ANALELE UNIVERSITĂȚII
"EFTIMIE MURGU" REȘIȚA
ANUL XIII, NR. 1, 2006, ISSN 1453 - 7397

Cornelia Victoria Anghel

The Gradient Method – A Matriceal Algorithm Used in Optimization

In the paper is the mathematical gradient method described and an eloquent numerical example is given. The Matriceal gradient method is used in optimization problems. As the result we obtain the optimum of the function. The advantage of this method is that in some conditions can minimized the order of the matrix.

1. Introduction

The gradient method is an algorithm used to solve the convex non-linear programs, with linear or non-linear restrictions, such as:

$$\begin{cases} \sum_{j=1}^{n} a_{ij} X_j \le b_i,\ 1 \le i \le m \\ \min f(x) \end{cases} \quad (1)$$

where f(x) function is convex and differentiable on the domain:

$$D = \left\{ x \in \Re^n \middle| \sum_{j=1}^{n} a_{ij} x_j \le b_i,\ 1 \le i \le n \right. \quad (2)$$

2. The gradient algorithm

The algorithm covers several steps:

An admissible initial random solution is determined

$$x^{1}{}'(x_1^1, x_2^1, ..., x_n^1) \in D$$

If the gradient $\nabla f(x^k) = 0, (k \le 1), then\ the\ po\text{int}\ \ x^k$ is the optimal solution

If the gradient $\nabla f(x^k) \ne 0,$ the multitude of coefficients

$$I^k = \left\{ i \middle| \sum_{j=1}^{n} a_{ij} x_j^{\ k} = b_j ,\ 1 \le i \le m \ \right\} \quad (3)$$

and the equation hyperplanes:

$$H_i^k = \left\{ x \in \Re^n \middle| \sum_{jj=1}^{n} a_{ij} x_j^{\ k} = b_i \ \right\} i \in I^k \quad (4)$$

that contain the point x^k

The vector $(-\nabla f(x^k))$ is projected on the liniar variety determined by the planes intersection
$H_i^k, i \in I^k$ and has the expression:

$$P_k[-\nabla f(x^k)] = [I - {}^t A_k (A_k {}^t A_k) A_k][-\nabla f(x^k)] \quad (5)$$

where Ak is the matrix build with the coefficients of the unknown from the hyperplanes equations $H_i^k, i \in I^k$.

The operator: $P_k = I - {}^t A_k (A_k {}^t A_k)^{-1} A_k$ (6)

Is called projection matrix
If we simultaneously obtain:

$$P_k[-\nabla f(x^k)] = 0, şi (A_k {}^t A_k)^{-1} A_k [-\nabla f(x^k)] \ge 0 \quad (7)$$

then x^k is the optimum solution of the program

If Pk$[-\nabla f(x^k)] \ne 0$, we note $y^k = P_k[-\nabla f(x^k)]$ and the number λ_0^k is determined
Such as

$$f(x^k + \lambda_0^k y^k) = \min\ f(x^k + \lambda y^k) \quad (8)$$

if $\lambda_0^k = 0$, then point x^k is an optimum solution.

if $\lambda_0^k \ne 0$, then from point x^k we create a projection in the direction of y^k, such as:

We

$$\widetilde{I}_1^k = \left\{ i \notin I^k \middle| \sum_{j=1}^{n} a_{ij} Y_j^k > 0 \ \right\} \quad (9)$$

and

$$\lambda_i = \frac{b_i - \sum_{j=1}^{n} a_{ij} x_j^{\ k}}{\sum_{j=1}^{n} a_{ij} y_j^{\ k}}, i \in I_1^k \qquad (9')$$

so that λ_i>0, $(\forall) i \in I_1^k$

We define the number:

$$\gamma^k = \begin{cases} \min \lambda i, dacă \\ i \in I_1^k \qquad\qquad I_1^k \neq \varnothing \quad \text{şi} \quad I_1^k = \varnothing \\ \infty, dacă \end{cases} \qquad (10)$$

si $\lambda^k = \min(\lambda_0^k, \gamma^k)$ (11)

and with its help we determine the point

$x^{k+1} = x^k + \lambda^k y^k$ (12)

which alows us to repete the algorythm from stage b)

3. Numerical example

We solbe the convex program: $\begin{cases} 4x_1 + 6x_2 - 7 \leq 0 \\ x_2 - 1 \leq 0 \\ -x_1 \leq 0 \\ -x_2 \leq 0 \end{cases}$

min [$f(x) = x_1^2 + x_2^2 - 2x_1 - x_2$]

Solution

We calculate the gradient of the function

$$\nabla f(x) = \frac{\partial f}{\partial x_1} i + \frac{\partial f}{\partial x_2} j, adică \nabla f(x) = (2x_1 - 2)i + (2x_2 - 1)j$$

or with the help of the matrix we find $\nabla f(x) = (2x_1 - 2, \quad 2x_2 - 1)$

We choose an initial admissible solution $x^1 = (\frac{1}{4}, 1)$, for which

$\nabla f(x^1) = (-\frac{3}{2}, 1) \neq 0$

We note the multitude of the coeffiecients of the unknown that occur in the gradient's expression $I^1 = \{1,2\}$ and of the matrix built with the coefficients of the unknown x_1 and x_2 from the first and the second equation

$$A_1 = \begin{pmatrix} 4 & 6 \\ 0 & 1 \end{pmatrix}$$

We determine the necessary elements for the projections

$${}^tA_1 = \begin{pmatrix} 4 & 0 \\ 6 & 1 \end{pmatrix}, \quad A_1{}^tA_1 = \begin{pmatrix} 52 & 6 \\ 6 & 1 \end{pmatrix}, \det(A_1{}^tA_1) = 16 \neq 0$$

thus the matrix is irreversible and its reverse is:

$$(A_1{}^tA_1)^{-1} = \frac{1}{16}\begin{pmatrix} 1 & -6 \\ -6 & 52 \end{pmatrix}$$

