

Comparison descent directions for Conjugate Gradient Method

Comparación de direcciones de descenso para el método de gradiente conjugado

F. Mesa  ; D. M. Devia- Narváez  ; G. Correa-Vélez 
 DOI: <https://doi.org/10.22517/23447214.24893>
 Artículo de investigación científica y tecnológica

Abstract— In the following manuscript we will show as a starting point a theoretical analysis of the gradient method, known as one of the first descent methods, and from this we will identify the strength of the conjugate gradient methods. Taking an objective function, we will determine the values that optimize it by means of different methods, indicating the differences of geometric type that these have. Different systems will be used, in order to serve as a test, obtaining their solution in each case and finding the speed at which they converge in accordance with the conjugate gradient methods proposed by Hestenes-Stiefel and Fletcher-Reeves.

Index Terms— Conjugate direction, descent, gradient, iteration, minimization, optimization, quadratic function, solution.

Resumen— En el siguiente manuscrito mostraremos como punto de inicio un análisis teórico del método de gradiente, conocido como unos de los primeros métodos de descenso, y a partir de ello identificar la fortaleza de los métodos del gradiente conjugado. Tomando una función objetivo determinaremos los valores que la optimizan mediante diferentes métodos indicando las diferencias de tipo geométrico que estos tengan. Se usarán distintos sistemas, con el fin de que sirvan de prueba obteniendo en cada caso su solución y encontrando la velocidad en que convergen de conformidad con los métodos de gradiente conjugado propuestos por Hestenes-Stiefel y Fletcher-Reeves.

Palabras claves— Descenso, dirección conjugada, función cuadrática, gradiente, iteración, minimización, optimización, solución.

I. INTRODUCTION

OPTIMIZATION is one of the most important tools in applied mathematics that is used in solving real life problems in different disciplines such as engineering and

Biology. Getting the resources available to solve a problem or perform a specific task are used in the best way, is without a doubt the number one objective of optimization. For this, we have unrestricted optimization algorithms, within which we can count on the method of the steepest descent or also called the gradient method, which allows us to optimize a quadratic function, which will be the objective function, using different directions of search that are descending (geometrically). The gradient method is of great importance since its speed of convergence is quite high, together with the possibility of solving problems whose objective functions have associated a large number of dimensions [1].

II. CONTENT

A. Quadratic forms

A quadratic form F can be defined as a scalar map whose domain corresponds to a finite vector space of dimension n , represented by (1):

$$F(\bar{x}) = \frac{1}{2} (a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + \dots + a_{nn}x_n^2) - (b_1x_1 + b_2x_2 + \dots + b_nx_n) + c$$

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

$$F(\bar{x}) = \frac{1}{2} \bar{x}^T A \bar{x} - \bar{b}^T \bar{x} + c \quad (1)$$

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

To classify the quadratic form F , we use the eigenvalues λ_i $i=1,2,\dots,n$ of its Hessian matrix A , and we identify its optimum as follows:

Positive definite quadratic form if $\lambda_i > 0 \ i = 1, \dots, n$. It has a global minimum, represented in fig. 1a.

This manuscript was submitted on November 29, 2019, and accepted on July 23, 2021.

F. Mesa is a professor in the department of mathematics at the Universidad Tecnológica de Pereira, Colombia, and director of the research group Applied mathematics and education code COL0189463. (femesa@utp.edu.co).

D. M. Devia-Narvaez is a professor in the department of mathematics at the Universidad Tecnológica de Pereira, Colombia and an active member of the research group Applied Mathematics and Education. (dmdevian@utp.edu.co).

G. Correa-Vélez is a professor in the department of mathematics at the Universidad Tecnológica de Pereira, Colombia and an active member of the research group Applied Mathematics and Education. (gecove@utp.edu.co)



Quadratic form defined negative if $\lambda_i < 0 \ i = 1, \dots, n$ It has a global maximum, represented in fig. 1b.

Positive semi-definite quadratic form if $\lambda_i \geq 0 \ i = 1, \dots, n$. It has infinite minimum points, represented in fig. 1c.

Negative semi-definite quadratic form if $\lambda_i \leq 0 \ i = 1, \dots, n$. It has infinite maximum points, Its graphic representation would correspond to the inverted fig. 1d.

