Contemporary Engineering Sciences, Vol. 10, 2017, no. 35, 1719 - 1726
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ces.2017.711189

Comparison of the Conjugate Gradient Methods of Liu-Storey and Dai-Yuan

Fernando Mesa

Department of Mathematics and GEDNOL
Universidad Tecnológica de Pereira
Pereira, Colombia

Pedro Pablo Cárdenas Alzate

Department of Mathematics and GEDNOL
Universidad Tecnológica de Pereira
Pereira, Colombia

Carlos Alberto Rodríguez Varela

Department of Mathematics
Universidad Tecnológica de Pereira
Pereira, Colombia

Abstract

The purpose of this paper is to present the capabilities of the conjugate gradient methods based on the theoretical analysis of the gradient method, the precursor of the descent methods. It indicates the geometric differences of these and the improvements made in the search for the optimal value of an objective function. Different test systems are proposed to solve, in order to obtain a solution that can determine the speed of convergence of the conjugate address proposed by Liu-Storey and Dai-Yuan [1].

Keywords: Gradient, descent, optimization, conjugate direction, iteration, solution, minimization.

1 Introduction

Optimization is an area of applied mathematics that allows modeling and solving real-life problems; its principles and methods are used to solve quantitative problems in disciplines such as physics, biology, engineering and economics. The main objective of optimization is the best use of available resources to accomplish a certain task [2]. It includes the study of optimality criteria for problems, the determination of algorithmic methods of solution, the study of the structure of such methods, and computer experimentation with methods both in test conditions and in real-life problems.

Within the algorithms of unrestricted optimization is the gradient method (or the most pronounced descent method), which seeks to minimize quadratic objective functions, from geometrically descendant search directions, this method is of theoretical interest and has been the pillar for the construction of the methods of the descent as it is the method of the conjugate gradient, which possesses a high convergence and is used for the minimization of objective functions with many associated dimensions.

2 Quadratic forms

The squared function is defined as the scalar functions defined on a vector space of dimension n in the following way:

$$f(\bar{x}) = \tfrac{1}{2}\left(a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + \cdots + a_{nn}x_n^2\right) - (b_1x_1 + b_2x_2 + \cdots + b_nx_n) + c$$

$$f(\bar{x}) = \bar{x}^T \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \bar{x} - [b_1\ b_2\ \cdots\ b_n]^T \bar{x} + c$$

$$f(\bar{x}) = \tfrac{1}{2}\bar{x}^T A\bar{x} - \bar{b}^T\bar{x} + c, \tag{1}$$

where c is a constant value and $\bar{x} = [x_1\ x_2 \cdots x_n]^T$.

We have that the eigenvalues of the square symmetric matrix A, hessian of f, define the location of the optimal point and the classification of the quadratic function, that is: Let λ_i (with $i = 1, 2, \ldots, n$) be the eigenvalues of the matrix A associated with the quadratic function. The classification of a quadratic form and its optimal is:

i) Positive definite quadratic form if $\lambda_i > 0$ with $i = 1, \ldots, n$. It has a global minimum.

ii) Negative definite quadratic form if $\lambda_i < 0$ with $i = 1, \ldots, n$. It has a global maximum.

iii) Positive semi-definite quadratic form if $\lambda_i \geq 0$ with $i = 1, \ldots, n$. It has infinite minimum points.

iv) Positive semi-definite quadratic form if $\lambda_i \leq 0$ with $i = 1, \ldots, n$. It has infinite maximum points.

v) Quadratic form indefinite if $\exists i, j : \lambda_i > 0$. It has a saddle point.

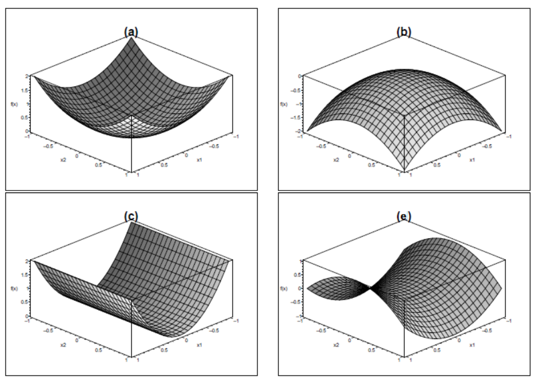

Figure 1: Graphic definition of a quadratic form in $\mathbb{R}^3$.

