Proceeding Series of the Brazilian Society of Applied and Computational Mathematics, Vol. 3, N. 1, 2015.
Trabalho apresentado no XXXV CNMAC, Natal-RN, 2014.

O uso do vetor q-gradiente como direção de busca em métodos de otimização contínua

Aline C. Soterroni, **Fernando M. Ramos**,
Laboratório Associado de Computação e Matemática Aplicada, LAC, INPE,
12227-010, São José dos Campos, SP
E-mail: alinecsoterroni@gmail.com, fernando.ramos@inpe.br,

Roberto L. Galski,
Centro de Rastreio e Controle de Satélites, CRC, INPE
12210-080, São José dos Campos, SP
E-mail: galski@ccs.inpe.br,

Érica J. C. Gouvêa, **Marluce Scarabello**
Programa de Doutorado em Computação Aplicada, CAP, INPE,
12227-010, São José dos Campos, SP
E-mail: ericagouvea@gmail.com, marluce.scarabello@inpe.br.

Resumo: *O q-cálculo surgiu da generalização de séries, funções, números especiais, entre outros, por meio de um parâmetro multiplicativo q e tal que, no limite $q \to 1$, retomam suas respectivas versões clássicas. Com base nos trabalhos de Euler e Heine, o reverendo inglês Frank Hilton Jackson desenvolveu, no ínicio do século XX, o q-cálculo de forma sistemática com destaque para a reintrodução do conceito de q-derivada, que passou a ser mais conhecida como derivada de Jackson, e a criação da q-integral. Nas últimas décadas, o q-cálculo tem conectado matemáticos e físicos em aplicações de mecânica estatística, teoria dos números e análise combinatória. Este trabalho mostra como o conceito de q-derivada pode ser usado na área de otimização contínua por meio do vetor q-gradiente, uma generalização do vetor gradiente clássico no contexto do q-cálculo.*

Palavras-chave: *q-cálculo, q-derivada, derivada de Jackson, q-gradiente*

1 Introdução

Generalizações no contexto do q-cálculo remontam aos trabalhos de Fermat, Euler, Heine e Gauss, mas foi no início do século XX que o reverendo inglês Frank Hilton Jackson contribuiu para o desenvolvimento do q-cálculo de forma sistemática (veja [1]). Dentre uma série de generalizações de funções, séries e números especiais, F. H. Jackson reintroduziu o operador q-derivada[1], amplamente conhecido como derivada de Jackson, e criou o conceito de q-integral definida (veja [2, 3, 4, 5]).

Seja $f(x)$ uma função diferenciável de uma única variável. A derivada clássica de f em relação a x é dada por

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}, \tag{1}$$

[1] O operador q-derivada também é conhecido como operador q-diferença, derivada de Jackson ou, simplesmente, q-derivada.

em que h é um infinitésimo. Substituindo $x+h$ por qx (q é um número real diferente de 1) na última equação e desconsiderando o limite, tem-se a definição da q-derivada dada por

$$D_q f(x) = \frac{f(qx) - f(x)}{qx - x}, \tag{2}$$

para $x \neq 0$ e $q \neq 1$. No limite $q \to 1$, a q-derivada retorna à derivada clássica, ou seja,

$$\frac{df(x)}{dx} = \lim_{q \to 1} \frac{f(qx) - f(x)}{qx - x}. \tag{3}$$

Enquanto a derivada clássica avalia o quanto uma dada função $f(x)$ é sensível a pequenas *translações* em sua variável independente, a q-derivada é baseada em *deformações* na variável independente, ou seja, em vez da variável independente x ser transladada por uma quantidade h, ela é dilatada ou contraída por uma quantidade q.

2 q-gradiente

Para funções diferenciáveis de n variáveis, $f(\mathbf{x})$, se a derivada parcial clássica de primeira ordem com respeito a uma variável x_i é dada por

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h}, \tag{4}$$

então a q-derivada parcial de primeira ordem com respeito a x_i é

$$D_{q_i, x_i} f(\mathbf{x}) = \frac{f(x_1, \ldots, q_i x_i, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{q_i x_i - x_i}. \tag{5}$$

Quando $x_i = 0$ ou $q_i = 1$, $\forall i = 1, 2, \ldots, n$, a q-derivada parcial de primeira ordem é dada pela derivada parcial clássica de primeira ordem

$$D_{q_i, x_i} f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_i}. \tag{6}$$

