Republic of Iraq
Ministry of Higher Education and scientific Research
Babylon University
College of Education for Pure Sciences
Department of Mathematic

# Numerical Optimization : Line Search Methods

A Research
Submitted to the Mathematic Department University of Babylon as a Partial Fulfillment of the Requirement for the Degree of Higher Diploma Education/ Mathematic

By

**Ahmed Lafta Qasim Lafta**

Supervised By

**Dr. Ahmed Sabah Al-jilawi**

1443 A.H 2021

" You must not lose faith in humanity.
Humanity is an ocean; if a few drops
of the ocean are dirty, the ocean does not become dirty. "

Mahatma Gandhi

# Supervisor's certificate

I certify that the research entitled
" **Numerical Optimization :Line Search Methods**"
for the student "**Ahmed Lafta Qasim**" was prepared under my supervision in University of Babylon / college of Education for pure science as a partial requirement for the Degree of Higher Diploma Education/ Mathematic.

Signature:

Name: Dr. Ahmed Sabah Al-jilawi

Scientific grade : Asst Prof

Date : / /2021

According to the available recommendation, I forward this project for debate by the examining committee.

Signature:

Name: Dr. Azal J. Musa

Head of mathematics Department

Title : Asst Prof

Date : / /2021

## Certification of Scientific Expert

I Certify that I have read the Scientific content of this **"Numerical Optimization :Line Search Methods**"and I have approved this project is qualified for debate.

Signature :

Name : Dr. Hasnaa H. Shahid

Title : Lecturer

Date : / / 2021

## Certification of Linguistic Expert

I Certify that I have read the project entitled " **Numerical Optimization :Line Search Methods**" and corrected its grammatical mistakes; therefore , it has qualified for debate.

Signature :

Name : Dr. Mais F. Al-jabbawi

Title : Lecturer

Date : / / 2021

## Examining Committee Certification

We certify that we have read this research entitled
**" Numerical Optimization :Line Search Methods "**
as examining committee examined the student "**Ahmed Lafta Qasim**" in its contents and that in our opinion it is adequate for the partial fulfillment of the requirement for the Degree of Higher Diploma Education/ Mathematics.

Chairman
Signature :
Name: Dr. Zahir A. H. Al-shanoon
Title : Prof
Date : / /2021

Member
Signature :
Name : Dr. Hawraa A. Al- murieb
Title : Lecturer
Date : / /2021

Member
Signature :
Name : Dr.Hussein A. Wasi
Title : Lecturer
Date : / /2021

Member / Supervisor
Signature :
Name: Dr. Ahmed S. Al-jilawi
Title : Asst Prof
Date : / /2021

Approved by the dean of college of education for pure science:

Signature

Name : Dr. Bahaa Hussein Salih Rabee

Title: Professor

Date : / /2021

## Dedication

To the people who I love and respect, my Father, Mother, Wife, brothers, and Sons . A special dedication is to Martyrs of the October Revolution

# Acknowledgments

I would like to express my thanks to the Merciful Allah and then thanks my supervisor **Dr. Ahmed Sabah Al-jilawi** for standing with me and supporting me to bring this work into big Success and thanks to all those who contributed to this wo

## Abstract:

The purpose of this research is to solve the optimization problems, one of the most important techniques is the line search technique. Initially, we implemented Bisection Method, Newton's Method, Secant Method, and Golden Section Method, Also, the line search strategy is one of two basics iterative approaches to find a local minimum $x^*$ of the objective function $f : R^n \to R$ . The line search approach find a descent direction along which the objective function $f$ will be and then computes a step size that determined how $x$ should move along. Also, the comparison between the results of the Steepest Descent Method and the Conjugate Gradient Method.

list of content


| Title | Page Number |
|---|---|
| CHAPTER ONE ( Preleminories of Optimization: Line Serch Method) | 1 |
| 1.1 Introuduction of optimization | 2 |
| 1.2 Introuduction of Line Search Method | 3 |
| 1.3 Definition Inner Product | 4 |
| 1.4 Definition Norm | 4 |
| 1.5 Definition Convex Sets | 4 |
| 1.6 Definition Convex Functions | 4 |
| 1.7 Definition Closed Function | 6 |
| 1.8 Definition Epigraph | 6 |
| 1.9 The Proximal Operator | 6 |
| 1.10 The Linear Programming Problem | 7 |
| 1.11 Bisection Method | 7 |
| 1.11.1 Bisection Algorithm | 7 |
| 1.7 Features of Bisection | 8 |
| Example 2 | 8 |
| 1.12 Newton's Method | 9 |
| 1.12.1 Features of Newton's Method | 9 |
| 1.13 Secant Method | 11 |
| 1.14 Golden Section Method | 11 |
| CHAPTER TWO (Line Search Technique) | 14 |
| 2.1 Introduction | 15 |
| 2.2 Step Length | 15 |
| 2.3 The Wolfe Conditions | 18 |
| 2.4 The Goldsteein Conditions | 20 |
| Chpter THREE (Comparison of Steepest Descent Method and Conjugate Gradient Method) | 21 |

| | |
|---|---|
| 3.1 Introduction | 22 |
| 3.2 The Steepest Descent Methods | 23 |
| 3.3 Conjugate Gradient Method | 24 |
| 3.4 The Comparison of Methods | 26 |
| 3.6 Backtracking-Armijo Line Search | 28 |
| 3.7 Computing a Search Direction $d^K$ | 29 |
| 3.8 The Method of Steepest Descent | 29 |
| Example 3 | 30 |
| Example4 | 31 |
| References | 34 |

# Chapter One
# Preleminories of Optimization: Line Serch Method

## 1 .1 Introduction of Optimization

Optimization, in general, is ″the problem of minimizing or maximizing a function under a set of constraints problems with optimization abound . Optimization is a precise procedure using design constraints and criteria to enable the planner to find the optimal solution ″ .[3]

″ Optimization techniques have been applied in numerous fields to deal with different practical problems. Optimization algorithms are a fundamental and successful tool in mathematical programming to reach a solution, generally with the assistance of a computer ″.[9]

″ Optimization algorithms start with an initial estimate of the value of the variables and by an iterative technique generates a sequence of improved estimates, or iterates, until an optimal solution is reached, Optimization problems of sorts arise in all quantitative disciplines from computer science and engineering to operations research and economics, A great algorithm should be efficient, fast, accurate, and robust. It should generate a good approximation of an optimal solution ″. [13]

## 1 .2 Introuduction of Line Search Method

Each new time Optimization teaching is being approached to hand issues that are a lot bigger and complex than before , as a result of the wide (and developing) Optimization of improvement in science. Improvement is a significant device in choice science and the investigation of actual frameworks. The way toward recognizing target factors and requirements for a given issue is known as modeling. This chapter is to find a better line search technique and different combinations of optimization and line search algorithms. Line search technique on the off chance that it looks for the base of a descent direction vector that, when processed iteratively with a sensible advance size [7] . The solution methods for unconstrained optimization problems can be broadly classified into gradient-based and nongradient-based search methods. As the name suggests, gradientbased methods require gradient information in determining the search direction [3] . The gradient-based methods discussed in this chapter are steepest descent. Davidon , Fletcher , Powell (DFP), Broyden, Fletcher, and Goldfarb, Shanno (BFGS), and Newton. Methods the search direction computed by these uses the gradient information, Hessian information, or a combination of these two.[2]

**1.3 Definition Inner Product [6]:**

A function $\langle\cdot,\cdot\rangle: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is an inner product if :

(i) $\langle x, x\rangle > 0, \langle x, x\rangle = 0 \Leftrightarrow x = 0$ (positivity).

(ii) $\langle x, y\rangle = \langle y, x\rangle$ (symmetric).

(iii) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ (additivity).

(iv) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$ for all $\alpha \in \mathbb{R}$ (homogeneity).

**1.4 Definition Norm [6]** :

A function $\|\cdot\|: \mathbb{R}^n \to \mathbb{R}$ is said to be a norm, if it satisfies the following Properties :

(i) $\| x \| \geq 0, \forall x \in \mathbb{R}^n; \| x \| = 0$ if and only if $x = 0$.
(ii) $\| \alpha x \| = |\alpha| \, \| x \|, \forall \alpha \in \mathbb{R}, \forall x \in \mathbb{R}^n$.

(iii) $\| x + y \| \leq \| x \| + \| y \|, \forall x, y \in \mathbb{R}^n$.

