambda N{+}l{+}1 + \lambda_N{}^2{}_{+l+1} + - - - + \lambda_{N+l+1}{}^{-(N+l+1)}]$

(this is because $\lambda_{N+l+k} \leq \lambda_{N+l+1}$ $\forall k = \frac{2}{2}, - - - i$)

$\leq \frac{c_1}{1-\lambda_N} \lambda_N - - - \lambda_{N+l}$ (since $1 + \lambda_{N+l+1} + \lambda_{N+l+1} + - - - - \lambda_{N+l+1}$

$\leq \infty$ $\sum_{i=0}^{i} \lambda_N = \frac{1}{1-\lambda N}$

$= C\,\lambda_N\lambda_{N+1} - - - - \lambda_{N+1}$ by denoting $C =: \frac{C_1}{1-\lambda N}$

Hence $\|x_i - x_{N+1}\| \xrightarrow{i,l\to\infty} 0$.

That means, the sequence (x_k) is a cauchy sequence and since the space E^n is complet (finite dimensional) the sequence(x_k) has a limit $x^* \varepsilon \Omega$, and

$\|x_{N+1} - x^* \| \leq C \lambda_N \lambda_{N+1} --- \lambda_{N+l}$ (5.13)

By theorem (5.1), $\lim f(x_k) = f(x^*) = \min_{x \varepsilon \Omega} f(x)$.

Therefore (5.1) converges to the solution and the rate of convergence is super linear. ||

Theorem 5.3. If the conditions of theorem 5.2 are fulfilled and besides, matrix f " satisfies the lipschitz, condition on the set Ω with constant R, then the sequence (5.1) (in which α_k and p_k are choosen by the method described above) converges to a solution at a quadratic rate.

Proof: If the second derivative f " (x) satisfies Lipschitz' condition with constant R, then inequality (5.12) takes the form $\|x_{k+1} - x_k\| \leq \frac{2R}{m} \|x_k - x_{k-1}\|^2$ --- (5.14)

(this is because $\|f''(x_{k-1} + \theta(x_k - x_{k-1}) - f''(x_{k-1})\|$

$$\leq R \|x_{k-1} + \theta(x_k - x_{k-1}) - x_{k-1}\|$$
$$= R \|\theta(x_k - x_{k-1})\|$$
$$\leq R \|x_k - x_{k-1}\| \text{ (Since } 0 \leq \theta \leq 1)$$

put $\beta_k := \frac{2R}{m} \|x_{k+1} - x_k\|$

Since $\|x_{k+1} - x_k\| \xrightarrow{k\to\infty} 0 \quad \forall \varepsilon > 0 \ \exists L(\varepsilon)$ such that $\beta_k < 1 \ \forall k > L(\varepsilon)$

From (5.14) with $k > L(\varepsilon)$, we have $\frac{m}{2R} \beta k \leq \beta_k - 1 \|x_k - x_{k-1}\|$

Or $\beta_k \leq \beta_{k-1} \frac{2R}{m} \|x_k - x_{k-1}\|$. This implies that $\beta k \leq \beta k^2 - 1$

Similarly we have $\beta_{k-1} \leq \beta_k^2 -_2$ and soon

In general we have $\beta_k \leq \beta_k^2 -_1 \leq \beta_k^{22} -2 \leq --- \beta^{2k-L} L$

Consequently, for any $i > L + 1, l = 0,1, ---$ we have

$$\|x_i - x_{L+1}\| \leq \sum_{k=L+1}^{i} \|x_{k+1} - x_k\| \leq \frac{m}{2R} \sum_{k=L+1}^{i-1} \beta^{2^{k-L}} \quad \text{(Since } \|x_{k+1} - x_k\| = \frac{m}{2R}\beta_k$$

$$= \frac{m}{2R} \sum_{s=1}^{i-1} \beta^{2^{k-L}}_L \qquad \text{and } \beta_k \leq \beta^{2^{k-L}}_L$$

Since $x_k \to x^*$, we have $\|x_{L+1} - x^*\| = \lim_{\infty} \|x_{L+1} - x_i\|$

$$\leq \frac{m}{2R} \sum_{s=1}^{\infty} \beta^{2s} \beta^{2l} \leq CL, \ C < \infty \text{ (since the series } \sum_{s=1}^{\infty} \beta^{2s} L \text{ (converges)}$$

Lemma 5.1. If a strongly convex function f(x) is being minimized and α_k in (5.1) is choosen under condition $f(x_k + \alpha_k p_k) = \min f(x_k + \alpha p_k)$ - - - (5.15)

$$0 \leq \alpha \leq 1$$

then $x_k \to x^*$ and $\alpha_k \to 1$ as $k \to \infty$

Proof: By Taylor's formula

$$f(x_{k+1}) - f(x_k) = \alpha x < f'(x_k), p_k > \frac{\alpha k^2}{2} < f''(x_{kC})\, p_k, p_k >$$

with the value of α_k satisfying (5.15), the right hand side of the last equality considered as a function of variable α mustattain its minimum.