Logo, a q-derivada parcial de primeira ordem de f com respeito à variável x_i pode ser naturalmente definida como

$$D_{q_i, x_i} f(\mathbf{x}) = \begin{cases} \dfrac{f(x_1, ..., q_i x_i, ..., x_n) - f(x_1, ..., x_i, ..., x_n)}{q_i x_i - x_i}, & x_i \neq 0 \text{ e } q_i \neq 1 \\ \dfrac{\partial f(\mathbf{x})}{\partial x_i}, & x_i = 0 \text{ ou } q_i = 1 \end{cases}. \tag{7}$$

Finalmente, se o gradiente clássico de uma função objetivo de n variáveis $f(\mathbf{x})$ é o vetor das n derivadas parciais de primeira ordem de f

$$\nabla f(\mathbf{x})^\top = \left[\frac{\partial f(\mathbf{x})}{\partial x_1} \; \ldots \; \frac{\partial f(\mathbf{x})}{\partial x_i} \; \ldots \; \frac{\partial f(\mathbf{x})}{\partial x_n} \right], \tag{8}$$

então o q-gradiente é o vetor nas n q-derivadas parciais de primeira ordem de f dado por (veja [6])

$$\nabla_{\mathbf{q}} f(\mathbf{x})^\top = \left[D_{q_1, x_1} f(\mathbf{x}) \; \ldots \; D_{q_i, x_i} f(\mathbf{x}) \; \ldots \; D_{q_n, x_n} f(\mathbf{x}) \right], \tag{9}$$

em que o parâmetro q é um vetor de n posições, ou seja, $\mathbf{q} = (q_1, \ldots, q_i, \ldots, q_n)$. Quando $q_i \to 1$, $\forall i$, o vetor q-gradiente retorna ao vetor gradiente clássico, isto é,

$$\nabla_{\mathbf{q}} f(\mathbf{x})\big|_{\mathbf{q}=\mathbf{1}} = \nabla f(\mathbf{x}). \tag{10}$$

DOI: 10.5540/03.2015.003.01.0428

A Figura 1 ilustra uma interpretação geométrica do vetor gradiente clássico e do vetor q-gradiente para a função $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 2 - (e^{-x^2} + 2e^{-(x-3)^2})$. Para funções de uma única variável, a interpretação geométrica do vetor gradiente em um ponto x é simplesmente a interpretação da derivada em x. Logo, o módulo do vetor gradiente em x é dado pelo coeficiente angular da reta *tangente* à curva de f nesse ponto e a direção de máximo crescimento será à direita de x, se o coeficiente angular for positivo, ou à esquerda, se for negativo. Analogamente, o vetor q-gradiente é dado pela reta *secante* à curva de f que passa pelos pontos $(x, f(x))$ e $(qx, f(qx))$. Se o coeficiente angular da reta secante for positivo (negativo), então a direção de máximo crescimento dada pelo vetor q-gradiente será à direita (esquerda).

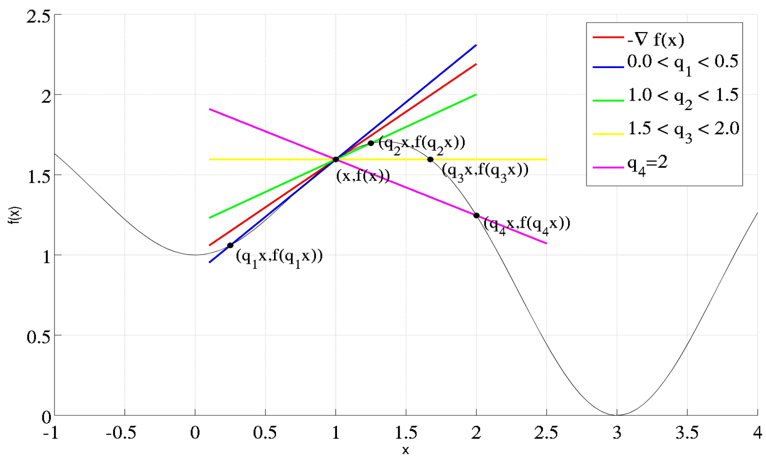

Figura 1: Interpretação geométrica do q-gradiente para uma função de uma única variável e diferentes valores do parâmetro q.