**1.5 Definition Convex Sets [4] .**

Let $S \subseteq \mathbb{R}^n$. If the line segment between any two points in $S$ lies in $S$, i.e

$$\lambda x_1 + (1-\lambda)x_2 \in S,\ \forall x_1, x_2 \in S, \forall \lambda \in [0,1]$$

then $S$ is said to be convex. It can be shown that a set $S \subseteq \mathbb{R}^n$ is convex if and only if for any $x_1, \dots, x_n \in S$, the convex combination $\sum_{i=1}^{n} \lambda_i x_i$ Where $\sum_{i=1}^{n} \lambda_i = 1, \lambda_i \geq 0, i = 1, \dots, n$, belongs to $S$.

**1.6 Definition Convex Functions [7].**

Let $S \subseteq \mathbb{R}^n$ be a nonempty convex set. If $f: S \to \mathbb{R}$ satisfies

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2),\ \forall x_1, x_2 \in S, \forall \lambda \in [0,1]$$

then $f$ is said to be a convex function on $S$. If the above inequality is true as a strict inequality for all $x_1 \neq x_2$ and for all $\lambda \in (0,1)$, then $f$ is called a strictly convex function on $S$. If there is a constant $b > 0$ such that for all $x_1, x_2 \in S$, and for all $\alpha \in [0,1]$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) - \frac{1}{2} b\lambda(1-\lambda)\|x_1 - x_2\|^2$$

then f is called strongly convex f hyp($f = \{(x,\alpha) \mid f(x) \geq \alpha, x \in S, \alpha \in \mathbb{R}\}$.

**Example 1 :** interval $\mathrm{A} = (1,2)$ is convex set

Solution :

$\forall\, x, y \in (1,2), \lambda \in [0,1]$

we need to prove $\lambda \mathrm{x} + (1-\lambda)\mathrm{y} \in (1,2)$

let $\mathrm{x} \in (1,2) \rightarrow 1 < \mathrm{x} < 2\}^{*}\lambda \rightarrow \lambda < \lambda\mathrm{x} < 2\lambda \ldots\ldots\ldots (1)$

let $\mathrm{y} \in (1,2) \rightarrow 1 < \mathrm{y} < 2\}^{*}(1-\lambda) \rightarrow (1-\lambda) <^{*} (1-\lambda)\mathrm{y} < 2^{*}(1-\lambda) \ldots\ldots (2)$

By take the sumof (1) and (2) we get :-

$\lambda + (1-\lambda) < \lambda\mathrm{x} + (1-\lambda)\mathrm{y} < 2\lambda + 2 - 2\lambda \rightarrow 1 < \lambda\mathrm{x} + (1-\lambda)\mathrm{y} < 2$

So , $\lambda\mathrm{x} + (1-\lambda)\mathrm{y} \in (1,2) = \mathrm{A}$

$(1,2)$ is convex

## 1.7 Definition Closed Function [8].

A convex function $f$ is a closed function if its epigraph epi $(f)$ is a closed set in $\mathbb{R}^n \times \mathbb{R}$. Equivalently, a function $f: \mathbb{R}^n \to \mathbb{R}$ is said to be closed if for each $\alpha \in \mathbb{R}$ , the set $\{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$ is closed.

## 1.8 Definition Epigraph [10].

Let $S \subseteq \mathbb{R}^n$ be a non empty convex set.The epigraph of a function $f: S \to \mathbb{R}$, denoted by epi $(f)$, is a subset of $\mathbb{R}^{n+1}$ defined by $\text{epi}(f) = \{(x, \alpha) \mid f(x) \leq \alpha, x \in S, \alpha \in \mathbb{R}\}$ The hypograph of $f$, denoted by hyp $(f)$, is a subset of $\mathbb{R}^{n+1}$ defined by

$$\text{hyp}(f) = \{(x, \alpha) \mid f(x) \geq \alpha, x \in S, \alpha \in \mathbb{R}\}.$$

## 1.9 The Proximal Operator

In this section, we introduce the proximal operator which will be fundamental in many of the algorithms.

**Definition:[13]** The proximal operator $\text{prox}_f: \mathbb{R}^n \to \mathbb{R}^n$ of a convex function $f: \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\text{prox}_f(y) = \underset{x \in \mathbb{R}^n}{\text{argmin}} \left( f(x) + \frac{1}{2} \| x - y \|_2^2 \right)$$

The function $x \mapsto f(x) + \frac{1}{2} \| x - y \|_2^2$ is strongly convex and has a unique minimizer for every
$y \in \mathbb{R}^n$. The proximal operator of $f$ with parameter $\alpha > 0$ is given by

$$\text{prox}_{\alpha f}(y) = \underset{x \in \mathbb{R}^n}{\text{argmin}} \left( f(x) + \frac{1}{2\alpha} \| x - y \|_2^2 \right)$$

From the following properties hold for all $\alpha > 0$.

- The point $x^*$ minimizes $f$ if and only if $\text{prox}_{\alpha f}(x^*) = x^*$.
- If $f(x) = \frac{1}{2} x^T A x + b^T x + c$ is a convex quadratic function, then

$\text{prox}_{\alpha f}(y) = (I + \alpha A)^{-1}(y - \alpha b)$.

- If $f(x) = b^T x + c$ is affine function. Then $\text{prox}_{n+}(y) = (y - \alpha b)$.
- If $f(x) = c$ is constant, then $\text{prox}_{\alpha f}(y) = y$.

If $f: \mathbb{R}^n \to \mathbb{R}$ is a closed convex function, the proximal point algorithm for minimizing $f$

- over $\mathbb{R}^n$ is given by

$$\begin{aligned} x^{k+1} &= \operatorname{prox}_{\alpha^k f}(x^k) \\ &= \operatorname*{argmin}_{x\in\mathbb{R}^n}\left(f(x) + \frac{1}{2\alpha^k}\|x - x^k\|_2^2\right). \end{aligned}$$

## 1. 10 The Linear Programming Problem [9].

The standard LP problem is given by

$$\begin{cases} \text{minimize} & \langle c, x\rangle \\ \text{subject to} & \langle a_i, x\rangle = b_i,\ i = 1, \dots, m \\ & x \geq 0, x \in \mathbb{R}^n \end{cases}$$

where $\mathbb{R}$ is the set of real numbers, $\mathbb{R}^n$ is the set of vectors of length $n$ with real entries, and the problem data is given by the vectors $a_i \in \mathbb{R}^n$, for $i = 1, \dots, m$, the numbers $b_i \in \mathbb{R}$, for $i = 1, \dots, m$, and the vector $c \in \mathbb{R}^n$. The inner product of two vectors $x$ and $y$ in $\mathbb{R}^n$ is defined as let $\mathbb{R}^n_+$ be the set of vectors $x \in \mathbb{R}^n$ such that $x \geq 0$.[8]

## 1.11 Bisection Method [15].

In bisection method we reduce begin with an interval so that $0 \in [a, b]$ and divide the interval in two halves,i.e. $\left[a, \frac{a+b}{2}\right]$ and $\left[\frac{a+b}{2}, b\right]$. A next search interval is chosen by comparing and finding which one has zero. This is done by evaluating the sign. The algorithm for this is given as follows: Choose $a, b$ so that $f(a)f(b) < 0$ [15] .

### 1.11.1 Bisection Algorithm [12].

1 $m = \frac{a+b}{2}$

2 $f(m) = 0$ Stop and return.

3 $f(m)f(a) < 0; b = m$
4 $f(m)f(b) < 0; a = m$
5 $|a - b| < \epsilon$ stop and return.

## Features of bisection

- Guaranteed Convergence.[9]
- Slow as it has linear convergence with $\beta = 0.5$.[11]
- At the $k^{\text{th}}$ iteration $\frac{|b-a|}{2^k} = \epsilon$ giving total evaluations as $k = \log\left(\frac{|b-a|}{2}\right)$ [16] .

