Academic Year : 2023 | Semester : 3

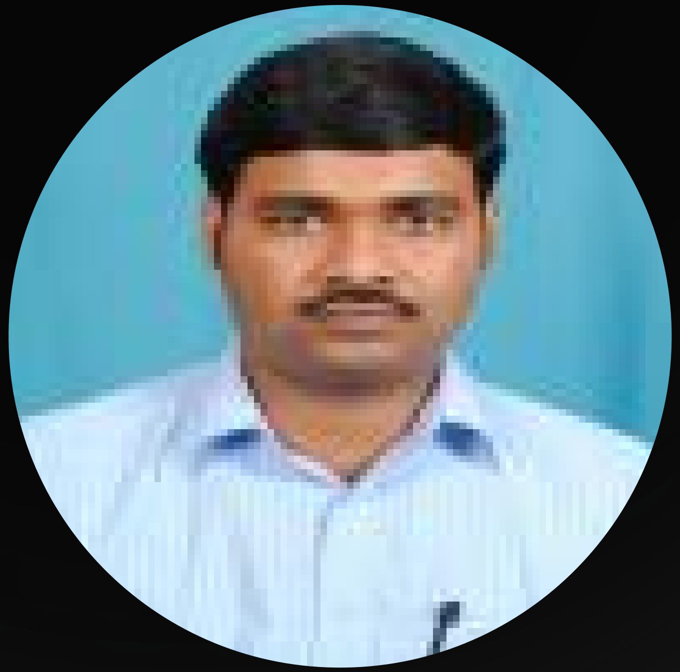
Research paper attached

# Exploratory Project

An adaptive projection BFGS method for nonconvex unconstrained optimization problems


| Name : | Branch : | Roll No. : | Institution : |
|---|---|---|---|
| Aastha Kandwal | Mathematics and Computing | 22124003 | IIT(BHU) Varanasi |

# ACKNOWLEDGEMENT



**Dr. Debdas Ghosh**

I would like to express my sincere appreciation to Dr. Debdas Ghosh for his unwavering support and guidance during my exploratory project on optimization. Throughout the semester, Dr. Ghosh provided invaluable mentorship, and his expertise significantly contributed to the depth and quality of my research.

I am particularly grateful for Dr. Ghosh's generosity in sharing a research paper on the adaptive projection BFGS method for nonconvex unconstrained optimization problems. His insightful discussions and clarifications have been instrumental in enhancing my comprehension of the intricacies involved in this advanced optimization technique.

Dr. Ghosh's commitment to my academic development has truly made a positive impact on my learning journey. I am fortunate to have had the opportunity to work under his guidance, and I sincerely appreciate the time and effort he invested in assisting me with my project. His dedication to fostering a conducive learning environment has inspired and motivated me throughout this endeavor. Thank you, Dr. Ghosh, for your invaluable contributions to my academic growth.

# QUASI-NEWTON METHOD

There are two drawbacks of Newton Methods :

1. Hessian inverse needs to be positive definitive
2. Local convergence ,i.e., initial point needs to be close enough to stationary point

In Quasi-Newton Method, we consider a positive definitive approximation of Hessian Inverse.

So far in solving unconstrained optimization problem, we have see two special methods:

1. Steepest descent method : $d_j = -\nabla f(x_j)$
2. Newton method : $d_j = -H(x_j)^{-1}\nabla f(x_j)$

SD is slower and Newton's method is faster in computing of optimal solution.
SD has global convergence whereas convergence of newton is local ,i.e., convergence can't be guaranteed to any arbitary initial point

Now in order to elevate the drawbacks of Newton's Method we consider the direction of motion :

$$d_j = -B_j \nabla f(x_j)$$

This $d_j$ is known as deflected gradient as if the direction of negative gradient has been deflected a bit by the action of positive action matrix $B_j$.
We will now devise a method that doesn't use the computation of inverse Hessian but constructs an approximation $B_j$ to actual $H(x_j)^{-1}$ whose accuracy gradually improves.
Iterative formula :

$$x_{j+1}=x_j+\alpha_j d_j$$

$$x_{j+1}=x_j+\alpha_j(-B_j \nabla f(x_j))$$

The generated sequence $\{x_j\}$ will hve descent poperty

# QUASI-NEWTON METHOD

Quasi-Newton will have fast convergence and local convergence like Newton Method.

