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

Método q-G: uma generalização do método da máxima descida

Aline C. Soterroni, **Fernando Manuel 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,

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

Resumo: *Este trabalho apresenta uma generalização do método da máxima descida, com base no conceito de q-derivada proveniente do q-cálculo, denominada método do q-gradiente ou método q-G. O q-cálculo surgiu da generalização de expressões matemáticas por meio de um parâmetro multiplicativo q dando origem a q-versões de funções, séries, operadores e números especiais, que no limite,* $q \to 1$*, retomam as suas respectivas versões clássicas. A principal ideia deste novo método é o uso da direção contrária à direção do vetor q-gradiente como direção de busca em problemas de otimização global contínua. O desempenho do método q-G foi avaliado através de um conjunto de seis funções teste da literatura e obteve bons resultados, sobretudo em funções multimodais.*

Palavras-chave: *q-cálculo, q-gradiente, otimização global, método q-G*

1 Introdução

O q-cálculo surgiu da generalização de expressões matemáticas por meio de um parâmetro multiplicativo q. Essas generalizações deram origem a versões análogas, também chamadas de q-versões de funções, séries, operadores e números especiais. No início do século XX, o reverendo inglês Frank Hilton Jackson desenvolveu o q-cálculo de forma sistemática. Em particular, generalizou os conceitos de derivada e integral no contexto do q-cálculo e reintroduziu a q-derivada, também conhecida como derivada de Jackson [3, 4, 5].

Neste trabalho estende-se a aplicabilidade do q-cálculo na área de otimização por meio do uso do vetor q-gradiente no método da máxima descida. Para funções contínuas e irrestritas de n variáveis, a q-versão do método da máxima descida, o método q-G, utiliza como direção de busca a direção contrária à direção do vetor q-gradiente. O vetor q-gradiente por sua vez é o vetor das q-derivadas parciais de primeira ordem da função objetivo. O método q-G retorna para a sua versão clássica à medida que o parâmetro q tende a 1. Quando $q \neq 1$ a direção de busca do método q-G pode ser tanto de descida quanto de subida, característica que permite ao método escapar de mínimos locais. Utilizando estratégias simples para o cálculo do parâmetro q, necessário para o cálculo do q-gradiente, e para o tamanho do passo dado na direção de busca,

DOI: 10.5540/03.2015.003.01.0427

tem-se um algoritmo estocástico para o método q-G destinado a problemas de otimização global contínua.

Para testar o desempenho deste novo método foi considerado um conjunto de seis funções teste unimodais e multimodais com 20 dimensões. Os resultados foram comparados com três Algoritmos Genéticos (AGs) desenvolvidos por [2] e [1].

2 O método q-G

Dada uma função $f(\mathbf{x})$ continuamente diferenciável de n variáveis, o gradiente de f é o vetor de n derivadas parciais de primeira ordem da função. Similarmente, o q-gradiente é um vetor das n q-derivadas parciais de primeira ordem de f e o parâmetro $\mathbf{q}$ é um vetor de n variáveis.

A q-derivada parcial de primeira ordem com relação à variável x_i é dada por [6]

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

A partir da Equação (1) define-se o vetor q-gradiente de f

$$\nabla_q f(\mathbf{x})^T = [D_{q_1,x_1} f(\mathbf{x}) \ldots D_{q_i,x_i} f(\mathbf{x}) \ldots D_{q_n,x_n} f(\mathbf{x})], \tag{2}$$

e quando $q_i \to 1, \forall i = 1,\ldots,n$, o vetor q-gradiente retorna ao vetor gradiente clássico.

Uma estratégia de otimização pode ser definida por um procedimento iterativo que, a partir de um ponto $\mathbf{x}^0$, gera uma sequência de soluções possíveis por meio da expressão $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha \mathbf{d}^k$, em que $\mathbf{d}^k$ é a direção de busca e α^k é o tamanho do passo dado nesta direção na iteração k.