**Example 2:** Converging $f(x) = \frac{x}{1+x^2}$ with $[a, b] = [-0.6, 0.75]$ and $\epsilon = 10^{-4}$.[19]

| Number | $a$ | $f(a)$ | $b$ | $f(b)$ | $m$ | $f(m)$ |
|---|---|---|---|---|---|---|
| 1 | −0.6 | −0.441176 | 0.75 | 0.48 | 0.075 | 0.07458 |
| 2 | −0.6 | −0.441176 | 0.075 | 0.07458 | −0.2625 | −0.24557 |
| 3 | −0.2625 | −0.245578 | 0.075 | 0.07458 | −0.09375 | −0.09293 |
| 4 | −0.09375 | −0.092933 | 0.075 | 0.07458 | −0.009375 | −0.00937 |
| 5 | −0.009375 | −0.009374 | 0.075 | 0.07458 | 0.032813 | 0.032777 |
| 6 | −0.009375 | −0.009374 | 0.0328 | 0.032777 | 0.011719 | 0.011717 |
| 7 | −0.009375 | −0.009374 | 0.0117 | 0.011717 | 0.001172 | 0.001172 |
| 8 | −0.009375 | −0.009374 | 0.0011 | 0.001172 | −0.004102 | −0.00410 |
| 9 | −0.004102 | −0.004101 | 0.0011 | 0.001172 | −0.001465 | −0.00146 |
| 10 | −0.001465 | −0.001465 | 0.0011 | 0.001172 | −0.000146 | −0.00014 |
| 11 | −0.000146 | −0.000146 | 0.0011 | 0.001172 | 0.000513 | 0.000513 |
| 12 | −0.000146 | −0.000146 | 0.0005 | 0.000513 | 0.000183 | 0.000183 |
| 13 | −0.000146 | −0.000146 | 0.0001 | 0.00018 | 0.000018 | 0.00001 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 14 | −0.000146 | −0.000146 | 0.0000 | 0.000018 | −0.000064 | −0.00006 |

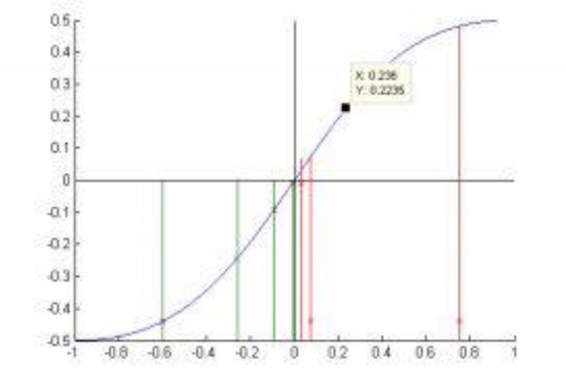



Figure 1: The bisection method

If the function is smooth and continuous, the speed of convergence can be improved by using the derivative information

## 1.12 Newton's Method [14].

In Newton's method does a linear approximation of the function and finding the $x$ -intercept of that approximation, thereby improving the performance of the bisection method. Linear approximation can be done by using Taylor's series.

$$
\begin{aligned}
f'(x_{k+1}) &= f(x_k) + f'(x_k)(x_{k+1} - x_k) \\
f'(x_{k+1}) &= 0 \\
x_{k+1} &= x_k - \frac{f(x_k)}{f'(x_k)}
\end{aligned}
$$

### 1.12.1 Features of Newton's method [13] .

- Newton's method has a quadratic convergence when the chosen point is close enough to zero. If the derivative is zero at the root, it has only
  local quadratic convergence.
- Numerical difficulties occur when the first-derivative is zero.
- If a poor starting point is chosen the method may fail to converge or diverge.

- If it is difficult to find analytical derivation, secant method may be used.
  For a smooth, continuous function when proper starting point is chosen, Newton's method can be real fast. The convergence of
  $f(x) = \frac{x}{1+x^2}$

with $x_0 = 0.5$ and $\epsilon = 10^{-4}$.[5]

| $X$ | $f(x)$ |
|---|---|
| 1 | 0.5 |
| 2 | -0.333333 |
| 3 | 0.083333 |
| 4 | -0.001166 |
| 5 | 0 |

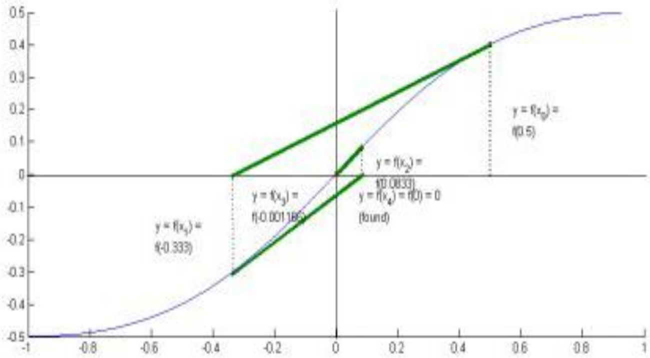



Figure 2: Newton's method converging with $x_0 = 0.5$

| | |
|---|---|
| 1 | 0.75 |
| 2 | $-1.928571$ |
| 3 | $-5.275529$ |
| 4 | $-10.944297$ |
| 5 | $-22.072876$ |
| 6 | $-44.236547$ |

| | |
|---|---|
| ⋮ | ⋮ |
| 20 | −725265.55757 |

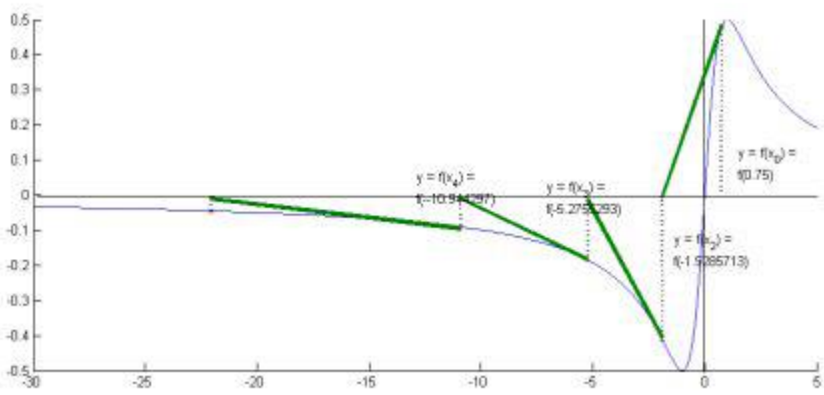



Figure 3: Newton's method diverging with $x_0 = 0.75$

Newton's method to find the minimum uses the first derivative of the approximation an equalling it to zero.

$$\begin{aligned} f'(x_{k+1}) &= f(x_k) + f'(x_k)(x_{k+1} - x_k) \\ f''(x_{k+1}) &= f'(x_k) + f''(x_k)(x_{k+1} - x_k) \\ f''(x_{k+1}) &= 0 \\ x_{k+1} &= x_k - \frac{f'(x_k)}{f''(x_k)} \quad .[13] \end{aligned}$$

## 1.13 Secant Method [20] .

When $f'$ is expensive or cumbersome to calculate, one can use secant's method to approximate the derivative. [5]

The derivation of this method comes by replacing first derivative in the newton's method by its approximation (finite differentiation), i.e

$f'(x_k) = \frac{f_k - f_{k-1}}{x_k - x_{k-1}}$, where $f_k = f(x_k)$

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f_k - f_{k-1}} f_k \quad .[6]$$

Just like Newton's method the secant's method to find the minimum is given by:

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f'_k - f'_{k-1}} f'_k$$

Convergence of secant method is super-linear with $\beta = 1.618$ . The table below shows the secant method convergence for $f = \frac{x}{1+x^2}$ with $x_0 = -0.6$ and $x_1 = 0.75$ as initial points

| Number | $x_{k-1}$ | $x_k$ | $f(x_{k-1})$ | $f(x_k)$ |
|---|---|---|---|---|
| 1 | $-0.6$ | $0.75$ | $-0.441176$ | $0.48$ |
| 2 | $0.75$ | $0.046552$ | $0.48$ | $0.046451$ |
| 3 | $0.046552$ | $-0.028817$ | $0.046451$ | $-0.028793$ |
| 4 | $-0.028817$ | $0.000024$ | $-0.028793$ | $0.000024$ |