The conditions for approximation of Hessian ,i.e., $B_j \approx H(x_j)^{-1}$ are :

1. $B_j$ is to satisfy Quasi Newton equation : $B_j y_{j-1} = s_{j-1}$
2. Curvature condition : $y_j(s_j)^T > 0$
3. $B_j$ is symmetric
4. Concstruct $B_j$ with the help of $B_{j-1}$ and $\nabla f(x_j)$

# BFGS METHOD

Algorithm :

(Quasi-Newton method with BFGS update)

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

Initialize :

Provide f(x0

$x^0$ -> initial point

$B^0$-> a symmetric and positive definite matrix

$\epsilon > 0$ -> tolerance level/termination scalar

k <-0 ( iteration number)

Main steps :

1. While $||g^k|| > \epsilon$

$d^k = -B^k g^k$

$\alpha^k = \text{argmin } f(x^k+\alpha d^k) \quad , \alpha > 0$

$x^{k+1} = x^k+\alpha^k d^k$

$s^k = x^{k+1}-x^k$

$y^k = g^{k+1}-g^k$

# BFGS METHOD

$$B^{k+1} = \left(I_n - \frac{s^k(y^k)^T}{(s^k)^T y^k}\right) B^k \left(I_n - \frac{y^k(s^k)^T}{(s^k)^T y^k}\right) + \frac{s^k(s^k)^T}{(s^k)^T y^k}$$

2. Output $x^k$ as a stationary point of f.

# RESEARCH PAPER

# ABSTRACT

## Existing Algorithm

The BFGS (Broyden-Fletcher-Goldfarb-Shanno) is a widely used and effective approach for addressing unconstrained optimization problems within quasi-Newton algorithms. This iterative optimization algorithm belongs to the family of quasi-Newton methods, renowned for their ability to optimize smooth, unconstrained objective functions by approximating the inverse Hessian matrix. The BFGS method, in particular, exhibits robust convergence properties and is widely embraced in various fields for its effectiveness in navigating complex, high-dimensional optimization landscapes.

## Shortcomings of Algorithm

While the BFGS algorithm is generally effective, it may fail for nonconvex problems under Wolfe conditions. Wolfe conditions are criteria used in line search algorithms to ensure a proper decrease in the objective function and gradient.

## Proposed Algorithm

An adaptive projection BFGS algorithm is proposed naturally which can solve nonconvex problems, and the following properties are shown in this algorithm:

1. The generation of the step size $\alpha_j$ satisfies the popular Wolfe condition
2. A specific condition is proposed which has sufficient descent property, and if the current point satisfies this condition, the ordinary BFGS iteration process proceeds as usual
3. Otherwise, the next iteration point $x_{j+1}$ is generated by the proposed adaptive projection method.

# INTRODUCTION

Consider the following model of optimization problem :

$$\min\{\varphi(x) \mid x \in R^n\} \tag{1.1}$$

where $\varphi : R^n \to R$ and $\varphi$ is continuously differentiable

There are many methods to solve problem (1.1) like conjugate gradient method and quasi-Newton method.

Note that : 1. $\varphi_j$ denotes the function value $\varphi(x_j)$

2. $g_j$ denotes the function value $g(x_j)$

The BFGS algorithm for solving the problem (1.1) is :

$$x_{j+1} = x_j + \alpha_j d_j \quad , \quad j \in N$$

$\alpha_j$ is the step size generated by inexact line search mechanism

$d_j$ is the search direction of current point $x_j$ , it is calculated by solving this equation : $B_j d_j = -g_j$

$B_j$ is an approximation to the Hessian matrix $\nabla^2 \varphi(x_j)$ at $x_j$

The most widely used quasi-Newton update formula, the Broyden-Fletcher-Goldfarb Shanno (BFGS) method , updates Bj via

$$B_{j+1} = B_j - \frac{B_j s_j s_j^T B_j}{s_j^T B_j s_j} + \frac{y_j y_j^T}{s_j^T y_j}$$

here $s_j = x_{j+1} - x_j$ and $y_j = g_{j+1} - g_j$

BFGS algorithm converges globally for convex problems under Weak Wolfe Powell (WWP) conditions.

So, two unanswered question about quasi-Newton method under the WWP conditions

1. Whether the Davidon-Fletcher-Powell (DFP) algorithm can converge globally for convex problems
2. Whether the Broyden-Fletcher-Goldfarb Shanno (BFGS) algorithm can converge globally for non-convex problems.