Métodos de otimização baseados em gradiente são caracterizados pelo cálculo da direção de busca e do tamanho do passo. O método da máxima descida, por exemplo, utiliza $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$ e o tamanho do passo α^k é determinado por uma técnica de busca linear que minimiza a função objetivo ao longo da direção $\mathbf{d}^k$. O método q-G utiliza $\mathbf{d}^k = -\nabla_q f(\mathbf{x}^k)$ como direção de busca com os valores do parâmetro $\mathbf{q}$ obtidos por meio da geração de $q_i^k x_i^k$'s $(i = 1,\ldots,n)$ de acordo com uma distribuição gaussiana, com média no ponto x_i^k e desvio padrão σ^k. Note que os valores de $q_i^k x_i^k$ são utilizados diretamente no cálculdo do q-gradiente na iteração k. O desvio padrão é inicialmente diferente de zero e tende a zero ao longo do procedimento iterativo por meio da expressão $\sigma^{k+1} = \beta \cdot \sigma^k$, em que $\beta \in (0,1)$ é o fator de redução. Quando $\sigma^k \neq 0$, os valores de q_i^k obtidos da geração de $q_i^k x_i^k$ podem ser quaisquer números reais e, associados aos valores da função objetivo, permitem que a direção de busca seja tanto de subida quanto de descida. Quando $\sigma^k \to 0$, tem-se $q_i^k x_i^k \to x_i^k$, logo $q_i^k \to 1$, $\forall i$. Quando os q_i^k's são numericamente iguais a 1, $\forall i$ $(i = 1, 2, \ldots, n)$, a direção de busca é dada pelo vetor gradiente clássico e o método q-G retorna ao método da máxima descida.

Para determinar o tamanho do passo α^k foi utilizada a estratégia de controle do passo por redução, em que um valor inicial para o tamanho do passo (α^0) é reduzido a cada iteração por um fator de redução $0 < \beta < 1$, ou seja, $\alpha^{k+1} = \beta \cdot \alpha^k$. Por simplicidade, considerou-se o mesmo fator de redução β utilizado na redução do desvio-padrão σ, responsável pela geração do parâmetro $\mathbf{q}$. A redução do tamanho inicial do passo α^0 ao longo das iterações permite passos grandes no início do procedimento iterativo (quando a busca é global, o desvio-padrão inicial σ^0 é diferente de zero e as direções de busca são quaisquer) e passos pequenos e tendendo a zero no final (quando a busca é local, com $\sigma^k \to 0$ e as direções de busca são de descida ou de máxima descida).

Um algoritmo para método q-G é dado a seguir. O ponto $\mathbf{x}^0$ é obtido de forma aleatória e o critério de parada pode ser dado pelo número de avaliações da função objetivo (NAFO) ou pela precisão desejada. O ponto de mínimo será o ponto $\mathbf{x}^k$ que atingir o menor valor da função objetivo ao longo da busca.

Algoritmo do Método q-G

Dados $f(\mathbf{x})$ contínua e diferenciável com $\mathbf{x} \in \mathbb{R}^n$, um ponto inicial $\mathbf{x}^0$ e os parâmetros livres σ^0, α^0 e $0 < \beta < 1$

1: **Faça** $k = 0$
2: **Faça** $\mathbf{x}_{melhor} = \mathbf{x}^0$
3: **Enquanto** não atingir um critério de parada, **faça**
4: Obtenha $\mathbf{q}^k$ segundo uma distribuição gaussiana com $\mu = \mathbf{x}^k$ e σ^k
5: $\mathbf{d}^k = -\nabla_{\mathbf{q}} f(\mathbf{x}^k)/|\nabla_{\mathbf{q}} f(\mathbf{x}^k)|$
6: $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$
7: **Se** $f(\mathbf{x}^{k+1}) < f(\mathbf{x}_{melhor})$ **então** $\mathbf{x}_{melhor} = \mathbf{x}^{k+1}$
8: $\sigma^{k+1} = \beta \cdot \sigma^k$
9: $\alpha^{k+1} = \beta \cdot \alpha^k$
10: $k = k + 1$
11: **Retorne** $\mathbf{x}_{melhor}$

O algoritmo do método q-G é de fácil implementação e possui apenas três parâmetros de ajuste.

3 Experimentos Computacionais

O conjunto de funções teste utilizado para verificar a performance do método q-G corresponde a três funções unimodais (*Ellipsoidal*, Schwefel e Rosenbrock) e três funções multimodais (Ackley, Rastrigin e Rastrigin Rotacionada) definidas em [2, 1]. Todas as funções são contínuas, não-lineares e de 20 variáveis e têm como mínimo global $f(\mathbf{x}^*) = 0$ em $\mathbf{x}^* = \mathbf{0}$, exceto para a Rosenbrock em que $\mathbf{x}^* = \mathbf{1}$. Além de serem utilizadas com frequência na área de otimização contínua, permitindo a comparação direta com outros algoritmos, esse conjunto de funções teste possui características que tornam a sua minimização difícil e que estão presentes em muitos problemas reais [1].

