



20 e 21 de outubro
Instituto Nacional de Pesquisas Espaciais - INPE
São José dos Campos - SP

The q -steepest descent method using the q -gradient vector

Aline Cristina Soterroni¹, Luiz Roberto Galski², Fernando Manuel Ramos³

¹Applied Computing Program – CAP
National Institute for Space Research – INPE

²Satellite Control Centre – CCS
National Institute for Space Research – INPE

³Computing and Applied Mathematics Associated Laboratory – LAC
National Institute for Space Research – INPE

`aline.soterroni@lac.inpe.br, galski@ccs.inpe.br, fernando@lac.inpe.br`

Abstract. *In this work we make use of the first-order partial q -derivatives of a function of n variables to calculate the q -gradient vector and take the negative direction as the search direction of unconstrained optimization methods. We present a q -version of the classical steepest descent method called the q -steepest descent method. This q -version is reduced to the classical version whenever the parameter q , that is used to calculate the q -gradient vector, is equal to 1. We applied the classical steepest descent method and the q -steepest descent method to an unimodal and a multimodal test function.*

Keywords: q -derivative, q -gradient vector, q -steepest descent method

1. Introduction

It is well-known that along the direction given by the gradient vector, the objective function $f(x)$ increases most rapidly. If the optimization problem is to minimize an objective function, then it is intuitive to use the steepest descent direction $-\nabla f(x)$ as the search direction in the optimization methods. Here we introduce the q -gradient vector that is similar to the usual gradient vector, but instead of the usual first-order partial derivatives we use the first-order partial q -derivatives obtained from the Jackson's derivative, also referred to as q -difference operator, or q -derivative operator or simply q -derivative. Frank Hilton Jackson gave many contributions related to basic analogues or q -analogues, especially on basic hypergeometric functions [Chaundy 1962], and in the beginning of nineteenth century he generalized the concepts of derivative and integration in the q -calculus context and created the q -derivative and the q -integration [Jackson 1908, Jackson 1909, Jackson 1910a, Jackson 1910b]. In the q -derivative, instead of the independent variable x of a function $f(x)$ be added by a infinitesimal value,

it is multiplied by a parameter q that is a real number different from 1. And in the limit, $q \rightarrow 1$, the q -derivative tends to the usual derivative. Our proposal approach has two advantages. On the one hand, the closer to 1 the value of the parameter q is, the closer to the classical gradient vector the q -gradient vector will be. For monomodal poorly scaled functions, to use a direction close, but not equal, to the steepest descent direction can reduce the zigzag movement towards the solution. On the other hand, when $q \rightarrow 1$ the negative of the q -gradient vector can make any angle with the negative of the classical gradient vector and the direction search can point to any direction. It can be interesting for multimodal functions because the search procedure can escape from local minima. In this work we use the q -gradient vector based on the q -derivative and use its negative as search direction in optimization methods. The paper is organized as follows. In the Section 2 the q -gradient vector is defined. In the Section 3 we present a q -version of the classical steepest descent method. Section 4 deals with the performance of the q -steepest descent method and the classical steepest descent method for numerical examples. And the Section 5 contains final considerations.

2. The q -gradient vector

Let $f(x)$ be a real-valued continuous function, the q -derivative of f is given by

$$D_q f(x) = \frac{f(x) - f(qx)}{(1-q)x}, \quad (1)$$

where $x \neq 0$ and $q \neq 1$. The parameter q is usually taken as a fixed real number $0 < q < 1$. However, this hypothesis can be weakened and q be a fixed real number different from 1 [Koekojev and Koekojev 1993]. And when $q \rightarrow 1$, the q -derivative tends to the usual derivative.