# MOTIVATION

The following coclusion can easily be deduced from strongly convex fuction :

$$S_j^T y_j \geq p||s_j||^2 > 0 \qquad (2.1)$$

where p>0 is a cofficient related to the strong convexity of the function

This inequality is difficult to satisfy for non-convex problems. Therefore, in order to obtain it, modified BFGS formulas :

$$B_{j+1} = \begin{cases} B_j - \dfrac{B_j s_j s_j^T B_j}{s_j^T B_j s_j} + \dfrac{y_j y_j^T}{y_j^T s_j} & \text{, if } \dfrac{y_j s_j}{||s_j||^2} > C_1 ||g_j||^{c_2} \\ B_j & \text{, otherwise} \end{cases}$$

where $C_1 > 0$ and $C_2 > 0$ are constants

Besides that, under certain conditions, the sufficient descent property can also get (2.1). Based Based on this, Yuan proposed a projection technique under a modified weak-Wolfe-Powell conditions. Their idea is to project the point that does not satisfy the sufficient descent condition onto a paraboloid by projection method, so as to establish the convergence of the algorithm defined as :

$$\{ x \in R^n \mid \boldsymbol{\lambda} \, || W_j - x ||^2 + ( W_j - x)^T g(W_j) = 0\} \ ,$$

$$\text{where } \boldsymbol{\lambda} > 0$$

The next point created by the projection technique is shown below :

$$x_{j+1} = x_j + \frac{\boldsymbol{\lambda} ||W_j - x_j||^2 + (W_j - x_j)^T g(W_j)}{||g(W_j) - g(x_j)||^2} [g(W_j) - g(x_j)]$$

# MOTIVATION

These observations motivate us to invent an adaptive projection technique to over come situations that do not satisfy the condition we specified, so that the modified BFGS method for nonconvex functions can converge globally under the general WWP line search condition, which can also update different projection formulas adaptively according to different problems.

Correspondingly, we define an index set that satisfy the special sufficient descent condition :

$$S\,D := \{j \mid d_j^T g_j \leq -p\alpha_j \|d_j\|^2 \|g_j\|^\alpha ,\ j \geq 0\},$$

where p is a positive constant , $\alpha \in (-\infty, +\infty)$ is a tuning parameter and $\|g_j\|^\alpha$ can be viewed as an adaptive term.

This definition also helps to explore the convergence rate of the algorithm. For points that do not satisfy the sufficient descent condition, we will use the projection method to get a new direction, so the definition of the adaptive projection surface needs to be given first:

$$\{x \in R^n \mid \mu \|V_j - x\|^2 \|g(x)\|^\alpha + (V_j - x)^T g(V_j) = 0\}, \qquad (2.2)$$

where $V_j = x_j + \alpha_j d_j$ and parameter $\mu > 0$

For the current point $x_j$ , the generation of the next point $x_{j+1}$ will be divided into two cases :

Case (A) : $j \in S\,D$. Iteration proceeds as usual

$$x_{j+1} = V_j$$

Case (B) : $j \notin S\,D$. First, we project the current point $x_j$ onto the surface (2.2), then the next iteration $x_{j+1}$ is defined below:

$$x_{j+1} = x_j + \frac{P_j}{\|g(V_j) - g(x_j)\|^2}[g(V_j) - g(x_j)] \qquad (2.3)$$

where $P_j = \mu \|V_j - x_j\|^2 \|g(x_j)\|^\alpha + (V_j - x_j)^T g(V_j)$

# ALGORITHM

Adaptive projection technique-BFGS algorithm for nonconvex unconstrained optimization problems (APT-BFGS)

**STEP 1**

Choose an initial point $x_0 \in R_n$, an initial symmetric and positive definite matrix $B_0 \in R_{nxn}$

**STEP 2**

Given the necessary parameters
$\rho > 0$,
$\zeta_1 \in (0, 1/2)$,
$\epsilon \in (0, 1/2)$,
$\zeta_2 \in (\zeta_1, 1)$,
$\mu \in (\zeta_2 \rho, +\infty)$,
$\alpha \in (-\infty, +\infty)$,
and let $j = 0$.

**STEP 3**

Calculate $\| g_j \|$,
if $\| g_j \| \leq \epsilon$ is true,
the algorithm stops

**STEP 4**

Get the direction $d_j$ by solving the linear equation :

$B_j d_j = -g_j$

**STEP 5**

Determine a step size $\alpha_j > 0$ that satisfies the WWP condition.