Quadratic form indefinite if $\exists i, j$ such that $\lambda_i > 0, \lambda_j < 0$. It has a saddle point, represented in fig. 1e.

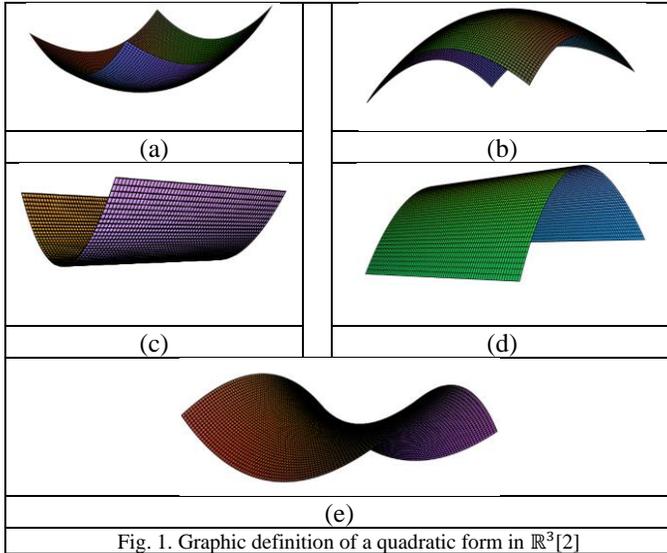


Fig. 1. Graphic definition of a quadratic form in \mathbb{R}^3 [2]

A. Gradient Method

This is a descent method which consists of choosing any point and from this through iterations we can build a sequence of points that is obtained by advancing on the line of maximum descent. This sequence will converge to a point very close to the solution

The problems that we can solve using this method are of the type:

$$\min_{x \in \mathbb{R}^n} F(\bar{x})$$

Here F is a quadratic form, $F: \mathcal{R}^n \rightarrow \mathcal{R}$ continuously differentiable, with an associated positive definite Hessian matrix A .

From the point \bar{x}^0 (initial position) we generate a succession of points \bar{x}^k given by (2)

$$\bar{x}^{k+1} = \bar{x}^k - \alpha_k \bar{Q}^k \quad (2)$$

α_k indicates the length of the step, known as the descent parameter (3), and we can find it by minimizing:

$$F(\bar{x}^k - \alpha_k \bar{Q}^k) = \min_{\alpha \in \mathbb{R}} F(\bar{x}^k - \alpha_k c)$$

$$\frac{d}{d\alpha} F(\bar{x}^k - \alpha_k \bar{Q}^k) = 0$$

$$\alpha_k = \frac{(\bar{b} - A\bar{x}_k)^T \bar{Q}^k}{(\bar{Q}^k)^T \bar{Q}^k} \quad (3)$$

The direction of maximum descent in \bar{x}^k of the quadratic form F is the gradient \bar{Q}^k given by equation (4) [3]:

$$\bar{Q}^k = \nabla F(\bar{x}^k) = -(\bar{b} - A\bar{x}_k) \quad (4)$$

We present the algorithm of this method in the following table I [4]:

TABLE I
GRADIENT METHOD ALGORITHM

Step	Description
1	Enter the quadratic function $F(\bar{x})$
2	Consider a point \bar{x}^0 . Do $k=0$
3	Choose the direction of maximum descent (gradient): $\bar{Q}^k = \nabla F(\bar{x}^k)$
4	Calculate the descent parameter: $\alpha_k = \frac{(\bar{b} - A\bar{x}_k)^T \bar{Q}^k}{(\bar{Q}^k)^T \bar{Q}^k}$
5	Do: $\bar{x}^{k+1} = \bar{x}^k - \alpha_k \bar{Q}^k$
6	Check convergence. if $\rightarrow \ \bar{Q}^k\ < \varepsilon$ the method stops and \bar{x}^k is the solution. Otherwise, do $k = k + 1$ and repeat from 2

B. Conjugated Gradient Method

We frequently find problems represented by systems that are sparse, since they arise when solving equations in partial derivatives in numerical form. It is there when we use this descent method, which helps us save memory by using only null elements, which they are to a large extent those observed in the matrix of coefficients that represents the system [3].