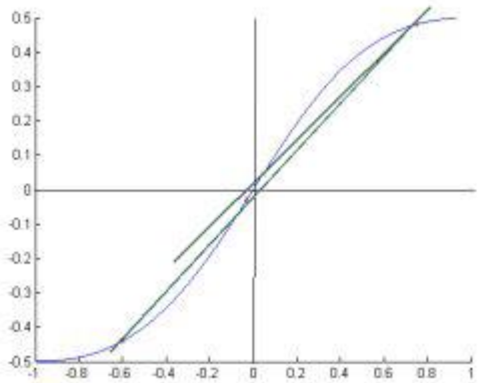



Figure 4: Secant's method converging with $x_0 = -0.6; x_1 = 0.75$

## 1. 14 Golden Section Method [17] .

The table below shows the Golden Section method convergence for $f = \frac{x}{1+x^2}$ with $x_0 = -0.6$ and $x_1 = 0.75$ as initial points

| Number | $a$ | $b$ | $f(a)$ | $f(b)$ |
|---|---|---|---|---|
| 0 | $-0.6$ | $0.75$ | $-0.08375$ | $0.222146$ |
| 1 | $-0.6$ | $0.234346$ | $-0.441176$ | $0.222146$ |
| 2 | $-0.6$ | $-0.084346$ | $-0.441176$ | $-0.08375$ |
| 3 | $-0.6$ | $-0.281308$ | $-0.441176$ | $-0.26068$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 20 | $-0.6$ | $-0.599911$ | $-0.441176$ | $-0.441146$ |

$$x_{\min} = -0.599966 f(x_{\min}) = -0.441165$$

# CHAPTER TWO
# Line Search Technique

## 2.1 Introduction

In mathematics, the conjugate gradient method is an algorithm for the numerical solution of particular systems of linear equations, namely those
whose matrix is positive-definite. The conjugate gradient method is often implemented as an iterative algorithm, applicable to sparse systems that are too large to be handled by a direct implementation or other direct methods such as the Cholesky decomposition. Large sparse systems often
arise when numerically solving partial differential equations or optimization problems**.[6]**

## 2.2 STEP LENGTH [15] .

In computing the step length $\alpha_k$, we face a tradeoff. We would like to choose $\alpha_k$ to give a substantial reduction of $f$, but at the same time we do not want to spend too much time making the choice. The ideal choice would be the global minimizer of the univariate function $\phi(\cdot)$defined by $\phi(\alpha) = f(x_k + \alpha p_k),\ \alpha > 0$ but in general, it is too expensive to identify this value .[8] To find even a local minimizer of $\phi$ to moderate precision generally requires too many evaluations of the objective function $f$ and possibly the gradient $\nabla f$. More practical strategies perform an inexact line search to identify a step length that achieves adequate reductions in $f$ at minimal cost [10] . Typical line search algorithms try out a sequence of candidate values for $\alpha$, stopping to accept
one of these values when certain conditions are satisfied. The line search is done in two stages: A bracketing phase finds an interval containing desirable step lengths, and a bisection or interpolation phase computes a good step length within this interval. Sophisticated line search algorithms can be quite complicated [1] .

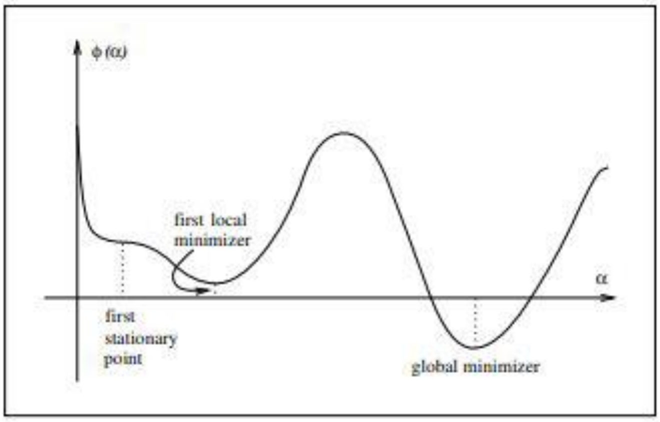



**Figure 3.1** The ideal step length is the global minimizer.

We now discuss various termination conditions for line search algorithms and show that effective step lengths need not lienear minimizers of the univariate function $\phi(\alpha)$ . A simple condition we could impose on $\alpha_k$ is to require a reduction in $f$, that is, $f(x_k + \alpha_k p_k) < f(x_k)$. That this requirement is not enough to produce convergence to $x^*$ is illustrated in Figure 3.2, for which the minimum function value is $f^* = -1$, but a sequence of iterates $\{x_k\}$ for which $f(x_k) = 5/k, k = 0,1, \ldots$ yields a decrease at each iteration but has a limiting function value of zero. The insufficient reduction in $f$ at each step causes it to fail to converge to the minimizer of this convex function. To avoid this behavior we need to enforce a sufficient decrease condition, a concept we discuss next [15] .

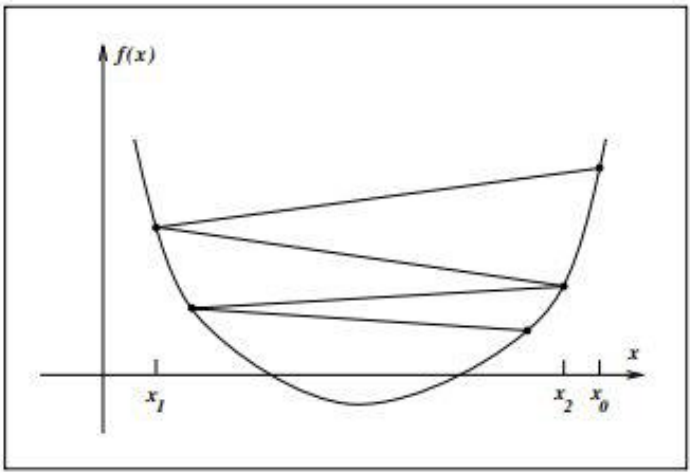



Figure 3.2 Insufficient reduction in $f$.

## 2.3 THE WOLFE CONDITIONS

A popular inexact line search condition stipulates that $\alpha_k$ should first of all give sufficient decrease in the objective function $f$, as measured by the following inequality: $f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f_k^T p_k$
for some constant $c_1 \in (0,1)$. In other words, the reduction in $f$ should be proportional to both the step length $\alpha_k$ and the directional derivative $\nabla f_k^T p_k$. Some times called the Armijo condition [13]. The sufficient decrease condition states that $\alpha$ is acceptable only if $\phi(\alpha) \leq l(\alpha)$. The intervals on which this condition is satisfied are. In practice, $c_1$ is chosen to
be quite small, say $c_1 = 10^{-4}$. To rule out unacceptably short steps we introduce a second requirement, called the curvature condition, which requires $\alpha_k$ to satisfy $\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k$ Note that the left-handside is simply the derivative $\phi'(\alpha_k)$, so the curvature condition ensures that the slope of $\phi$ at $\alpha_k$ is greater than $c_2$ times the initial slope $\phi'(0)$. This makes sense because if the slope $\phi'(\alpha)$ [14].

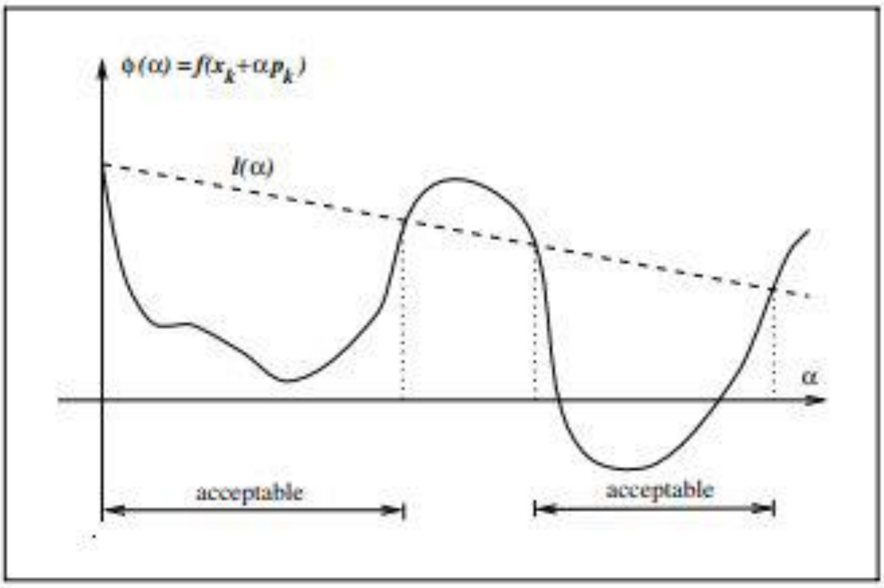



**Figure 3.3** Sufficient decrease condition.

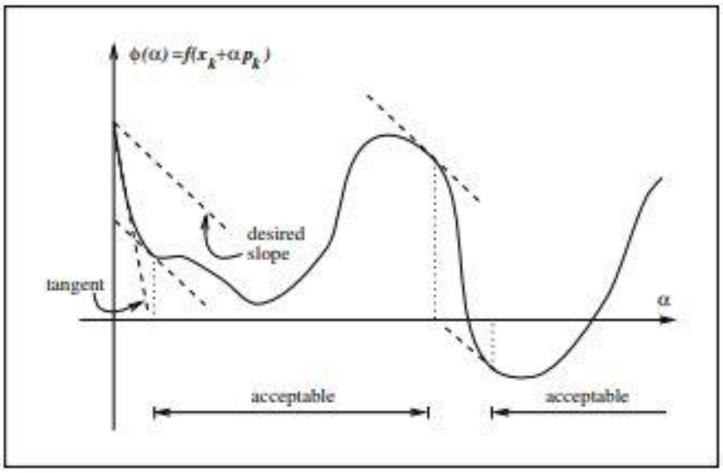



**Figure 3.4** The curvature condition.

is strongly negative, we have an indication that we can reduce $f$ significantly by moving further along the chosen direction [14]. On the other hand, if $\phi'(\alpha_k)$ is only slightly negative or even positive, it is a sign that we cannot expect much more decrease in $f$ in this direction, so it