O vetor gradiente clássico (neste caso a derivada) de f em $x = 1$ é dado pela semi-reta em vermelho (Figura 1) que tem coeficiente angular positivo. Se considerarmos a minimização dessa função a partir de uma estratégia de resolução baseada na direção de gradiente como, por exemplo, o método da máxima descida, então a direção de busca contrária à direção do gradiente clássico é à esquerda de x, direção do mínimo local dessa função. O vetor q-gradiente, por sua vez, pode ser dado pelas demais semi-retas que passam por $(x, f(x))$ e $(q_i x, f(q_i x))$ $(i = 1, 2, 3, 4)$. Note que o coeficiente angular de cada semi-reta pode ser positivo ou negativo dependendo do parâmetro q_i utilizado. Para $q_4 = 2$, por exemplo, o coeficiente angular da q-derivada é negativo em $x = 1$ (semi-reta em magenta) e a direção contrária à direção do q-gradiente é à direita de x, ou seja, a direção do mínimo global dessa função. Note que existem valores de q_i para os quais o coeficiente angular do q-gradiente ou é nulo ou é positivo.

A Figura 2 ilustra a direção contrária à direção do vetor gradiente clássico, direção de máxima descida (reta em vermelho), e a direção contrária à direção do vetor q-gradiente para diferentes valores do parâmetro $\mathbf{q}$ (semi-retas em preto), juntamente com as curvas de nível de uma função bidimensional $(n = 2)$ $f : \mathbb{R}^n \to \mathbb{R}$ dada por[2]

$$f(\mathbf{x}) = \sum_{i=1}^{n} (x_i - o_i)^2, \tag{11}$$

com $\mathbf{o} = (10, 10)$ e $\mathbf{x} = (12, 12)$. Para traçar as semi-retas numeradas de 1 a 12 na Figura 2, foram tomados $\mathbf{q}$ simétricos em relação a $\mathbf{1}$ e, respectivamente, iguais a $(0.5, 1.5)$, $(0.6, 1.4)$, $(0.7, 1.3)$, $(0.8, 1.2)$, $(0.9, 1.1)$, $(0.95, 1.05)$, $(0.96, 1.04)$, $(0.97, 1.03)$, $(0.98, 1.02)$, $(0.99, 1.01)$ e $(0.999, 1.001)$. Já para a semi-reta 12 tem-se $\mathbf{q} = (-10, -3)$. À medida que os valores de $\mathbf{q} = (q_1, q_2)$ se aproximam de 1, pela esquerda para a coordenada q_1 e pela direita para a

[2] Essa função nada mais é que a função quadrática $f(\mathbf{x}) = \sum_{i=1}^{n} x_i^2$ deslocada e com mínimo global em $\mathbf{o}$.

DOI: 10.5540/03.2015.003.01.0428

coordenada q_2, a direção contrária à direção do vetor q-gradiente se aproxima da direção de máxima descida, como mostram as semi-retas de 1 a 11. Note que a semi-reta 11 e a semi-reta determinada pelo gradiente clássico (reta vermelha) são praticamente coincidentes. Finalmente, a semi-reta 12 mostra que a direção contrária à direção do q-gradiente pode ser uma direção de subida.

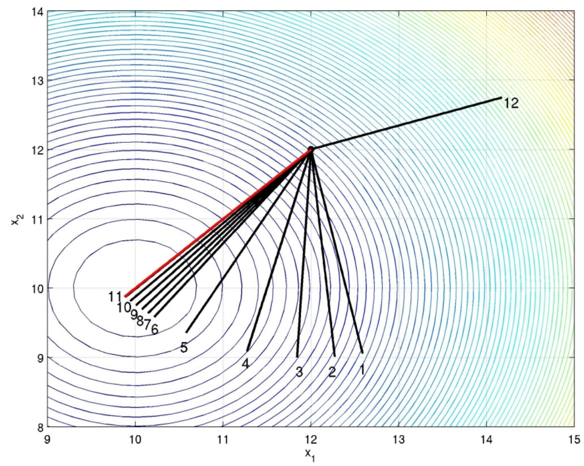

Figura 2: Direções contrárias às direções do gradiente clássico (vermelho) e do q-gradiente (preto) para diferentes valores do parâmetro $\mathbf{q}$.