The conjugate gradient method initially requires the construction of an orthogonal base (Gramm-Schmidt method), determining with it the best solution or simply the one that is most efficient. The interesting thing about this method is the way in which the base is built guarantees the orthogonality of each element with respect to the previous one and automatically all the previous ones also satisfy this condition [5].

As a great advantage of this method is its speed of convergence, since it is faster than that of the descent method, as we can see in fig. 2

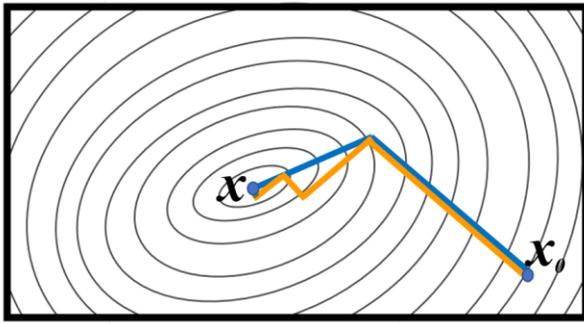


Fig. 2. Comparison of descent directions, steepest descent method (Orange), conjugate gradient (Blue) [6]

This method has the same approaches as the previous gradient method, since it is used to solve problems of optimization without restrictions:

$$\min_{x \in \mathbb{R}^n} F(\bar{x})$$

and starts from an initial position \bar{x}^0 generating the sequence of points \bar{x}^k given in (5):

$$\bar{x}^{k+1} = \bar{x}^k - \alpha_k \bar{Q}^k \quad (5)$$

The descent parameter that was previously in the (3) is modified by (6):

$$\alpha_k = \frac{(\bar{b} - A\bar{x}_k)^T \bar{a}^k}{(\bar{a}^k)^T \bar{a}^k} \quad (6)$$

Like the direction of descent that will be \bar{a}^k and is calculated by (7):

$$\bar{a}^k = -\bar{Q}^k + \beta_k \bar{a}^k, \bar{a}^0 = -\bar{Q}^0 \quad (7)$$

In this part β_k is a scalar known as the parameter of the conjugate gradient, which corresponds to different values depending on the conjugate gradient algorithm that is chosen for the solution of the problem [7]. Any conjugate gradient algorithm has a very simple general structure as illustrated below in the table II:

TABLE II
CONJUGATE GRADIENT METHOD ALGORITHM

Step	Description
1	Enter the quadratic function $F(\bar{x})$
2	Consider a point \bar{x}^0 . Do $k=0$
3	Choose the direction of maximum descent (gradient): $\bar{Q}^k = \nabla F(\bar{x}^k), \bar{a}^0 = -\bar{Q}^0$
4	Calculate the descent parameter: $\alpha_k = \frac{(\bar{b} - A\bar{x}_k)^T \bar{Q}_k}{(\bar{Q}_k)^T \bar{Q}_k}$
5	Do: $\bar{x}^{k+1} = \bar{x}^k - \alpha_k \bar{Q}^k$
6	Calculate $\bar{Q}^{k+1} = \nabla F(\bar{x}^{k+1})$
7	Calculate $\bar{y}^k = \bar{Q}^{k+1} - \bar{Q}^k \text{ y } \bar{s}^k = \bar{x}^{k+1} - \bar{x}^k$

- 8 Calculate β_k according to the conjugate gradient algorithm used
$$\bar{a}^{k+1} = -\bar{Q}^{k+1} + \beta_k \bar{a}^k$$
- 9 Check convergence. If $\rightarrow \|\bar{Q}^k\| < \varepsilon$ the method stops and \bar{x}^k is the solution. Otherwise, do $k = k + 1$ and repeat from 4.

B. Conjugate Gradient of Fletcher – Reeves

This method is an improvement of the descending gradient, which, using conjugated vectors, seeks to give a solution in fewer iterations, for the use of this method it is necessary to know the first displacement, since from this we can proceed to find the vectors directional that will be conjugated with each other [8].