That means $\varphi(\alpha_k) = \alpha_k < f'(x_k), pk > \frac{\alpha_k^2}{2} < f''(x_{kC})\, p_k, p_k >$ attains its minimum at α_k.

Then $\varphi'(\alpha_k \quad) = 0$

Or $< f'(x_k), p_k > \alpha_k < f''(x_{kC})\, p_k, p_k > = 0$

This implies that $1 \geq \alpha_k = - \frac{< f'(x_k), p_k >}{< f''(x_{kC}) p_k, p_k >}$

$$\geq - \frac{< f'(x_k), p_k >}{M \| p_k \|^2} \quad \text{(by 5.6)}$$

$$\geq \frac{< f''(xk)\, p_k, p_k >}{M \|p_k\|^2} \quad \text{(by 5. 7)}$$

$$\geq \frac{m \|p_k\|^2}{M \|p_k\|^2} = \frac{m}{M} \quad \text{(by 5.6)}$$

Thus $\alpha_k \geq C > 0$.

There fore in a similar way as in the proof theorem 5.1, we can show that $\Psi_k(x_k) \xrightarrow{k \to \infty} 0$. That means the sequence (5.1) in which α_k is choosen under condition (5.15) converges to the solution. At the same time by theorem 5.2, $\| p_k \| \xrightarrow{k \to 0} 0$ and $\|F_k\| \xrightarrow{k \to \infty} 0$

Now it remains to show that $\alpha_k \xrightarrow{k \to \infty} 1$.

We know that $f(x_{k+1}) = \psi_k(\alpha_{k+1}) + \frac{\alpha k^2}{2} < F_k P_k, P_k >$

But $\psi_k(x_{k+1}) = < f'(x_k), x_{k+1} - x_k > \frac{1}{2} < f''(x_k)(x_{k+1} - x_k), x_{k+1} - x_k >$

$= < f'(x_k), x_k - x_k > + < f'(x_k), x_{k+1} - x_k > + \frac{1}{2} < f''[x_k)(x_{k+1} - x_k) + x_k - x_k], (x_{k+1} - x_k) + x_k - x$

$= < f'(x_k), x_k - x_k > + < f'(x_k), x_{k+1} - x_k > + \frac{1}{2} < f''(x_k)(x_{k+1} - x_k) x_{k+1} - x_k > + \frac{1}{2} < f''(x_k)(x_{k+1} - x_k), x_k - x_k > + \frac{1}{2} < f''(x_k)(x_k - x_k), x_{k+1} - x_k > + \frac{1}{2} < f''(x_k)(x_k - x_k), x_k - x_k >$

$= < f'(x_k), p_k > + f''(x_k), x_{k+1} - x_k > + < f''(x_k)(x_k - x_k), x_{k+1} - x_k > + \frac{1}{2} < f''(x_k)(x_{k+1} - x_k), x_{k+1} - x_k) + \frac{1}{2} < f'(x_k) p_k, p_k >$

$= < f'(x_k), p_k > + \frac{1}{2} < f''(x_k) p_k, p_k > + < f'(x_k) + f''(x_k)\, p_k, x_{k+1} - x_k > + \frac{1}{2} < f''(x_k)(x_{k+1} - x_k), x_{k+1} - x_k >$

$= \psi_k(x_k) + < \psi'_k(x_k), x_{k+1} - x_k > \frac{1}{2} < \psi''(x_k)(x_k)(x_{k+1} - x_k), x_{k+1} - xk)$

But $x_{k+1} - x_k = x_k + \alpha_k p_k - x_k = x_k - x_k + \alpha_k p_k$

$$= - p_k + \alpha_k p_k$$

$$= (\alpha_{k-1}) p_k$$

Then $f(x_{k+1}) = \psi_k(x_k) + < \psi_k'(x_k), x_{k+1} - x_k > +(\frac{\alpha_{k-1}}{2})^2 < \psi_k''(x_k)p_k > + \frac{\alpha_k^2}{2} < F_k p_k, p_k >$