3 Gradient method

The gradient method is a descent method in which you begin to iterate at an arbitrary point and continue following the line of maximum descent, obtaining a succession of points until you get a point close enough to the solution [3]. This method is used to solve optimization problems without restrictions of type:

$$\min_{x \in \mathbb{R}^n} f(\bar{x}), \tag{2}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a quadratic, continuous and differentiable function, its associated matrix A is positive definite. The method starts from an initial position $\bar{x}^0$ generating a sequence of points $\bar{x}^k$ according to the equation:

$$\bar{x}^{k+1} = \bar{x}^k - \alpha_k \bar{g}^k, \tag{3}$$

where α_k is the descent parameter, which indicates the step length and is obtained from the minimization:

$$\begin{aligned} &f(\bar{x}^k - \alpha_k \bar{g}^k) = \min_{\alpha \in \mathbb{R}} f(\bar{x}^k - \alpha_k c) \\ &\frac{d}{d\alpha} f(\bar{x}^k - \alpha_k \bar{g}^k) = 0 \\ &\alpha_k = \frac{(\bar{b} - A\bar{x}_k)^T \bar{g}^k}{(\bar{g}^k)^T \bar{g}^k}, \end{aligned} \tag{4}$$

and $\bar{g}^k$ is the gradient, direction of maximum descent in $\bar{x}^k$ of the quadratic function f

$$\bar{g}^k = \nabla f(\bar{x}_k) = -(\bar{b} - A\bar{x}_k) \tag{5}$$

The algorithm of the method is as follows:

i) Enter the quadratic function $f(\bar{x})$.

ii) Consider a point $\bar{x}^0$. Do $k = 0$.

iii) Choose the direction of maximum descent (gradient):

$$\bar{g}^k = \nabla f(\bar{x}^k)$$

iv) Calculate the descent parameter:

$$\alpha_k = \frac{(\bar{b} - A\bar{x}_k)^T \bar{g}^k}{(\bar{g}^k)^T \bar{g}^k}$$

v) Do $\bar{x}^{k+1} = \bar{x}^k - \alpha_k \bar{g}^k$

vi) Check convergence. If $\rightarrow ||\bar{g}^k|| < \epsilon$ the method is stopped and $\bar{x}^k$ it is the solution. Otherwise, do $k = k + 1$ and repeat from 4.

4 Conjugate Gradient Method

The conjugate gradient method is a particular case of descent method, the latter is especially indicated for the resolution of dispersed systems (linear systems whose coefficient matrix has a significant number of zeros), such systems frequently arise when the equations are solved numerically in partial derivatives. This method in general manages to save memory and operations by operating only on non-zero elements. The basic idea behind the conjugate gradient method is to construct a base of orthogonal vectors and use it to search the solution more efficiently. Such a procedure generally would not be advisable because the construction of an orthogonal base using the Gramm-Schmidt procedure requires, in selecting each new element of the base, to ensure its orthogonality with respect to each of the vectors previously constructed. The great advantage of the conjugate gradient method is that when using this procedure, it is enough to ensure the orthogonality of a new member with respect to the last one that has been built, so that this condition is automatically fulfilled with respect to all the previous ones [4].

The conjugate gradient method can also be used to solve unrestricted optimization problems such as energy minimization, among others, in addition to exceeding the most pronounced descent method, as can be seen in Figure 2.

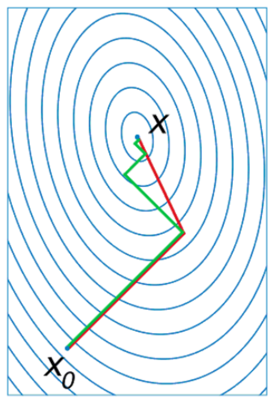

Figure 2: Comparison of descent directions, steepest descent method (green), conjugate gradient (red).