**STEP 6**

Let $V_j = x_j + \alpha_j d_j$

**STEP 7**

If $j \in S$ D.
Let $x_{j+1} = V_j$ ,
$s_j=x_{j+1}-x_j$ and $y_j=g(x_{j+1})-g(x_j)$ ,
then skip to Step 9.

**STEP 8**

Else $j \notin S$ D.
Define the next iteration $x_{j+1}$ as shown in (2.3),
$s_j=x_{j+1}-x_j$ and $y_j=g(x_{j+1})-g(x_j)$

**STEP 9**

Determine $B_{j+1}$ by the ordinary BFGS formula

$$B_{j+1}=B_j - \frac{B_j s_j s_j^T B_j}{s_j^T B_j s_j} + \frac{y_j y_j^T}{s_j^T y_j}$$

**STEP 10**

Let $j = j + 1$ and,
return to Step 3

# REFERENCES

Research papers :

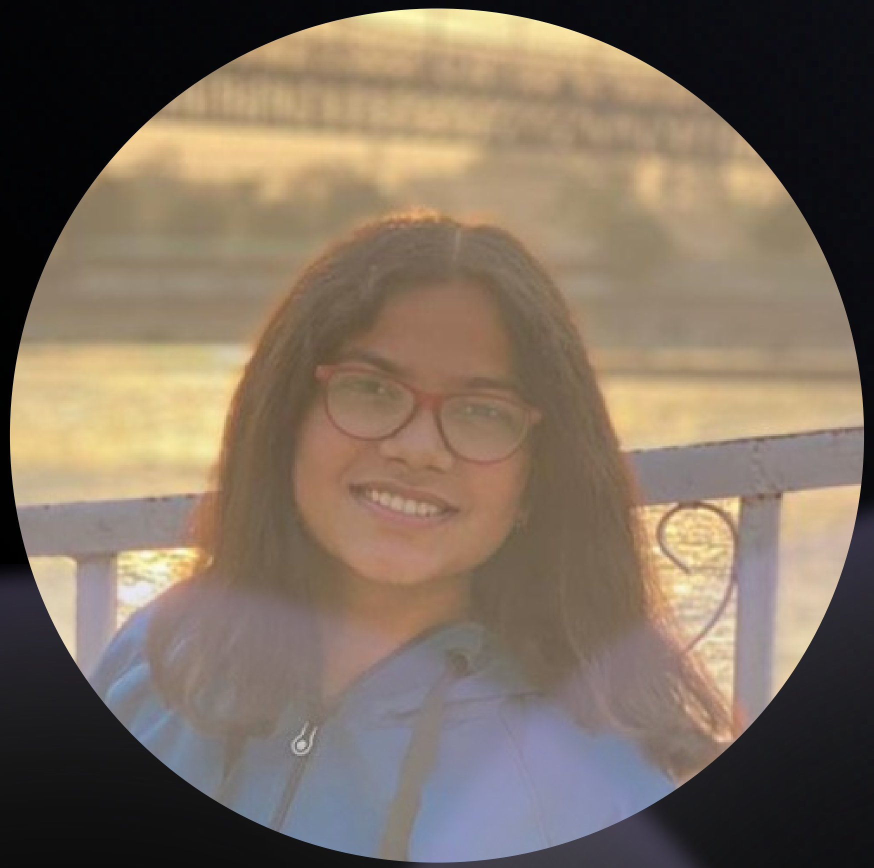

1. Gonglin Yuan, Xiong Zhao, Kejun Liu & Xiaoxuan Chen :An adaptive projection BFGS method for nonconvex unconstrained optimization problems (2023)
2. Andrei, N.: An unconstrained optimization test functions collection. Adv. Model. Optim. 10(1), 147- 161 (2008)
3. Li, D., Fukushima, M.: On the global convergence of the BFGS method for nonconvex unconstrained optimization problems. SIAM J Optim. 11(4), 1054-1064 (2001)
4. Yuan, G., Li, P., Lu, J. The global convergence of the BFGS method with a modified WWP line search for nonconvex functions. Numerical Algorithms, pp. 1-13 (2022)
5. Yuan, G., Sheng, Z., Wang, B., Hu, W., Li, C.: The global convergence of a modified BFGS method for nonconvex functions. J Comput. Appl. Math. 327, 274-294 (2018)


Video References:


1. Lecture 33,34 & 35 Quasi Newton Method-IIT Kanpur NPTEL
2. Lecture 22 Inexact Line Search-IIT Kanpur NPTEL

# T.H.A.N.K Y.O.U

**Submitted By**



**Aastha Kandwal**