makes sense to terminate the line search. Typical values of $c_2$ are 0.9 when the search direction $p_k$ is chosen by a Newton or quasi- Newton method, and 0.1 when $p_k$ is obtained from a Nonlinear conjugate gradient method[4]

The sufficient decrease and curvature conditions are known collectively as the Wolfe conditions [12]. We illustrate them in Figure 3.5 and restate them here for future reference:

$$\begin{aligned} f(x_k + \alpha_k p_k) &\leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k \\ \nabla f(x_k + \alpha_k p_k)^T p_k &\geq c_2 \nabla f_k^T p_k \\ \text{with } 0 < c_1 < c_2 < 1 \end{aligned}$$

A step length may satisfy the Wolfe conditions without being particularly close to a minimizer of $\phi$, as we show in Figure 3.5. We can, however, modify the curvature condition to force $\alpha_k$ to lie in at least a broad neighborhood of a local minimizer or stationary point of $\phi$. The strong Wolfe conditions require $\alpha_k$ to satisfy

$$\begin{aligned} f(x_k + \alpha_k p_k) &\leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k \\ |\nabla f(x_k + \alpha_k p_k)^T p_k| &\leq c_2 \left|\nabla f_k^T p_k\right| \end{aligned}$$

with $0 < c_1 < c_2 < 1$. The only difference with the Wolfe conditions is that we no longer allow the derivative $\phi'(\alpha_k)$ to be too positive. Hence, we exclude points that are far from stationary points of $\phi$.

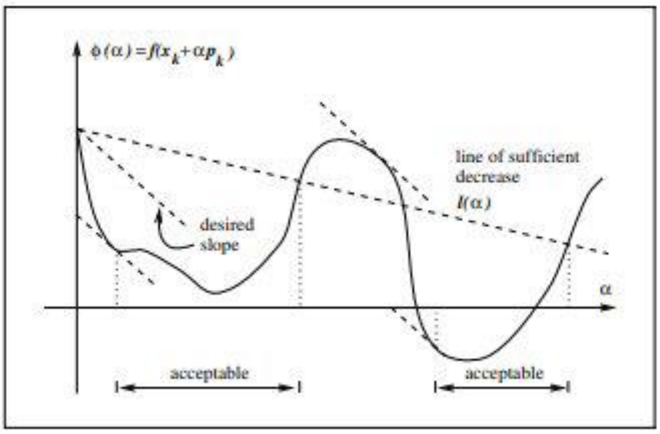



**Figure 3.5** Step lengths satisfying the Wolfe conditions.

It is not difficult to prove that there exist step lengths that satisfy the Wolfe conditions for every function $f$ that is smooth and bounded below [12] .

## 2.4 THE GOLDSTEIN CONDITIONS

Like the Wolfe conditions, the Goldstein conditions ensure that the step length $\alpha$ achieves sufficient decrease but is not too short. The Goldstein conditions can also be stated as a pair of inequalities, in the following way:

$$f(x_k) + (1-c)\alpha_k \nabla f_k^T p_k \le f(x_k + \alpha_k p_k) \le f(x_k) + c\alpha_k \nabla f_k^T p_k$$

with $0 < c < 1/2$. The Goldstein and Wolfe conditions have much in common, and their convergence theories are quite similar. The Goldstein conditions are often used in Newton-type methods but are not well suited for quasi- Newton methods that maintain a positive definite Hessian approximation.[13]

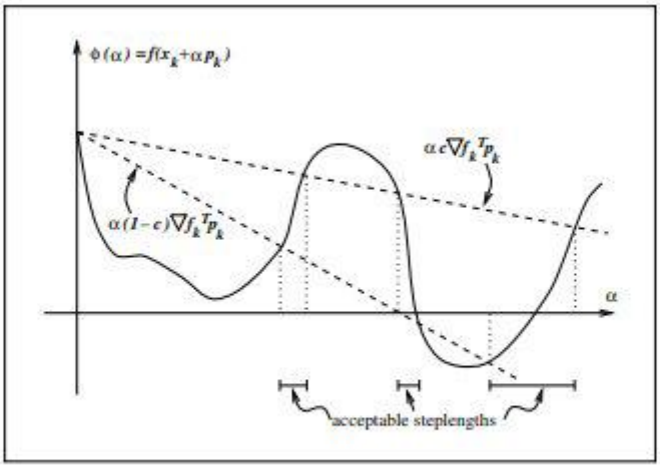



**Figure 3.6** The Goldstein conditions.

# CHAPTER THREE

## Comparison of Steepest Descent Method and Conjugate Gradient Method

## 3.1 Introduction

Computer algorithms are important methods for numerical processing. In all implementations it is important to make them more efficient and decrease complexity but without loss of efficiency. Cryptographic algorithms are one of most important methods for computer science therefore efficiency and complexity of these is crucial for optimal processing [14]. Knowledge systems demand fast and precise methods for information processing [4], [6]. Similarly decision support models need devoted techniques for lower complexity [12]. This publication present comparison of steepest descent method and conjugate gradient method. These methods are used for solving systems of linear equations. In our publication, we analyze, which method is faster and how many iteration required each method. First, we describe these methods, than we compare them and make conclusions [7]. Consider the following unconstrained optimization problem (m) f(x) [11] . Any optimization algorithm starts by an initial point $x^{\circ}$ and performs series of iterations to reach the optimal point [11]. At any iteration the next point is given by the sum of old point and the direction in which to search of a next point multiplied by how far to go in that direction. Thus $x_{k+1} = x_k + \alpha^k d^k$ Where $d^k$, is the search direction and $\alpha^k$ is a positive scalar determining how far to go in that direction, it is called the step length[20]. Additionally, the new point must be such that the function value (the function which we are optimizing) at that point should be less than or equal to the previous point[11]. This is quite obvious because if we are moving in a different direction where the function value is increasing we are not really moving towards the minimum [11] . Thus

$$\mathrm{f}(x_{k+1}) < f(x_k)$$

$$f(x_k + \alpha^k d^k \leq f(x_k)$$

## 3.2 The Steepest Descent Method

The steepest descent method formulated by Stiefel[5]. We will present the mathematical description of the method of steepest descent and we will make implementation in the form of code. Consider the system linear equations in the general form $Ax = B$ [5] . We suppose that matrix $A$ is symmetric and positive definite. With $x^*$ we denote solution of the system , i.e.: $Ax^* = b_x$ [5] . The methods of gradient constructed a sequence of successive approximations to determine the solutions $x^*$ with using formula: $x^{(k+1)} = x^{(k)} + c_k r^{(k)}$

In iterative process this vector can be determined in each iteration of the algorithm as a suitable combination of the previous values, as described is in next formula: $r^{(k+1)} = r^{(k)} - c_k A r^{(k)}$

The coefficients $c_k$ iterations occurring in the equation 3 are determined on the basis of the following equality:

$$\|x^* - x^{(k+1)}\|_B = \inf_{c_k} \|x^* - x^{(k)} - c_k r^{(k)}\|_B$$

In contrast, norm $\|\cdot\|_B$ is defined as $\| x \|_B \stackrel{\text{def}}{=} \sqrt{x^T B x}$