Now since $< \psi_k'(x_k), x_{k+1} - x_k > \geq 0$ and $< \psi''_k(x_k)p_k, p_k >$

$= < f''(x_k)p_k, p_k > \geq m \ ||p_k||^2$ and $< F_k p_k, p_k > = 0 \ (||p_k||^2$ (Since $||F_k|| \longrightarrow 0$ as $||p_k|| \longrightarrow 0$) we have that the minimum of $f(x_{k+1})$

$- \psi_k(x_k)$ can be attained only as $\alpha_k \longrightarrow 1$.

Other wise with any K, we should have $1 - \alpha_k \geq \beta > 0$ and at the same time

$f(x_{k+1}) - \psi_k(x_k) = 0 \, (|| p_k ||^2) || > 0$ Where as with $\alpha = 1$, the difference $f(x_k) - \psi_k(x_k) = \frac{1}{2} < F_k P_k, p_k > = 0 \, (|| p_k ||^2)$, (Since $x_{k+1} - x_k = 0$ ($\alpha = 1$) and $\alpha_{k-1} = 0$ The terms $< \psi'_k(x_k), x_{k+1} - x_k > 0$ and $(\frac{\alpha_{k-1}}{2})^2 < \psi''_k(x_k)p_k, p_k > = 0$)

That is with sufficiently great K we should always have $f(x_k) < f(x_{k+1})$; which is a contradiction to the choice of α_k(since $f(x_{k+1}) = \min_{0 \leq \alpha \leq 1} f(x_k + \alpha p_k)$)

Theorem 5.4 If the function f(x) satisfies the requirements of theorem 5.1 and in method (5.1) parameter α_k is choosen under condition (5.15), then $x_k \rightarrow x^*$ at a super linear rate.

Proof: By estimates (5.3) and (5.8), we have

$\psi_k(x_{k-1}) \leq \alpha_k \psi_k(x_k) \leq \alpha_k^2 \psi_k((x_k) \leq \frac{\alpha_k^2}{2} m \ ||p_k||^2$

That is $\psi_k(x_{k+1}) \leq -\frac{m}{2} \ ||x_{k+1} - x_k||^2$ --- (5.16)

(Since $||\alpha_k p_k||^2 = ||x_{k+1} - x_k||^2$)

More over $\psi_k(x_{k+1}) = < f'(x_k), x_{k+1} - x_k > + \frac{1}{2} < f''(x_k) x_{k+1} - x_k \, x_{k+1} - x_k >$

$\geq < f'(x_k), x_{k+1} - x_k >$ (Since $< f''(x_k) x_{k+1} - x_k, x_{k+1} - x_k > \geq 0$

$= < f'(x_k), x_{k+1} - x_{k-1} > + < f'(x_k), x_{k-1} - x_k >$

Since at a point x_k the minimum of f(x) is attained in the direction $x_{k-1} - x_{k-1}$, we have $< f'(x_k), x_{k-1} - x_k > \geq 0$. Hence $\psi_k(x_{k+1}) \geq < f'(x_k), x_{k+1} - x_{k-1} >$

$= < f'(x_{k-1}), x_{k+1} - x_{k-1} > + < f'(x_k - f''(x_{k-1}), x_{k+1} - x_{k-1} >$

By using Lagranges formula for operators, we have

$\psi_k(x_{k+1}) \geq < f''(x_{k-1}) + f''(x_{k-1}) \ (x_k - x_{k-1}), x_{k+1} - x_{k-1} > + < \Phi_k (x_k - x_{k-1}), x_{k+1} - x_{k-1} >$

Where $\Phi_k = f''(x_{k-1} + \theta(x_k - x_{k-1})) - f''(x_{k-1}), \theta \, \varepsilon \, [0,1]$

using $x_k - x_{k-1} = \alpha_k -(x_{k-1} - \bar{x}_{k-1})$ we find that $\psi_k(x_{k+1}) \geq < f'(x_{k-1}) + f''(x_{k-1})(x_{k-1} - x_{k-1}) + \Phi_k(x_k - x_{k-1}), x_{k+1} - x_{k-1} >$