Note that the q -derivative is not defined at $x = 0$. Therefore, for real-valued functions differentiable at $x = 0$, the q -derivative can be given by [Koekojev and Koekojev 1993]

$$D_q f(x) = \begin{cases} \frac{f(x) - f(qx)}{(1-q)x}, & x \neq 0, \quad q \neq 1 \\ \frac{df(0)}{dx}, & x = 0. \end{cases} \quad (2)$$

For a real-valued continuous function of n variables the gradient vector of f is the vector of the n first-order partial derivatives of f . Similarly, the q -gradient vector of f is the vector of the n first-order partial q -derivatives of f . Before introducing the q -gradient vector it is convenient to define the first-order partial q -derivative of a real-valued continuous function of n variables differentiable at $\mathbf{x} = 0$ with respect to the variable x_i [P. M. Rajkovic and Stankovic 2005]

$$D_{q,x_i} f(\mathbf{x}) = \frac{f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, qx_i, x_{i+1}, \dots, x_n)}{(1-q)x_i}, \quad (3)$$

with $x_i \neq 0$ e $q \neq 1$. A modified notation was used. Similarly to the Equation (2), we can

define the first-order partial q -derivative of f considering the $x_i = 0$ and $q = 1$ as follows

$$D_{q,x_i}f(\mathbf{x}) = \begin{cases} \frac{f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, qx_i, x_{i+1}, \dots, x_n)}{(1-q)x_i}, & x_i \neq 0, \quad q \neq 1 \\ \frac{\partial f}{\partial x_i}(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n), & x_i = 0 \\ \frac{\partial f}{\partial x_i}(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n), & q = 1. \end{cases} \quad (4)$$

Therefore, let $f(\mathbf{x})$ be a real-valued continuous function of n variables and differentiable at $\mathbf{x} = 0$, and the parameter $\mathbf{q} = (q_1, \dots, q_i, \dots, q_n)$. We introduce here the q -gradient as

$$\nabla_{\mathbf{q}}f(\mathbf{x}) = [D_{q_1, x_1}f(\mathbf{x}) \dots D_{q_i, x_i}f(\mathbf{x}) \dots D_{q_n, x_n}f(\mathbf{x})]. \quad (5)$$

And in the limit, $q_i \rightarrow 1$, for all i ($i = 1, \dots, n$), the q -gradient vector returns to the usual gradient vector.

3. The q -steepest descent method

A general optimization strategy is to consider an initial set of variables, \mathbf{x}_0 , and apply an iterative procedure given by $\mathbf{x}_k = \mathbf{x}_{k-1} + \alpha_k \mathbf{s}_k$, where k is the iteration number, \mathbf{x} is the vector of variables, α is the steplength and \mathbf{s} is a search direction vector. This process continues until either no additional reduction in the value of the objective function can be made or the solution point has been approximated with sufficient accuracy [Vanderplaats 2007].

In the steepest descent method the search direction \mathbf{s}_k is given by the negative of the gradient vector at the point \mathbf{x}_k , $-\nabla f(\mathbf{x}_k)$, and the steplength α_k can be found by a one-dimensional search performed in the direction \mathbf{s}_k . Similarly, the search direction for the q -steepest descent method is given by negative of the q -gradient vector $-\nabla_{\mathbf{q}}f(\mathbf{x}_k)$ (Equation (5)). Besides that, we have to obtain the parameter $\mathbf{q} \in \mathbf{R}^n$. Our strategy is to generate random numbers by log-normal distribution with a fixed mean $\mu = 1$ and a variable standard deviation σ . The initial standard deviation σ_0 should be a real positive number different from zero and during the iterative procedure it is reduced to zero by $\sigma_k = \beta \cdot \sigma_{k-1}$, where β is the reduction factor. When σ tends to 0, the parameter q tends to 1 and the q -gradient vector tends to the usual gradient vector. In other words, when $\sigma \rightarrow 0$ the q -steepest descent method is reduced to the classical steepest descent method. We have chose this distribution because of its multiplicative effects [E. Limpert and Abbt 2001]. In addition, only positive numbers can be generated, when σ decreases with $\mu = 1$ the mode approaches to 1, and the skewed shape distribution for $\mu = 1$ shows the same likelihood for a log-normal random number to occur in the intervals $(0, 1)$ or $(1, \infty)$. The optimization algorithm for the q -steepest descent method is given below.