At the same time it fulfills the condition the alternation of the matrix $A$, i.e.: $A \cdot B = b \cdot A$ [3] . It is not difficult to show that searched coefficients $c_k$ are described by the equation:

$$c_k = \frac{\left(r^{(k)}, B\left(x^* - x^{(k)}\right)\right)}{(r^{(k)}, Br^{(k)})}$$

Just for some matrix $B$, you can determine the value of $B\left(x^* - x^{(k)}\right)$. For example, if $B = A^{(p)}, p \in \mathbb{N}$ then there is following equality:

$$A^{(p)}\left(x^* - x^{(k)}\right) = A^{(p-1)}\left(b - Ax^{(k)}\right) = A^{(p-1)} r^{(k)}$$

Also, it can be described in specific cases. For example, for $p = 1$, method is called steepest descent, while $p = 2$ method minimizes the Euclidean norm, and we call it the method of least residuum. where

$$c_k = \frac{\left(r^{(k)}, Ar^{(k)}\right)}{(Ar^{(k)}, Ar^{(k)})}$$

Example. Apply Steepest Descent Method to minimize function $f(x) = 2{x^2}_1 + 2{x^2}_2 + 6x_1 + 13$ with a strating point $x_{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ,the gradient and Hessian of the function $= \nabla f(x) = \begin{bmatrix} 4x_1 + 6 \\ 4x_2 \end{bmatrix}$ and $A = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$

### 3.3 Conjugate Gradient Method

We use conjugate gradient method to solve the system of linear equations given in the form of $Ax = b$ where $A$ is a positive definite matrix with $n \times n$ sizes[19] .As result of operation of this method we obtain a sequence of vectors starting from a vector initial $x_0$ $x_0 \to x_1 \to \cdots \to x_n$ which stops after accurate calculations at most $n$ steps, so we get the required solution in the form of $x_{n+1} = x$. Since the calculations performed occurs rounding error value, so the value of $x_{n+1}$ is genera lly not very accurate result[19]. Therefore, in this method, as in all other iterative methods we conduct further steps $x_k \to x_{k+1}$, respectively, until we get the exact solution. Since the amount of work at each step of the workload of the matrix multiplication $A$ by the vector, and therefore this method is not suitable for use in the matrix band and solid. In contrast, as much as possible, this method is suitable for a medium-sized array diluted. The general idea of our method is based on the fact that the functional $F: \mathrm{R}^n \to \mathrm{R}$ which has a form $F(z) = \frac{1}{2}(b - Az)^T A^{-1}(b - Az) = \frac{1}{2} z^T b^T z - Az + \frac{1}{2} b^T A^{-1} b$ is minimized with accurate solution $x$, thus $0 = F(x) = \min_{z \in \mathbf{R}^n} F(z)$ This equality follows from the fact that the matrix $A$ is positive definite, and thus the matrix $A^{-1}$ also has the same property [20] . While corresponding with the vector $z$ residualny vector $r : Az = -b$, disappears only $z := x$. It reminds us that the method of steepest descent in which a string of $x_1 \to x_2 \to$ ... is found with 1-dimensional minimize the functional F in the direction of the gradient: $x_{k+1}: F(x_{k+1}) = \min_u F(x_k + ur_k)$ where $r_k := DF(x_k)^T = b - Ax_k$ However, in the case of conjugate gradient method, we must to carry out the minimization of the $k$ dimensions:

$$x_{k+1}: F(x_{k+1}) = \min_{u_1,\dots,u_k} F(x_k \quad + u_1 r_1 + \cdots + u_k r_k)$$

$$r_i \quad := b - Ax_i \text{ for } i \leq k$$

It's very easy to compute $x_{k+1}$ [16] . In addition, vectors $r_i$ are ortogonal, and hence lineary independent provided that $r_k \neq 0$. Becouse, in space R, there are no more than $n$ in depended vectors, so in exact calculations, there is the smallest $k \leq n+1$ such that $r_k$ is equal zero. Vector $x_k$ corresponds to number $r_k$. Now we well show step $kx_k \rightarrow r_{k+1}$ of our method. The matrix dimensions $n \times k$ we denote $R_k$, $R_k :=$ $(r_1, r_2, \dots, r_k)$ With the k-dimensional minimilizing point are allowed all vectors from $z$ being in the form of

$z = x_k + R_k u,\ u \in \mathrm{R}^k$ where $r = r_k + AR_k u$ is residual vector. This will be an example of the linear method.

$Ax = b$ is a linear system and it is described by

$$\mathrm{Ax} = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

An initial guess of $x_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is made.

The residual will be found which is computed from the formula $\mathrm{r}_0 = \mathrm{b} - \mathrm{Ax}_0$. so in our case $\mathrm{r}_0 = \begin{bmatrix} -5 \\ -3 \end{bmatrix}$

As this is the first teration the residual vector will be used as the initial search direction. Now $\alpha_0$ is calculated using the equation $\alpha_0 = \frac{\mathbf{r}_0^{\mathrm{T}} \mathbf{r}_0}{\mathbf{p}_0^{\mathrm{T}} \mathbf{A} \mathbf{p}_0} = \frac{[-5 \quad -3]\begin{bmatrix} -5 & -3 \\ & -3 \end{bmatrix}}{[-5 \quad -3]\begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} -5 \\ -3 \end{bmatrix} = \frac{34}{173}}.$ With our first alpha found we can now compute

$\mathrm{x}_1 = \mathrm{x}_0 + \alpha_0 \mathrm{p}_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \frac{34}{173} \begin{bmatrix} -5 \\ -3 \end{bmatrix} = \begin{bmatrix} 0.017 \\ 1.411 \end{bmatrix}.$

With the new x now calculated one must calculate the next residual using the following equation:

$$\mathrm{r}_1 = \mathrm{r}_0 - \alpha_0 \mathrm{Ap}_0 = \begin{bmatrix} -5 \\ -3 \end{bmatrix} - \frac{34}{173} \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -5 \\ -3 \end{bmatrix} = \begin{bmatrix} 0.5 \\ -0.25 \end{bmatrix}$$

With our new residual we now must compute $p_1$ but first we need to find $\beta_0$.

$$\beta_0 = \frac{r_1^T r_1}{r_0^T r_0} = \frac{[.75 \quad -.25]\begin{bmatrix} .75 \\ -0.25 \end{bmatrix}}{[-5 \quad -3]\begin{bmatrix} -5 \\ -3 \end{bmatrix}} = 0.0183$$

With $\beta_0$ now the next search direction $p_1$ can be calculated:

$$p_1 = r_1 + \beta_0 p_0 = \begin{bmatrix} .75 \\ -.25 \end{bmatrix} + 0.0183 \begin{bmatrix} -5 \\ -3 \end{bmatrix} = \begin{bmatrix} 0.6585 \\ -0.3049 \end{bmatrix}$$

Table.1

## 3.4 COMPARISON OF METHODS [18] .

| | | Conjugate gradient method | | | Stecpest descent method | | |
|---|---|---|---|---|---|---|---|
| N | n | Iteration | Time in ms | Time in ti | Iteration | Time in ms | Time in ti |
| 5 | 25 | 13 | 1 | 1983 | 128 | 0 | 1780 |
| 10 | 100 | 33 | 36 | 68281 | 405 | 34 | 64634 |
| 15 | 225 | 49 | 286 | 530934 | 837 | 272 | 505238 |
| 20 | 400 | 65 | 1680 | 3110493 | 1413 | 1621 | 3002271 |
| 25 | 625 | 79 | 6499 | 12032438 | 2093 | 6329 | 11716167 |
| 30 | 900 | 94 | 20053 | 37122809 | 2909 | 19606 | 36295497 |

$$r_{k+1} := r_k + \left(-Ar_k + e_k(r_k - r_{k-1})\right)/q_{k+1}$$
$$x_{k+1} := x_k + \left(-r_k + e_k(x_k - x_{k-1})\right)/q_{k+1}$$
$$e_{k+1} := q_{k+1} \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$$

## 3.5 THE COMPARISON OF METHODS

A. Analysis of efficiency

There is a big difference between presented methods. The steepest descent method is several times faster than conjugate gradient method [4] . However, if we look on iteration of this methods in Table.1 we see, that conjugate method converge after less number of iteration[4]. For example, for $n = 20$, number of iterations of conjugate gradient method equals 65, and achieve the desired accuracy - 1413

## Exact Line Search:

In early days, $\alpha^k$ was picked to minimize

$$\text{(ELS) m } \; f(x_{k+}\alpha d_k) \quad \text{s.t. } \alpha \geq 0$$

Although usable, this method is not considered cost effective.