We calculate:

$$(A_1{}^tA_1)^{-1}A_1 = \frac{1}{16}\begin{pmatrix} 1 & -6 \\ -6 & 52 \end{pmatrix}\begin{pmatrix} 4 & 6 \\ 0 & 1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1 & 0 \\ -6 & 4 \end{pmatrix}$$

The product

$${}^tA_1(A_1{}^tA_1)^{-1}A_1 = \frac{1}{4}\begin{pmatrix} 4 & 0 \\ 6 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -6 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The projection operator:

$$P_1 = I - {}^tA_1(A_1{}^tA_1)^{-1}A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

thus $P_1[-\nabla f(x)] = 0$

We determine a vector

$$U' = (A_1{}^tA_1)^{-1}A_1[-\nabla f(x')] = \frac{1}{4}\begin{pmatrix} 1 & 0 \\ -6 & 4 \end{pmatrix}\begin{pmatrix} \frac{3}{2} \\ -1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} \frac{3}{2} \\ -13 \end{pmatrix}$$

because $U'_2 = -13 < 0$, x' is not c\solution.

In order to continue the algorythm in matrix A1 the second line is reversed and we obtain the matrix:

$$B_1 = (4,6), {}^tB_1 = \begin{pmatrix} 4 \\ 6 \end{pmatrix}, B_1{}^tB_1 = 52, iar (B_1{}^tB_1)^{-1} = \frac{1}{52}$$

We calculate the necessary elements for the projection operator.

$$(B_1{}^t B_1)^{-1} B_1 = \frac{1}{26}(2,3)$$

$${}^t B_1 (B_1{}^t B_1)^{-1} B_1 = \frac{1}{13}\begin{pmatrix} 4 & 6 \\ 6 & 9 \end{pmatrix}$$

The projection matrix will be:

$$P_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{1}{13}\begin{pmatrix} 4 & 6 \\ 6 & 9 \end{pmatrix} = \frac{1}{13}\begin{pmatrix} 9 & -6 \\ -6 & 4 \end{pmatrix}$$

We apply the gradient projection:

$$P_2\left[-\nabla f(x') \right] = \frac{1}{13}\begin{pmatrix} 9 & -6 \\ -6 & 4 \end{pmatrix}\begin{pmatrix} \frac{3}{2} \\ -1 \end{pmatrix} = \frac{1}{13}\begin{pmatrix} \frac{39}{2} \\ -\frac{26}{2} \end{pmatrix} = \begin{pmatrix} \frac{3}{2} \\ -1 \end{pmatrix} \neq 0$$

We note with $y_1 = P_2\left[-\nabla f(x') \right] = {}^t(\frac{3}{2}, -1)$

And from x^1 we descend in the direction of y1.
The half-line gets intersected.

$$D_1 = \{x_\lambda^1 = x^1 + \lambda y_1 = (\frac{1}{4} + \frac{3\lambda}{2}, (1-\lambda)); \ \lambda > 0\}$$

with the hyperplanes H3 and H4

$$-\left(\frac{1}{4} + \frac{3\lambda}{2}\right) = 0 \text{ and we obtain } \lambda_3 = -\frac{1}{6}$$

$-(1-\lambda)=0$ $\lambda_4 = 1 > 0$

The value of function f on the semidreapta D1 is:

$$\varphi'(\lambda) = f(\frac{1}{4} + \frac{3\lambda}{2}, 1-\lambda) = (\frac{1}{4} + \frac{3\lambda}{2})^2 + (1-\lambda)^2 - 2(\frac{1}{4} + \frac{3\lambda}{2}) - 1 + \lambda$$

We derive and get:

$$\varphi'(\lambda) = \frac{13}{4}(2\lambda - 1)$$

By cancelling the derivative we find $\lambda 0 \frac{1}{2} > 0$

We determine λ'=min(λ0, λ4)=min($\left(\frac{1}{2},1\right)=\frac{1}{2}$

With its help we determine the new solution

$$x^2 = \lambda^1 x^1 = {}^t\left(\frac{1}{4}+\frac{3}{2}.\frac{1}{2};1-\frac{1}{2}\right) = {}^t\left(1,\frac{1}{2}\right)$$

we calculate $\nabla f(x^2) = {}^t(0,0)$

It is specified that $x^2 = {}^t\left(1,\frac{1}{2}\right)$ is the optimum solution, for which $f_{\min} = -\frac{5}{4}$

4. Conclusions

The method of the projecting gradient is a matriceal method used to determine the optimum of the function. It's advantage is that in some conditions it can decrease the order of the matrix.

References

[1]. C.V. Anghel, N. Gillich sa. „*Cercetari Operationale*", Ed. MODUS, Resita, 2005

[2]. Anghel I, Anghel C.V. *Algebra liniara. Programare liniara*; Ed. „Eftimie Murgu" U.E.M.R., Reşiţa, 2003

[3]. The MathWorks, Inc. – *MATLAB* Version 5.3, 1999.

Address:

- PhD Eng. Cornelia Victoria Anghel "Eftimie Murgu" University of Romania, Piaţa "Traian Vuia", nr. 1-4, Reşiţa, canghel@uem.ro