For this method we proceed with an algorithm exactly like the previous method; however, we define the conjugate gradient parameter β_K in (8), containing the gradient parameters of the quadratic, present and previous function, parameters previously proposed in the conjugate gradient algorithm[9].

$$\beta_K = \frac{(\bar{Q}^{k+1})^T \bar{Q}^{k+1}}{(\bar{Q}^k)^T \bar{Q}^k} \quad (8)$$

C. Conjugate gradient of Hestenes-Stiefel

The conjugate gradient method has received a lot of attention and has been widely used in recent years. Although the pioneers of this method were Hestenes and Stiefel (1952) [8], the current interest starts from Reid (1971) posing it as an iterative method, which is the way it is most often used in the news [5].

As in the previous case, we reformulate Bk according to (9) to obtain its algorithm as follows:

$$\beta_K = \frac{(\bar{y}^k)^T \bar{Q}^{k+1}}{(\bar{Q}^k)^T \bar{s}^k} \quad (9)$$

III. ANALYSIS AND RESULTS

Next, different test systems for the Fletcher-Reeves Conjugate Gradient and Hestenes-Stiefel Conjugate Gradient methods will be presented.

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 (1).

For the stop criterion is used $\|\bar{Q}^k\| < \varepsilon = [10]^{-3}$ and as starting point \bar{x}^0 which is a column vector with n rows and all its components equal to one.

Proposed system for $n = 2$

$$A = \begin{bmatrix} 3 & -2 \\ -2 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$$

System for $n=3$

$$A = \begin{bmatrix} 30 & 5 & 8 \\ 5 & 100 & -3 \\ 8 & -3 & 50 \end{bmatrix}, \quad b = \begin{bmatrix} 7 \\ 5 \\ 10 \end{bmatrix}$$

A. Fletcher-Reeves Conjugate Gradient Test Systems.

The different test systems for the Fletcher-Reeves conjugate gradient method are presented below in Tables III, IV and V.

TABLE III
RESULTS OBTAINED GC FLETCHER-REEVES

Interactions	System dimensions
2	9
3	14

TABLE IV
RESULTS OBTAINED FOR $n=2$

i	x			Q
	1	1	3	6
0				
1	23.636	37.272	43.636	-21.818
2	36.404	35.416	0.162	11.143
3	37.129	38.898	0.6408	-0.1335
4	39.208	38.849	0.0071	0.3021
5	39.361	3.974	0.1397	-0.0239
6	3.981	39.726	0.0021	0.0714
7	39.851	39.941	0.0327	-0.0063
8	39.955	39.936	0.0007	0.0164
9	39.965	39.986	0.0077	-0.0016

i	d	α	β
0	3	6	
1	50.849	-0.7392	0.4545
2	0.23	11.044	0.251
3	0.6653	-0.0157	0.3152
4	0.0515	0.301	0.3125
5	0.143	-0.0042	0.296
6	0.0135	0.071	0.3142
7	0.0336	-0.0015	0.302
8	0.0032	0.0162	0.3073
9	0.0079	-0.0004	0.3063

TABLE V
RESULTS OBTAINED FOR $n=3$

i	x		
	1	1	1
0			
1	0.589237	-0.106777	0.486546
2	0.366001	0.117465	0.178653
3	0.290245	0.016762	0.159177
4	0.230833	0.061875	0.154669
5	0.210748	0.036907	0.162774
6	0.195204	0.051019	0.168272
7	0.188931	0.043359	0.170707
8	0.184161	0.047714	0.172294
9	0.182206	0.045356	0.173053
10	0.180743	0.046708	0.17354
11	0.18013	0.045978	0.173777
12	0.17968	0.0464	0.173927

i	Q
0	-36
1	-14.035.592
2	-5.996.602
3	-3.06
4	-1.471.729
5	-0.809206
6	-0.457403
7	-0.250393
8	-0.141776
9	-0.07739
10	-0.044175
11	-0.024047
12	-0.013828
13	-0.007507
14	-0.004347

i	d
0	-36
1	-14.060.543
2	-6.026.304
3	-3.08
4	-1.495.777
5	-0.813482
6	-0.464264
7	-0.251751
8	-0.143862
9	-0.077822
10	-0.044809
11	-0.024185
12	-0.014022
13	-0.007551
14	-0.004406

B. Hestenes-Stiefel Conjugate Gradient Test Systems

The different test systems for the Hestenes-Stiefel conjugate gradient method are presented below in Tables VI, VII, VIII and IX.