Algorithm 1

- Step 1: Initialize randomly the set of variables $\mathbf{x}_0 \in \mathbb{R}^n$, set the mean $\mu = 0$, the initial standard deviation σ_0 and the reduction factor β .
- Step 2: Set $k := 1$.
- Step 3: Generate the parameter $\mathbf{q} = (q_1, \dots, q_i, \dots, q_n)$ by a log-normal distribution with mean μ and standard deviation σ_k .
- Step 4: Compute the search direction $\mathbf{s}_k = -\nabla_{\mathbf{q}} f(\mathbf{x}_k)$.
- Step 5: Find the steplength α_k by one-dimensional search.
- Step 6: Compute $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{s}_k$.
- Step 7: If the stopping conditions are reached, then STOP; Otherwise go to Step 8.
- Step 8: Reduce the standard deviation $\sigma_k = \beta \cdot \sigma_{k-1}$.
- Step 9: Set $k := k + 1$, and go to Step 3.

4. Numerical Examples

We applied the classical steepest descent method and the q-version presented here (Algorithm 1) for the same functions, initial set of variables, stopping condition, and strategy to find the steplength. We considered the unimodal Rosenbrock function [Shang and Qiu 2006] and the highly multimodal Rastrigin function [Ballester and Carter 2004]. We generated 50 different initial variables \mathbf{x}_0 by a uniform distribution in the interval $(-2.048, 2.048)$ for the Rosenbrock function, and the interval $(-5.12, 5.12)$ for the Rastrigin function, and we used the same set for the steepest descent method and the q-steepest descent method. The stopping condition was the maximum number of function evaluations 10^5 . And for the one-dimension searches we used the golden section method by the code found in [W. H. Press and Flannery 1996] with the fractional precision tolerance equal to 10^{-8} . The steplength used to bracket the minimum before applying the golden search method in [W. H. Press and Flannery 1996] was set at 1% of the interval width considered to generate the initial points of each function. For the numerical derivatives we used the Ridder's method whose code is also found in [W. H. Press and Flannery 1996]. Besides that, we considered the parameter q numerically equal to 1 when $q \in [1 - \text{tol}, 1 + \text{tol}]$ with $\text{tol} = 10^{-4}$.

The Rosenbrock function is a well-known test function for numerical optimization problems given by

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2. \quad (6)$$

The minimum is $\mathbf{x}^* = (1, 1)$ with $f(\mathbf{x}^*) = 0$. We used the mean $\mu = 1$ and initial standard deviation $\sigma_0 = 0.5$ with reduction factor $\beta = 0.999$ for generating the parameter \mathbf{q} in the Step 3 of the Algorithm 1. The performance results for the Rosenbrock function are shown in Figure 1.

It can be seen from the results shown in Figure 1 that the steepest descent method converges more slowly than the q-steepest descent method. When the stopping condition is reached, the mean of the best function values for the steepest descent method is equal to $4.0813 \cdot 10^{-4}$ and for the q-steepest descent method it is equal to $4.9961 \cdot 10^{-10}$. Then, the q-steepest descent method has a better performance than the classical method.

The Rastrigin function is given by

$$f(\mathbf{x}) = 10n + \sum_{i=1}^n (x_i - 10 \cos(2\pi x_i)), \quad (7)$$

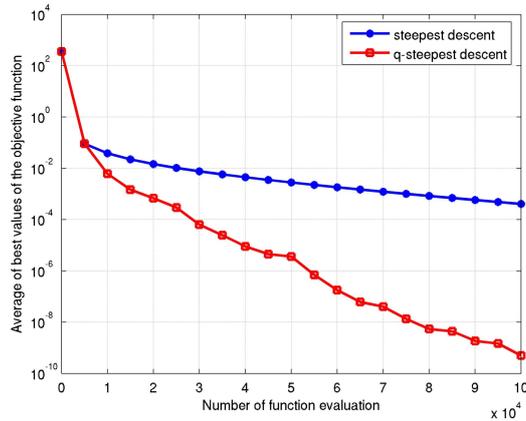


Figure 1. Average of the best function values, found in 50 independent runs of each algorithm, versus the number of function evaluations for Rosenbrock function.

where n is the number of variables. The global minimum is located at $\mathbf{x}^* = \mathbf{0}$ with $f(\mathbf{x}^*) = 0$. We set the mean $\mu = 1$ and initial standard deviation $\sigma_0 = 5.0$ with reduction factor $\beta = 0.995$ for generating the parameter q . The Figure 2 shows the performance results for the Rastrigin function with two variables.