Para funções unimodais cujas curvas de nível formam elipsóides alongados, o uso de uma direção de busca que é de descida, mas não exatamente de máxima descida, pode reduzir o movimento de ziguezague durante a busca pelo mínimo global e acelerar a convergência do método. Para funções multimodais, direções de busca que não são de descida podem permitir que um algoritmo escape de mínimos locais e possivelmente caminhe na direção do mínimo global da função. É fácil observar que o parâmetro $\mathbf{q}$ é um parâmetro chave para métodos baseados em q-gradiente. Quando $\mathbf{q}$ é diferente e não muito próximo de $\mathbf{1}$, a direção de busca dada pela direção contrária à direção do vetor q-gradiente pode ser tanto de descida quanto de subida, e métodos baseados nessa direção podem explorar o espaço de busca e escapar de mínimos locais, uma vez que é possível caminhar em direções de subida. Já quando $\mathbf{q}$ tende a $\mathbf{1}$, a direção de busca dada pelo vetor q-gradiente tende à direção de máxima descida, e os métodos baseados em q-gradiente retomam suas respectivas versões clássicas e realizam buscas locais. Um método baseado em q-gradiente pode ser mais estocástico ou mais determinístico, dependendo dos valores do parâmetro $\mathbf{q}$ utilizados ao longo do procedimento iterativo.

Esses exemplos simples mostram que o uso do vetor q-gradiente como direção de busca em métodos de otimização oferece mecanismos para escapar de mínimos locais. Além disso, a transição entre busca global e busca local pode ser controlada pelo parâmetro $\mathbf{q}$, desde que estratégias adequadas para a sua obtenção e a determinação do tamanho do passo sejam incorporadas ao algoritmo de otimização.

3 Algoritmo geral para métodos baseados em q-gradiente

Uma estratégia geral para a resolução de métodos de otimização irrestrita é dada pelo procedimento iterativo $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$, em que $\mathbf{x}^k \in \mathbb{R}^n$, $\mathbf{d}^k \in \mathbb{R}^n$ é a direção de busca e $\alpha^k \in \mathbb{R}$ é o tamanho do passo (distância percorrida ao longo da direção de busca $\mathbf{d}^k$ na iteração k). Observe que uma sequência de pontos $\{\mathbf{x}^k\}$ é gerada a partir de um ponto inicial $\mathbf{x}^0$.

É intuitivo pensar na direção de busca em um método de otimização como a direção contrária à direção do vetor gradiente da função objetivo no ponto em que se está na busca, pois o vetor

DOI: 10.5540/03.2015.003.01.0428

gradiente fornece a direção de máximo crescimento da função e, consequentemente, sua direção contrária fornece a direção de máximo decrescimento. Logo, para os métodos de otimização baseados no conceito de q-gradiente, deve-se substituir o vetor gradiente da função objetivo pelo vetor q-gradiente e utilizar a sua direção contrária na determinação da direção de busca.

Um algoritmo geral para métodos baseados em q-gradiente para funções diferenciáveis $f : \mathbb{R}^n \to \mathbb{R}$ é apresentado a seguir. O ponto inicial $\mathbf{x}^0 \in \mathbb{R}^n$ pode ser obtido por meio de sorteio uniforme no espaço de busca. O algoritmo retorna $\mathbf{x}_{melhor} \in \mathbb{R}^n$ como o menor valor de f ao longo do procedimento iterativo, isto é, $f(\mathbf{x}_{melhor}) < f(\mathbf{x}^k),\ \forall k$.

Algoritmo (geral) para métodos baseados em q-gradiente

Dados $f(\mathbf{x})$ contínua e diferenciável com $\mathbf{x} \in \mathbb{R}^n$ e um ponto inicial $\mathbf{x}^0$
1: **Faça** $k = 0$
2: **Faça** $\mathbf{x}_{melhor} = \mathbf{x}^0$
3: **Enquanto** não atingir um critério de parada, **faça**
4: Obtenha o parâmetro $\mathbf{q}^k$
5: Obtenha a direção de busca $\mathbf{d}^k$ com base no vetor q-gradiente (- $\nabla_{\mathbf{q}} f(\mathbf{x}^k)$)
6: Obtenha o tamanho do passo α^k
7: $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$
8: **Se** $f(\mathbf{x}^{k+1}) < f(\mathbf{x}_{melhor})$ **então** $\mathbf{x}_{melhor} = \mathbf{x}^{k+1}$
9: $k = k + 1$
10: **Retorne** $\mathbf{x}_{melhor}$