TABLE VI
RESULTS OBTAINED GC HESTENES-STIEFEL

Interactions	System dimensions
2	2
3	3
4	4
5	5

TABLE VII
RESULTS OBTAINED FOR $n=2$

i	x			Q
	1	1	3	6
0				
1	23.636	37.272	43.636	-21.818
2	4	4	0	0

i	d	α	β
0	3	6	-
1	59.504	0.9917	0.4545

TABLE VIII
RESULTS OBTAINED FOR n=3

<i>i</i>	<i>x</i>		
0	1	1	1
1	0.589237	-0.106777	0.486546
2	0.308488	0.036849	0.104164
3	0.179194	0.046263	0.174104

<i>i</i>	<i>Q</i>		
0	-36	-97	-45
1	-14.035.592	14.191.246	-19.361.546
2	-3.272.216	0.085052	2.434.438

<i>i</i>	<i>d</i>			
0	-38	-57	8	46
1	-358.961	132.257	1.511.123	5.896.596
2	140.535	0.82727	2.848.809	463.497
3	108.342	0.17977	-1.078.084	2.116.537
4	5.24E-15	8.52E-14	2.74E-15	1.73E-15

<i>i</i>	<i>alpha</i>	<i>beta</i>
0	-	-
1	0,01141008	5,32362825
2	0,01730629	1,24352268
3	0,03570331	-1,4900-14

TABLE IX
RESULTS OBTAINED FOR n=4

<i>i</i>	<i>x</i>			
0	1	1	1	1
1	0.38903	0.08354	112.862	173.959
2	-174.743	0.87071	202.801	524.913
3	-123.671	0.90078	306.332	693.356
4	-0.93943	0.90571	276.751	751.431

<i>i</i>	<i>Q</i>			
0	-38	-57	8	46
1	-168.573	417.838	111.031	35.919
2	325.962	-600.468	206.821	158.898
3	715.713	-0.03669	-182.346	903.819

<i>i</i>	<i>d</i>			
0	-38	-57	8	46
1	-358.961	132.257	151.112	589.659
2	140.535	0.8272	28.488	463.497
3	108.342	0.1797	-107.808	211.653
4	5.24E-15	8.52E-14	2.74E-15	1.73E-15

<i>i</i>	<i>alpha</i>	<i>beta</i>
0	-	-
1	0.0160781	311.616.601
2	0.05951814	86.791.544
3	0.03634179	719.957.007
4	0.02743856	-9.26E-15

C. Comparison of methods

The different simulations corresponding to the descent methods will be shown below.

Test system n = 2

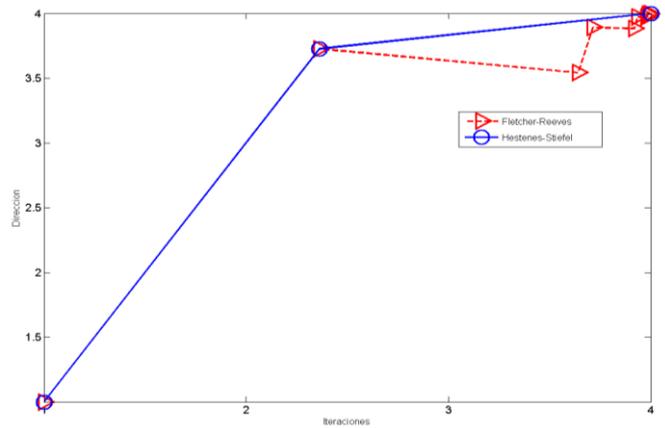


Fig. 3 . Direction vs. Iterations. Red Fletcher-Reeves, Blue Hestenes-Stiefel.