## Inexact Line Search Methods:

- Formulate a criterion that assures that steps are neither too long nor too short.
- Pick a good initial stepsize.

Construct sequence of updates that satisfy the above criterion after very few steps .

## Backtracking Line Search:

1 Given \alphainit $> 0$ (e.g., $\alpha$ init $= 1$ ), let $\alpha(0) =$ \alphainit and $1 = 0$.

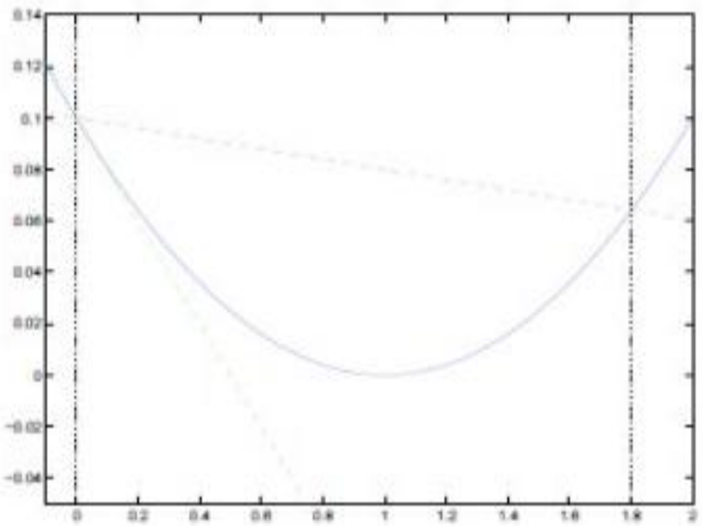
2 Until $f(x_{k+}\alpha d_k) < f_k$

i) set $\alpha^{i+1} = T\alpha^i$ where T $\in (0,1)$ is fixed (e · g · T $= 1/2$)

ii) increment I by 1 .

3 Set $\alpha^k = \alpha^i$

This method prevents the step from getting too small, but it does not prevent steps that are too long relative to the decrease in f. To improve the method, we need to tighten the requirement f($x_{k+}\alpha d^k) < f_k$
To prevent long steps relative to the decrease in f, we require the Armijo condition f($x^k + \alpha^k p^k) < f(x)^k + \alpha^k \beta + [g^k]^t p^k$
for some fixed $\beta \in (0,1)$( e.g., $\beta = 0.1$ or even $\beta = 0.0001$ ). That is to say, we require that the achieved reduction if f be at least a fixed fraction $\beta$ of the reduction promised by the first –oder Taylor approximation of $f$ at $x^k$.[16]




## 3.6 Backtracking-Armijo Line Search:

1 Given $\alpha_{\text{init}} > 0$ (e.g., $\alpha_{\text{init}} = 1$ ), let $\alpha^0 = \alpha_{\text{init}}$ and i $= 0$.
2 Until $\leq f\big(x^k + \alpha^i * p^k \leq f(x^k) + \alpha^i \beta [g^k]^T P^K$

i) set $\alpha^{(I+1)} = T\alpha^i =$, where T $\in (0,1)$ is fixed ( e.g., T $= \frac{1}{2}\Big)$,
ii) increment J by 1 .

3 set $\alpha^k = \alpha^{(J)}$

## 3.7 Computing a Search Direction $d^K$

**Method of Steepest Descent:**

The most straight-forward choice of a search direction, $P^K = - g^k$, is called steepest-descent direction.[3]

- $P^K$ is a descent direction.
- $P^K$ solves the problem

$$\text{m } p \in R^n m_k(x^k + p) = f^k + [g^k]^T P$$
$$\text{s.t. } \| p \|_2 = \|g^k\|_2$$

$p^k$ is cheap to compute. Any method that uses the steepest-descent direction as a search direction is a method of steepest descent. Intuitively, it would seem that $p^k$ is the best search- direction one can find. If that were true then much of optimisation theory would not exist [17] .

## 3.8 The Method of Steepest Descent :

When it is not possible to find the minimium of afunction analytically, and therefore must use an iterative method for obtaining an approximate solution, Newton's Method can be an effective method, but it can also be unreliable [17]. Therefore, we now consider another approach.
Given a function f: $R^n \rightarrow$ R that is differentiable at $x^{\circ}$, the direction of steepest descent is the vector $-\nabla f(x0)$. To see this, consider the function $\varphi(t) = f(x0 + tu)$ where $u$ is a unit vector; that is, $\| u \| = 1$.
Then, by the Chain Rule $\emptyset$ , $= \frac{df}{dx1}\frac{\partial x1}{\partial t} + \cdots + \frac{\partial f}{\partial xn}\frac{dxn}{dt}$

$$= \frac{df}{dx1}u_1 + \cdots + \frac{\partial f}{\partial xn}u_n$$

$$= \nabla f(x0 + tu) \cdot u$$

And therefore

$$\varphi'(0) = \nabla f(x0) \cdot u = \| \nabla f(x0) \| \cos\theta$$

where $\theta$ is the angle between ∇f(x0) and u. It follows that $\varphi'(0)$ is minimized when $\theta = \pi$, which yields

$$u = -\frac{\nabla f(x0)}{\|\nabla f(x0)\|}$$

$$\varphi(0) = -\| \nabla f(x0) \|$$

We can therefore reduce the problem of minimizing a function of several variables to a singlevariable minimization problem, by finding the minimum of $\varphi$(t) for this choice of u [12] .That is, we find the value of $t$, for $t > 0$, that minimizes $\varphi 0(t) = f(x0 - t\nabla f(x0))$

After finding the minimizer to, we can set $x_1 = x_0 - t_0 \nabla f(x_0)$ and continue the process, by searching from $x$ in the direction of $-\nabla f(x_1)$ to obtain $x_2$ by minimizing $\varphi 1(t) = f(x_1 - t\nabla f(x_1)$ , and so on. This is the Method of Steepest Descent: given an initial guess xO, the method computes a sequence of iterates $\{x_k\}$, where $x_{k+1} = x_k - t_k \nabla f(x_k), k = 0,1,2, \ldots$ where tk > 0 minimizes the function $\varphi k(t) = f(xk - t\nabla f(xk))$

**Example 3 :** Apply Steepest Descent Methad to minimize function

$$f(x) = 2x_1^2 + 2x_2^2 + 6x_1 + 13$$

with a strating point $x_{10} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ The gradient and Hessian f th fanltion

$$\nabla f(x) = \begin{bmatrix} 4x_1 + 6 \\ 4x_2 \end{bmatrix}$$

$$\text{and } A = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

The gradient of $f$ at $x_{(0)}$ is

$$\nabla f(x_{(0)}) = \begin{bmatrix} 10 \\ 8 \end{bmatrix}$$

The step length $\alpha_{(0)}$ for quadratic function is computed as

$$= \frac{100+64}{[4032]\begin{bmatrix}10\\3\end{bmatrix}} = \frac{164}{400+256} = 0.25$$

$$x(1) = \left(x_{(0)} - \alpha_0 \nabla f\left(x_{(0)}\right)\right)$$

$$= \begin{bmatrix}1\\2\end{bmatrix} - \frac{1}{4}\begin{bmatrix}10\\8\end{bmatrix} = \begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}2.5\\2\end{bmatrix}$$

$$g_{(1)} = \nabla f\left(x_{(1)}\right) = \begin{bmatrix}4(-1.5)+6\\4(0)\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$$

since the gradient of $f$ at $(x_1$ ) is zero. Thus the steep descent algorithm stop

* The minimium point is $x^* = x_{(1)} = \begin{bmatrix}-1.5\\0\end{bmatrix}$

* The minimium value of the objectin function is $f(x^*) = 2(-1.5)^2 + 2(0) + 6(-1.5) + 13 \ = 8.5$