Regras para satisfazer o passo 3 do algoritmo podem se basear no número máximo de avaliações da função objetivo, distância entre o ponto atual da busca e o ponto anterior, diferença entre o valor da função objetivo no ponto atual da busca e no mínimo global (veja [7]) ou, ainda, pode-se utilizar como critério de parada a expressão $||\nabla_{\mathbf{q}} f(\mathbf{x}^k)|| \leq \varepsilon$, para $\varepsilon > 0$ pequeno, desde que $q \to 1$ e o vetor q-gradiente tenha convergido para o vetor gradiente clássico pois, como visto na Figura 1, existem valores para o parâmetro $\mathbf{q}^k$ e o ponto $\mathbf{x}^k$ em que o q-gradiente é nulo, embora $\mathbf{x}^k$ não represente necessariamente um ponto estacionário da função objetivo. Logo, quando $q \to 1$, métodos baseados em q-gradiente retomam suas versões clássicas e a condição necessária de otimalidade de primeira ordem pode ser aplicada.

O uso do q-gradiente na determinação da direção de busca requer o cálculo dos parâmetros q_i $(i = 1, 2, \ldots, n)$. Uma estratégia simples consiste na geração de números aleatórios segundo uma distribuição de probabilidade que pode ser uniforme, log-normal ou gaussiana (veja [6]). Parâmetros q_i quaisquer têm uma maior probabilidade de gerarem uma direção busca, baseada no vetor q-gradiente, tanto de descida quanto de subida, enquanto q_i próximos de 1 tendem a gerar uma direção de busca próxima ou igual à direção de máxima descida. Como o objetivo é realizar uma transição entre busca global e busca local, as estratégias propostas em [6] utilizam uma distribuição de probabilidade com desvio padrão σ variável, ou seja, inicialmente grande e que tende a zero ao longo do procedimento iterativo por meio da expressão $\sigma^{k+1} = \beta\sigma^k$, em que $0 < \beta < 1$ é o fator de redução. Dessa forma, a influência do desvio padrão sobre o comportamento de métodos baseados em q-gradiente é similar à da temperatura no Recozimento Simulado, ou seja, valores altos de σ implicam em uma maior probabilidade de geração de direções de busca de subida e, consequentemente, uma maior exploração do espaço de busca, enquanto valores baixos de σ implicam em direções de descida e uma intensificação da busca.

Uma vez definida a direção, é preciso saber o quanto caminhar nela. Em [6] também são descritas estratégias para obtenção do tamanho do passo α, dentre elas a busca linear via seção áurea e a de um passo que é reduzido sistematicamente ao longo do procedimento iterativo. São as estratégias de obtenção do parâmetro q e do tamanho do passo que, combinadas, completam o algoritmo para métodos de otimização baseados em q-gradiente.

DOI: 10.5540/03.2015.003.01.0428

4 Comportamento de métodos baseados em q-gradiente

Com o objetivo de ilustrar o comportamento de métodos baseados em q-gradiente, considere a função unimodal Rosenbrock e a multimodal Rastrigin para o caso bidimensional e com mínimo global, respectivamente, localizado em $\mathbf{x} = (1, 1)$ e $\mathbf{x} = (0, 0)$. A Figura 3 ilustra o comportamento do método da máxima descida (trajetória em azul) e de um método baseado em q-gradiente (trajetória em vermelho) em que a direção de busca é a direção contrária à direção do vetor q-gradiente para q_i gerados segundo uma distribuição de probabilidade log-normal com desvio padrão variável e tamanho do passo obtido via seção áurea (veja [8]).