Test system n=3

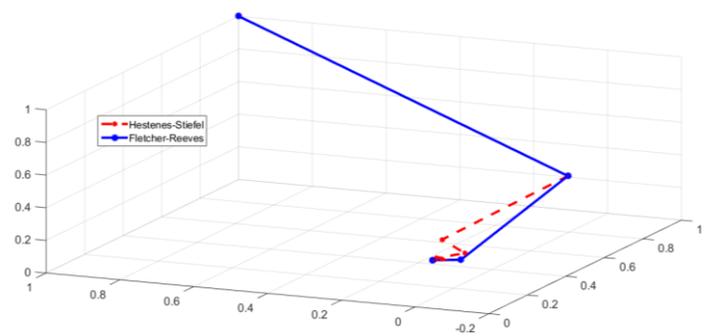


Fig. 4. Direction vs. Iterations. Red Fletcher-Reeves, Blue Hestenes-Stiefel

In the previous graphs (fig.3 and fig. 4) it is possible to observe the behavior of the Fletcher - Reeves and Hestenes-Stiefel methods for spaces of dimensions n = 2 and n = 3. In them we observe that more directions and observation points can be calculated by Fletcher-Reeves, versus the Hestenes-Stiefel results.

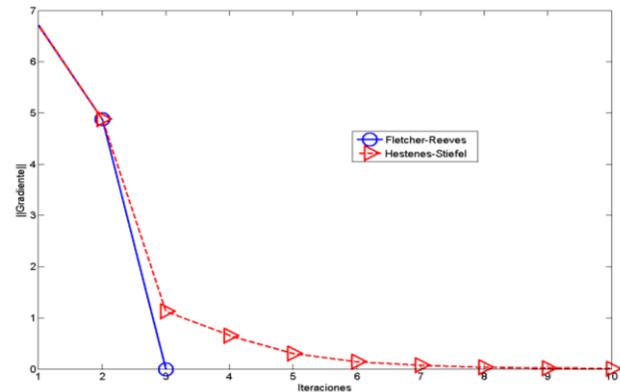


Fig. 5. Standard Gradient vs. Iterations. Red Fletcher-Reeves, Blue Hestenes-Stiefel.

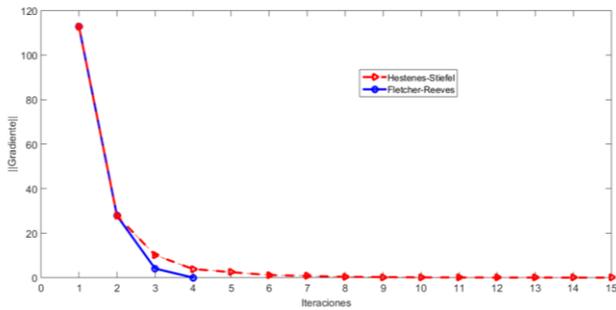


Fig. 6. Standard Gradient vs. Iterations. Red Fletcher-Reeves, Blue Hestenes-Stiefel.

In the graphs represented in fig. 5 and fig. 6 the norm for R^2 and R^3 can be seen, concluding that the convergence is faster with Hestenes-Stiefel than with the method proposed by Fletcher-Reeves.

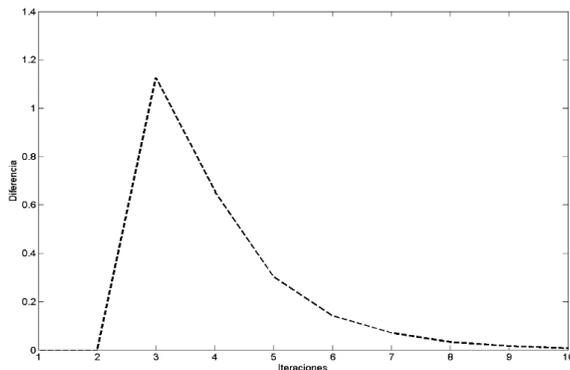


Fig. 7. Difference Standard Gradient vs. Iterations.

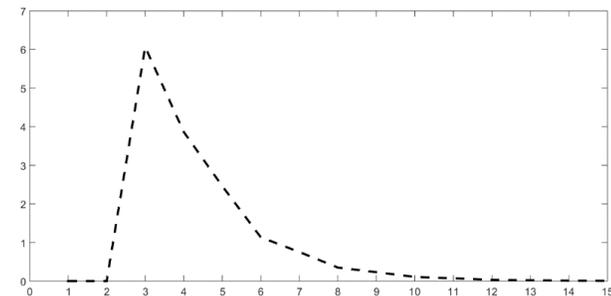


Fig. 8. Difference Standard Gradient vs. Iterations.