5 Conjugated Gradient Liu-Storey

The conjugate gradient method is a useful and powerful approach to solve large scale minimization problems. Liu and Storey developed a conjugate gradient method, which has good numerical performance but not a result of global convergence in traditional line searches such as Armijo, Wolfe and Goldstein line searches. The algorithm of the present method is the same as the one proposed above, highlighting only the conjugate gradient parameter formulated in (6) containing the parameters of conjugate direction, gradient of the quadratic function, present and previous and the difference of these parameters previously proposed in the algorithm of the conjugate gradient [5].

$$\beta_k = -\frac{(\bar{y}^k)^T \bar{g}^{k+1}}{(\bar{g}^k)^T \bar{d}^k} \tag{6}$$

6 Conjugated Gradient Dai-Yuan

This method produces a descent search direction in each iteration and converges globally whenever the line search satisfies Wolfe's inaccurate conditions. Dai and Yuan performed numerical experiments for two combinations of the new method and the Hestenes-Stiefel conjugate gradient method, obtaining similar convergence capabilities. The algorithm of the present method is equal to the one proposed above, highlighting only the conjugate gradient parameter formulated in (7) containing the gradient of the quadratic function, the difference of gradients and the difference of positions, parameters previously proposed in the conjugate gradient algorithm [6].

$$\beta_k = -\frac{(\bar{g}^{k+1})^T \bar{g}^{k+1}}{(\bar{y}^k)^T \bar{s}^k} \tag{7}$$

7 Test Systems

To check the effectiveness of the method, 4 systems of 2 and 3 variables are selected, each of them described by their associated matrices A and b according to the quadratic form proposed in equation (1). For the stop criterion, $\|\bar{g}^k\| < \epsilon = 10^{-3}$ is used and as starting point $\bar{x}^0$, which is a column vector with n rows and all its components equal to 1. Now, the proposed system for $n = 2$ is

$$A = \begin{pmatrix} 1 & -2 \\ -2 & 2 \end{pmatrix}, \qquad B = \begin{pmatrix} 4 \\ 8 \end{pmatrix}$$

Therefore, the results ($n = 2$) obtained are $X = [X1 \quad X2]$

$$x1 = [1.0000 \quad 1.0000; \quad -62.5714 \quad -100.7143; \quad -12.0000 \quad -8.0000];$$

$$d = 1.0e + 03 * [0.0050 \quad 0.0080; \quad 1.2860 \quad 2.3576; \quad 0.0000 \quad -0.0000;];$$

Gradient

$$g = [5.0000 \quad 8.0000; \quad -134.8571 \quad 84.2857; \quad 0.0000 \quad -0.0000];$$

Constants

$$\text{alpha1} = [-12.7143; \quad 0.0393;];$$

$$\text{beta1} = [-22.3499; \quad 0.0000;];$$

$$s = [-63.5714 \quad -101.7143; \quad 50.5714 \quad 92.7143;];$$

The graphs of figures 3 and 4 show the descent directions of the two optimization methods for spaces in 2 and 3 dimensions respectively. In the graphs it can be seen that the method proposed by Liu-Storey calculates more directions and points of solution in comparison with the method proposed by Dai-Yuan. In figures 4 and 5, the norm $\mathbb{R}^n$ was plotted for being $n = 2$ and $n = 3$ dimensions respectively, and it can be concluded that the Dai-Yuan method converges rapidly compared to the method proposed by Liu-Storey.

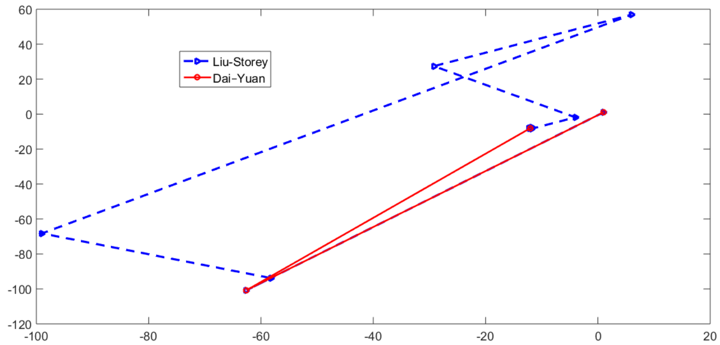

Figure 3: Test system for $n = 2$. Directions Red Dai-Yuan, Blue Liu-Storey.