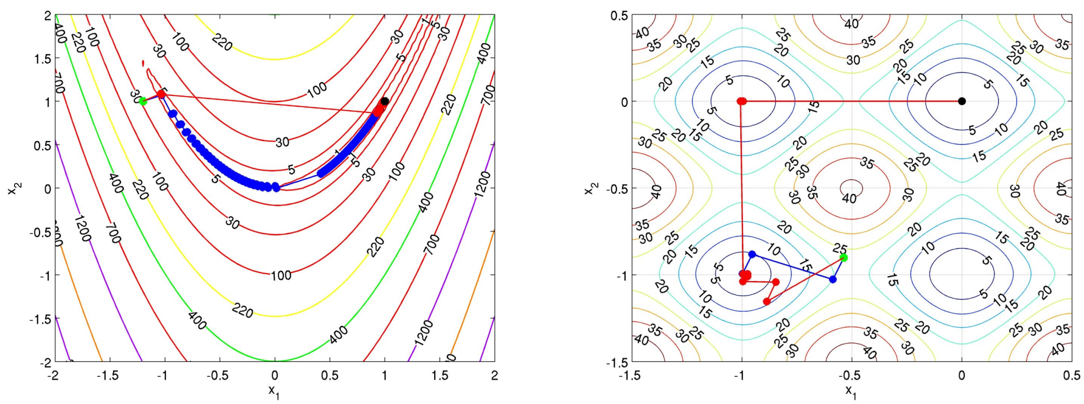

Figura 3: Trajetórias do método da máxima descida (azul) e de método baseado em q-gradiente (vermelho) para a função unimodal Rosenbrock (esquerda) e a função multimodal Rastrigin (direita). Em verde tem-se os pontos iniciais das buscas e em preto os pontos de mínimo de cada função.

A função Rosenbrock é conhecida por possuir o ponto de mínimo localizado em um vale longo, estreito e parabólico aplanado, o que dificulta a convergência dos metódos de otimização, sobretudo métodos baseados em gradiente clássico. Note, na Figura 1 (esquerda), que, enquanto o método da máxima descida apresenta uma trajetória em ziguezague (azul), o método baseado em q-gradiente (vermelho) é capaz de se mover em direções que não são de máxima descida e se aproxima do mínimo dessa função em um número menor de iterações.

A função Rastrigin é multimodal e possui um envoltório parabólico com múltiplos mínimos locais. Como esperado, o método da máxima descida converge para o mínimo local mais próximo do ponto inicial da busca. Já o método baseado em q-gradiente explora dois mínimos locais antes de atingir o mínimo global desta função.

5 Considerações Finais

Este trabalho mostra como o q-gradiente, uma generalização do vetor gradiente clássico, apresenta potencial de aplicação na resolução de problemas de otimização global contínua. A combinação dessa nova direção de busca com estratégias adequadas para a obtenção do parâmetro q e do tamanho do passo mostram que métodos baseados em q-gradiente realizam uma busca global no início do procedimento iterativo e uma busca local no final, com a presença de mecanismos que permitem que esses métodos escapem de mínimos locais e caminhem, a cada iteração, na direção do mínimo global.

Trabalhos em andamento incluem o método do q-gradiente ou, simplesmente, método q-G, uma generalização do método da máxima descida em que a direção de busca é a direção contrária à direção do vetor q-gradiente e, também, o método do q-gradiente conjugado, ou método q-GC, uma generalização do método dos gradientes conjugados em que a direção de busca é uma combinação linear de direções contrárias às direções de q-gradiente. Essas q-versões vêm sendo sistematicamente comparadas com outros algoritmos em problemas multidimensionais e multimodais com bons resultados. Ambas as versões se baseiam no algoritmo geral apresentado neste trabalho.

Referências

[1] V. Kac e P. Cheung, "Quantum calculus", Springer-Verlag, New York, 2002.

[2] F.H. Jackson, A generalization of the functions $\Gamma(n)$ and x^n, *Proc. Roy Soc. London*, (1904) 64-72.

[3] F.H. Jackson, On q-functions and a certain difference operator, *Trans. Roy Soc. Edin.*, 46 (1908) 253–281.

[4] F.H. Jackson, On q-definite integrals, *Quart. J. Pure and Appl. Math.*, 41 (1910) 193–203.

[5] F.H. Jackson, q-Difference Equations, *American Journal of Mathematics*, 32 (1910) 307–314.

[6] A.C. Soterroni, "O método do q-gradiente para otimização global", Tese de Doutorado em Computação Aplicada, INPE-São José dos Campos, 2012.

[7] A. Izmailov e M. Solodov, "Otimização, Volume 2: Métodos computacionais", IMPA, Rio de Janeiro, 2007.

[8] A.C. Soterroni, R.L. Galski e F.M. Ramos, The q-gradient vector for unconstrained continuous optimization problems, em "Operations Research Proceedings 2010" (B. Hu, K. Morash, S. Pickl e M. Siegle, eds.), pp. 365–370, Springer-Verlag Berlin Heidelberg, 2011.