**Example 4 :** We apply the Method of Steepest Descent to the function

$$f(x, y) = 4x^2 - 4xy + 2y^2$$

with initial point $x0 = (2,3)$. We first compute the steepest descent direction from

$$\nabla f(x, y) = (8x - 4y, 4y - 4x)$$

to obtain

$$\nabla f(x0) = \nabla f(2,3) = (4,4)$$

We then minimize the function

$$\varphi(t) = f((2,3) - t(4,4)) = f(2 - 4t, 3 - 4t)$$

by computing

$$\varphi'(t) = -\nabla f(2 - 4t, 3 - 4t) \cdot (4,4)$$

$$= -\left(8(2 - 4t) - 4(3 - 4t), 4(3 - 4t) - 4(2 - 4t)\right) \cdot (4,4)$$

$$= -(16 - 32t - 12 + 16t, 12 - 16t - 8 + 16t) \cdot (4,4)$$

$$= -(-16t + 4,4) \cdot (4,4)$$

$$= 64\mathrm{t} - 32$$

This strictly convex function has a strict global minimum when $\varphi'(t) = 64t - 32$, or $t = 1/2$, as can be seen by noting that $\varphi''(t) = 64 > 0$.
We therefore set
$x1 = x0 - 1/2\nabla f(x0) = (2,3) - 1/2(4,4) = (0,1)$. Continuing the process, we have

$$\nabla f(x1) = \nabla f(0,1) = (-4,4)$$

and by defining $\varphi(t) = f((0,1) - t(-4,4)) = f(4t, 1 - 4t)$
we obtain

$$\varphi'(t) = -(8(4t) - 4(1 - 4t), 4(1 - 4t) - 4(4t)) \cdot (-4,4) = -(48t - 4, -32t + 4) \cdot (-4,4) = 320^{*}t - 32$$

We have $\varphi'(t) = 0$ when $t = 1/10$, and because $\varphi"(t) = 320$, this critical point is a strict global minimizer. We therefore set

$$x_2 = x_1 - 1/10\nabla f(x_1) = (0,1) - 1/10(-4,4) = \left(\frac{2}{5}, \frac{3}{5}\right)$$

Repeating this process yields $x_3 = (0,2/10)$. We can see that the Method of Steepest Descent produces a sequence of iterates $xk$ that is converging to the strict global minimizer of $f(x, y)$ at $x^{*} = (0,0)$ The following theorems describe some important properties of the Method of Steepest Descent [14] .

B. Analysis of time

On graph of time, we can see that, for example, for $n = 15$ time in ms of conjugate gradient method equals 286 and time in ti of steepst descent method equals 271 . After that, we could see, that when number of $n$ increases there is substantial difference between methods. For example, when $n = 20$, time in ms of conjugate gradient method is equals 1680 , time in ti of achieve the desired accuracy - 1621 iterations [9] .

## V. CONCLUSIONS

We have compared two method implemented to solve systems of linear equations. The steepest descent method is faster method, because it solve equations in less amount of time. Conjugate gradient method is slower, but more productive, because, it converges after less iterations. So, we can see, that one method can be used, when we want to find solution very fast and another can be converge to maximum in less iteration [6] .

## REFERENCES

[1] O. Axelsson, G. Lindskog On the Rate of Convergence of the Freconditioned Conjugate Gradient Method, ,,Numerische Mathematik", 1986, nr 48, s. 499 – 523

[2] Mordecai Avriel Nonlinear Programming: Analysis and Methods, 2003, Dover Publishing, ISBN 0-486-43227-0

[3] D. Polap, M. Woźniak, , C. Napoli, and E. Tramontana, "Real-Time Cloud-based Game Management System via Cuckoo Search Algorithm" International Journal of Electronics and Telecommunications, vol. 61 , No. 4 , pp. 333-338, 2015, DOI: 10.1515/eletel-2015-0043.

[4] P. Artiemjew, Stability of Optimal Parameters for Classifier Based on Simple Granules of Knowledge, Technical Sciences, vol. 14, no. 1, pp. 57-69. UWM Publisher, Ulsztyn 2011.

[5] Knyazev, Andrew V.; Lashuk, Ilya Steepest Descent and Conjugate Gradient Methods with Variable Preconditioning. SIAM Journal on Matrix Analysis and Applications. 29 (4): 1267. doi: 10.1137/060675290

[6] Z. Marszałek, "Novel Recursive Fast Sort Algorithm," in Communications in Computer and Information Science, vol. 639 , pp. 344 – 355,2016, DOI: 10.1007/978-3-319-46254-7_27.

[7] G. Capizzi, F. Bonanno, C. Napoli, "Hybrid neural networks architectures for SOC and voltage prediction of new generation batteries storage", International Conference on Clean Electrical Power (ICCEP), pp. 341344, IEEE, 2015, DOI: 10.1109/ ICCEP. 2011.6036301.

[8] C. Napoli, F. Bonanno, G. Capizzi, "An hybrid neuro-wavelet approach for long-term prediction of solar wind", Proceedings of the International Astronomical Union, vol. 6 , no. 274 , pp. 153-155, IAU, 2015, DOI: 10.1017/S174392131100679X.

[9] Yvan Notay Flexible Conjugate Gradients, 2000, SIAM Journal on Scientific Computing. 22 (4): 1444. doi:10.1137/S1064827599362314
[10] D. Połap, M. Woźniak, , C. Napoli, and E. Tramontana, "Is swarm intelligence able to create mazes?" International Journal of Electronics and Telecommunications, vol. 61, No. 4, pp. $305-310$,2015, DOI: 10.1515/ eletel- $2015-0039$

[11] C. Napoli, G. Pappalardo, E. Tramontana, "An agent-driven semantical identifier using radial basis neural networks and reinforcement learning", $XV$ Workshop Dagli Oggetti agli Agenti (WOA), CEUR-WS, vol. 1260, 2015, DOI: 10.13140/2.1.1446.7843.

[12] D. Połap, M. Woźniak, "Flexible Neural Network Architecture for Handwritten Signatures Recognition" International Journal of Electronics and Telecommunications, vol. 62 , no. 2, pp. $197-202$,2016, DOI: 10.1515/ eletel $-2016-0027$

[13] Jan A. Snymam Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms, 2005, Springer Publishing, ISBN 0-387-24348-8

[14] J. Toldinas, R. Damasevicius, A. Venckauskas, T. Blazauskas, and J. Ceponis, "Energy Consumption of Cryptographic Algorithms in Mobile Devices" ELEKTRONIKA IR ELEKTROTECHNIKA, vol. 20, pp. 158161,2014 .

[15] A. BEN-TAL. AND A. NEMIROVSK1, Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications, MPS-SIAM Series on Optimization, SIAM, 2001.

[16] H. Y. BENSON, A. SEN, D. F. SHANNO, AND R. J. VANDERBEI, Interior-point algorithms, penalty methods and equilibrium problems, Technical Report ORFE-03-02, Operations Research and Financial Engineering, Princeton University, 2003.

[17] S. BENSON AND J. Moré, A limited-memory variable-metric algorithm for bound constrained minimization, Numerical Analysis Report P909-0901, ANL, Argonne, IL, USA, 2001.

[18] D. P. BERTSEKAS, Constrained Optimization and Lagrange Multiplier Methods, Academic Press, New York, 1982.

[19] —, Nonlinear Programming, Athena Scientific, Belmont, MA, second ed., 1999.

[20] M. BERZ, C. BischOE, C. F. CORLISS, AND A. GRIEWANK, eds., Computational Differentiation: Techniques, Applications, and Tools, SIAM Publications, Philadelphia, PA, 1996.

## الخلاصه :


الغرض من هذا البحث هو حل مشاكل التحسين، ومن اهم التقنيات هي تقنية (Line Search Technique). في البداية نحن اختبرنا ( Bisection Method, (Newton's Method, Secant Method, and Golden Section Method,

أيضا استخدمنا ستراتجية (Line Search Technique) أحد النهجين التكراريين الأساسيين للعثور على حد أدنى محلي $x^*$ لدالة الهدف $f : R^n \to R$.

استخدام (line search technique) للبحث عن اتجاه نزول تكون فيه الوظيفة الموضوعية $f$ ثم حساب حجم الخطوة الذي يحدد كيفية تحرك $x$ على طول كذلك مقارنة النتائج بين

(The Steepest Descent Method and The Conjugate Gradient Method).

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل كلية التربية للعلوم الصرفة
قسم الرياضيات

# التحسين العددي: طرق البحث الخطية

بحث مقدم

الى مجلس كلية التربية للعلوم الصرفة/ جامعة بابل
كجزء من متطلبات نيل درجة الدبلوم العالي تربية/ رياضيات

مقدم من قبل الطالب
احمد لفته قاسم لفته

بإشراف الدكتور
احمد صباح الجيلاوي

2021 1433