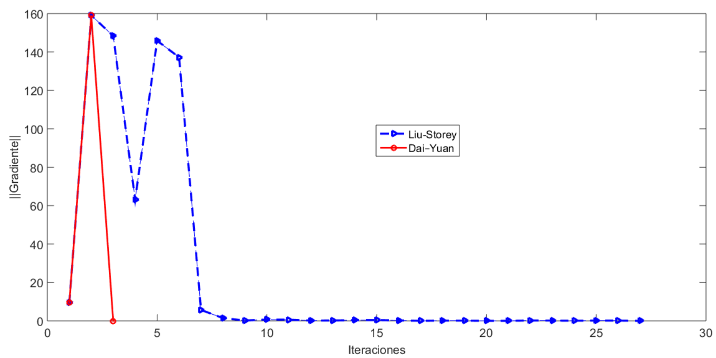

Figure 4: Gradient norm vs. Iterations. Red Dai-Yuan, Blue Liu-Storey.

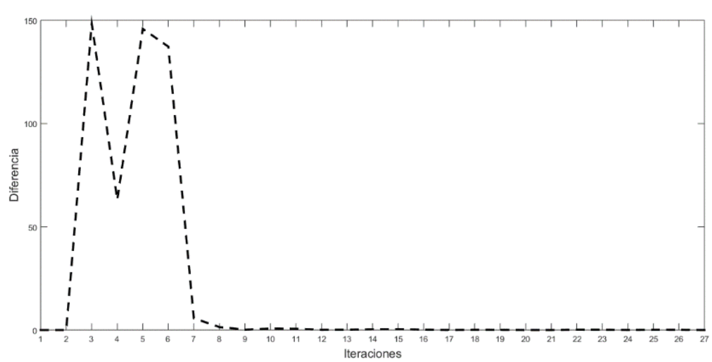

Figure 5: Gradient norm vs. Iterations: Difference

8 Conclusion

Due to its use of conjugate search addresses the Liu-Storey method should converge on N interactions or less for the case of a poorly conditioned quadratic function (those whose contours are highly eccentric or distorted) the method may need more interactions to converge . The reason for this is the cumulative effect of rounding errors. To avoid these problems, it is recommended to reinitialize the method periodically after a certain number of steps, taking the steep descent direction as the new search direction. Despite their limitations, the methods proposed by Dai-Yuan and Liu-Storey are far superior to the steepest descent method and to pattern search methods.

Acknowledgements. We would like to thank the referee for his valuable suggestions that improved the presentation of this paper and our gratitude to the Department of Mathematics of the Universidad Tecnológica de Pereira (Colombia) and the group GEDNOL.

References

[1] N. Andrei, *Conjugate Gradient Algorithms for Unconstrained Optimization*, ICI Technical Report, Vol. 13, 2008.

[2] X. Dupre, Apprentissage d'un réseau de neurones, *Machine Learning, Statistiques et Programmation*, Vol. 14, 2017.

[3] M. Fontelos, Fundamentos matemáticos de la Ingeniería, *Librería-Editorial Dykinson*, Vol. 01, 2007.

[4] C. Cardona, J. Henao, Predicción de los precios de contratos de electricidad usando una red neuronal con arquitectura dinámica, *Innovar*, **20** (2010).

[5] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients, *The Computer Journal,* **7** (1964), 149-154. https://doi.org/10.1093/comjnl/7.2.149

[6] A. Doucet, N. Gordon, An introduction to sequential monte carlo methods, Chapter in *Sequential Monte Carlo Methods in Practice,* Springer, 2001, 3-14. https://doi.org/10.1007/978-1-4757-3437-9_1

Received: December 4, 2017; Published: December 24, 2017