Finally, fig. 7 and fig. 8 gather the iteration-to-iteration difference between the standards of the grades in fig. 5 and fig. 6 respectively. The difference is clearly seen in the iteration after the method improved by Hestenes-Stiefel has already found the optimum of the systems proposed for 2 and 3 dimensions.

IV. CONCLUSIONS

When the quadratic form has distorted or too eccentric contours, more interactions will be required in order for the

Fletcher-Reeves method to converge. This is because by rounding the errors they result in the need for more interactions.

For a greater effectiveness of the Fletcher-Reeves method, we must periodically restart the method from an appreciable number of steps, in which the new search direction corresponds to that of the steep descent.

The results show us that the Fletcher-Reeves method is a better optimization method compared to the different search methods using patterns in particular than the Hestenes-Stiefel method [10].

REFERENCES

- [1]. J. Nocedal and S. J. Wright, "Numerical optimization", Springer, New York, 2006. doi: <https://doi.org/10.1007/978-0-387-40065-5>
- [2]. M. A. Fontelos, "Fundamentos matemáticos de la Ingeniería", Librería-Editorial Dykinson, 2007.
- [3]. U. Kindelán, "Resolución de sistemas lineales de ecuaciones: Método del gradiente conjugado", Universidad Politécnica de Madrid, 2007
- [4]. A. Katrutsa, M. Botchev, G. Ovchinnikov and I. Oseledets, "How to optimize preconditioners for the conjugate gradient method: a stochastic approach", Numerical Analysis, 2017.
- [5]. A. C. Ledesma, "Resolución de Grandes Sistemas de Ecuaciones Lineales", Instituto de Geofísica UNAM, 2006.
- [6]. B. Calvo and G Santafé, "scmp: Statistical comparison of multiple algorithms in multiple problems" *The R Journal*, (8) 1, 2016.
- [7]. N. Andrei, "Scaled conjugate gradient algorithms for unconstrained optimization", Computational Optimization and Applications 38(3) 401-416, 2007. doi: 10.1007/s10589-007-9055-7
- [8]. R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients", *The computer journal*, 7(2), 149-154, 1964.
- [9]. F. Mesa, P. P. Cardenas-Alzate and C. A. Rodriguez Varela, "Comparison of the Conjugate Gradient Methods of Liu-Storey and Dai-Yuan", *Contemporary Engineering Sciences*, 10(35), 1719, 1726.
- [10]. R. Fletcher, "Practical Methods of Optimization", Wiley, New York, 1980.



Fernando Mesa, university professor of the Mathematics Department of the Technological University of Pereira, Colombia and director of the applied mathematics and education research group. Master of Science (Msc) in physical instrumentation (2007). Member of the plasma laboratory of the National University of Colombia, located in Manizales, Colombia. Member of the group of non-linear differential equations "GEDNOL" at the Technological University of Pereira, Colombia. Director of the research group in Applied Mathematics and Education "GIMAE" at the Technological University of Pereira, Colombia.

ORCID: <https://orcid.org/0000-0002-3418-5555>



Diana Marcela Devia Narváez.

Associated teacher of mathematics and physics at Universidad Tecnológica de Pereira, Colombia. PhD in engineering (2012). Master of Science (Msc) in the faculty of Physics - Science (2010). Member of plasma laboratory from Universidad Nacional de Colombia

located in Manizales, Colombia. Member of nonlinear differential equations group “GEDNOL” at Universidad Tecnológica de Pereira, Colombia. Area of expertise: material processing through assisted plasma techniques, structural characterization mechanical of materials, simulation and modeling of material’s physical properties.

ORCID: <https://orcid.org/0000-0002-0447-4663>



German Correa Vélez, university lecturer of the department of Mathematics at the Universidad Tecnológica de Pereira, Colombia. Master of Science (Msc) in mathematics (2008). Member of nonlinear differential equations group “GEDNOL” at Universidad Tecnológica de Pereira, Colombia. Member of the

research group in Applied Mathematics and Education “GIMAE” at the Universidad Tecnológica de Pereira.

ORCID: <https://orcid.org/0000-0002-5244-3095>