Since $\psi_{k-1}(x_{k-1}) = \min_{x \varepsilon \Omega} \psi_{k-1}(x)$, we have by the characterizing theorem of convex optimization $< \psi'_{k-1}(\bar{x}_{k-1}), \bar{x}_{k-1} - x > \leq 0 \ \forall \ x \varepsilon \Omega$. consequently

$< \psi'_{k-1}(\bar{x}_{k-1}), x_{k-1} - x_{k+1} > = < f'(x_{k-1}) f''(x_{k-1})(x_{k-1} - x_{k-1}), x_{k-1} - x_{k+1} > \leq 0$

Or $< f'(x_{k-1}) + f''(x_{k-1})(x_{k-1} - x_{k-1}), x_{k+1} - x_{k-1} > \geq 0$

Since $< (\alpha_{k-1} - 1) f''(x_{k-1})(x_{k-1} - x_{k-1}) + \Phi_k(x_k - x_{k-1}), x_{k+1} - x_{k-1} >$

$= < (\alpha_{k-1} - 1) f''(x_{k-1}) \left(\frac{x_k - x_{k-1}}{\alpha_{k-1}} \right) + \Phi_k(x_k - x_{k-1}), x_{k+1} - x_{k-1} >$

$= < \left(\frac{\alpha_{k-1} - 1}{\alpha_{k-1}} f''(x_{k-1}) + \Phi_k \right)(x_k - x_{k-1}), x_{k+1} - x_{k-1} >$

We have $\psi_k(x_{k+1}) \geq < \left(\frac{\alpha_{k-1} - 1}{\alpha_{k-1}} f''(x_{k-1}) + \Phi_k \right)(x_k - x_{k-1}), x_{k+1} - x_{k-1} >$

Hence setting $b_k =: \| \alpha_{k-1} - 1 f''(x_{k-1}) + \Phi_k) \|$, We have

$- \psi_k(x_{k+1}) \leq b_k \|x_k - x_{k-1}\| \|x_{k+1} - x_{k-1}\|$

As $\left(\frac{1 - \alpha_{k-1}}{\alpha_{k-1}} \right)(x_k - x_{k-1}) = \left(\frac{1}{\alpha_{k-1}} - 1 \right)(x_k - x_{k-1})$

$= \left(\frac{x_{k-1} - x_{k-1}}{x_k - x_{k-1}} - 1 \right)(x_k - x_{k-1})$

$= x_{k-1} - x_{k-1} - (x_k - x_{k-1}) = x_{k-1} - x_k$

Now denoting $[1 - \alpha_{k-1}] b_k =: C_k$, we have $- \psi_k(x_{k+1}) \leq b_k \|x_k - x_{k-1}\| \| x_{k+1} - x_k \| + C_k \|x_k - x_k - 1\|^2$

Since $\alpha_k \rightarrow 1$, $\|p_k\| \rightarrow 0$ (Lemma 5.1) , we have $b_k \rightarrow 0$ $c_k \rightarrow 0$

Comparing this with (5.16), we estabilish that $\|x_{k+1} - x_k\|^2 \leq \xi_k \|x_k - x_{k-1}\| \|x_{k+1} - x_k\| + \delta_k \|x_k - x_{k-1}\|^2$

Where $\xi_k =: \frac{2b_k}{m}$, $\delta_k =: \frac{2c_k}{m}$

(this is because by (5.16), we have $- \psi_k(x_{k+1}) \geq \frac{m}{2} \|x_{k+1} - x_k\|^2$

$\Rightarrow \frac{m}{2} \|x_{k+1} - x_k\|^2 \leq b_k \|x_k - x_{k-1}\| \|x_{k+1} - x_k\| + c_k \|x_k - x_{k-1}\|^2$

$\Rightarrow \|x_{k+1} - x_k\|^2 \leq \frac{2b_k}{m} \|x_k - x_{k-1}\| \|x_{k+1} - x_k\| + 2c_k \|x_k - x_{k-1}\|^2$

Finally, having solved the quadratic inequality obtained for $\|x_{k+1} - x_k\|$, we find that $\|x_{k+1} - x_k\| \leq \mu k \|x_k - x_{k-1}\|$

where $\mu_k =: \frac{\xi_k}{2} + \sqrt{\frac{\xi_k^2}{4} + \delta_k} \xrightarrow{k \to \infty} 0$

The remaining part of the proof is performed just as in theorem 5.1. ||