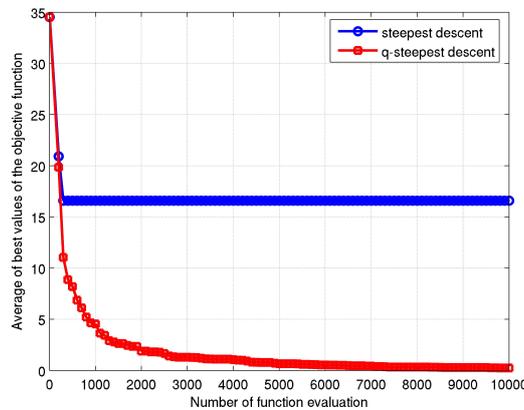


Figure 2. Average of the best function values, found in 50 independent runs of each algorithm, versus the number of function evaluations for Rastrigin function.

The Figure 2 displays the premature convergence of the steepest descent method. It was expected because the methods based on gradient vector move toward the local minimum closest to the initial point [Nocedal and Wright 1999]. As you can see, the q-steepest descent method has a slowly convergence. However, in the q-steepest descent method among 50 different initial points, 42 reached the global minimum of the Rastrigin function. The mean of the best function values for the steepest descent method is equal to 16.5958 and for the q-steepest descent method it is equal to 0.1989.

5. Final Considerations

In this work we used the q-gradient vector based on the first-order partial q-derivatives of a function of n variables and took its negative direction as the search direction in the

q-version of the steepest descent method called the q-steepest descent method. For the monomodal Rosenbrock function the steepest descent method converged more slowly than the q-steepest descent method. And for the highly multimodal Rastrigin function the q-steepest descent method moved toward the global minimum for several initial points. The performance results show the advantage of using the negative of the q-gradient vector as the direction search in unconstrained optimization methods with a stochastic strategy for the parameter q . Particularly, for the multimodal function it could be seen that the q-steepest descent method has mechanisms to escape from the many local minima and move toward the global minimum.

Acknowledgement

This work was supported by National Council for Scientific and Technological Development (CNPq), Brazil.

References

- Ballester, P. J. and Carter, J. N. (2004). An effective real-parameter genetic algorithm with parent centric normal crossover for multimodal optimisation. 3102:901–913.
- Chaundy, T. W. (1962). Frank hilton jackson. 37:126–128.
- E. Limpert, W. A. S. and Abbt, M. (2001). Log-normal distributions across the sciences: keys and clues. 51:341–352.
- Jackson, F. H. (1908). On q-functions and a certain difference operator. 46:253–281.
- Jackson, F. H. (1909). A q-form of taylor’s theorem. 38:62–64.
- Jackson, F. H. (1910a). On q-definite integrals. 41:193–203.
- Jackson, F. H. (1910b). q-difference equations. 32:305–314.
- Koekojev, J. and Koekojev, R. (1993). A note on the q-derivative operator. 176:627–634.
- Nocedal, J. and Wright, S. J. (1999). *Numerical optimization*. Springer-Verlag, New York.
- P. M. Rajkovic, S. D. M. and Stankovic, M. S. (2005). On q-newton-kantorovich method for solving systems of equations. 168:1432–1448.
- Shang, Y. and Qiu, Y. (2006). A note on the extended rosenbrock function. 14:119–126.
- Vanderplaats, G. N. (2007). *Multidiscipline design optimization*. Vanderplaats Research and Development, INC., Monterey.
- W. H. Press, S. A. Teukolsky, W. T. V. and Flannery, B. P. (1996). *Numerical recipes in Fortran 90: The art of parallel scientific computing*. Cambridge University Press, New York.