O método q-G é comparado com os Algoritmos Genéticos (AGs) G3-PCX, SPC-vSBX e SPC-PNX. O G3-PCX, desenvolvido por [2], é uma versão de AG com codificação real com bons resultados para as funções unimodais *Ellipsoidal*, Schwefel e Rosenbrock. Os resultados do G3-PCX serviram de referência para os AGs com codificação real SPC-vSBX e SPC-PNX, desenvolvidos por [1]. Os resultados do G3-PCX para as funções Ackley e Rastrigin Rotacionada foram obtidos por [1] a partir do código do G3-PCX.

Os resultados do método q-G foram obtidos para o melhor ajuste de parâmetros encontrados sobre cada função teste. Os melhores valores para σ^0, α^0 e β estão ilustrados na Tabela 1. Os valores de σ^0 e α^0 foram normalizados pelo maior comprimento linear do espaço de busca, $L = \sqrt{\sum_{i=1}^{n}(\mathbf{x}_{max_i} - \mathbf{x}_{min_i})^2}$, para os valores de $\mathbf{x}_{max_i}$ e $\mathbf{x}_{min_i}$ definidos [1].

Funções	$\sigma^{\mathbf{0}}/L$	$\alpha^{\mathbf{0}}/L$	β
Ellipsoidal	$0,0045$	$0,42$	$0,8600$
Schwefel	$0,0011$	$0,011$	$0,9970$
Rosenbrock	$0,0055$	$0,0055$	$0,9995$
Ackley	$0,075$	$0,045$	$0,9000$
Rastrigin	$0,46$	$0,0066$	$0,9995$
Rastrigin Rotacionada	$0,66$	$0,011$	$0,9990$

Tabela 1: Parâmetros usados pelo método q-G para o primeiro conjunto de funções teste.

Foram realizadas 50 execuções independentes do algoritmo do método q-G para 50 pontos iniciais distintos. Assim como em [2] e [1], o intervalo de inicialização utilizado pelo método q-G foi $[-10; -5]$. Os critérios de parada utilizados por todos os algoritmos foram: (i) atingir a precisão desejada 10^{-20}, ou seja, o procedimento iterativo termina quando o erro da função é menor que a precisão desejada ($f(\mathbf{x}_{melhor}) - f(\mathbf{x}^*) \leq 10^{-20}$); ou (ii) atingir um NAFO máximo, isto é, o algoritmo para se atingir 10^6 avaliações da função objetivo.

As Tabelas 2 e 3 ilustram uma comparação entre os algoritmos sobre o conjunto de funções teste. Cada execução independente que atingir a precisão desejada 10^{-20} antes de completar as 10^6 avaliações da função objetivo tem o NAFO armazenado. Logo, apenas os valores de NAFO das execuções de sucesso são ordenados e exibidos nas colunas "Melhor" (menor), "Mediano" e "Pior" (maior). Se a precisão desejada não é atingida antes de 10^6 avaliações, então o melhor valor da função objetivo encontrado dentre todas as execuções independentes é informado na coluna "$f(\mathbf{x}_{melhor})$". A coluna "Sucesso" exibe o número de execuções independentes que atingiram a precisão desejada (para as funções unimodais) ou que alcançaram a bacia de atração do mínimo global (para as funções multimodais). Os melhores valores estão em negrito.

Função	Algoritmo	Melhor	Mediano	Pior	$f(\mathbf{x}_{melhor})$	Sucesso
Ellipsoidal	**G3-PCX**	**5.826**	**6.800**	**7.728**	$\mathbf{10^{-20}}$	**10/10**
	SPC-vSBX	49.084	50.952	57.479	10^{-20}	10/10
	SPC-PNX	36.360	39.360	40.905	10^{-20}	10/10
	q-G	**5.905**	**7.053**	**7.381**	$\mathbf{10^{-20}}$	**50/50**
Schwefel	**G3-PCX**	**13.988**	**15.602**	**17.188**	$\mathbf{10^{-20}}$	**10/10**
	SPC-vSBX	260.442	294.231	334.743	10^{-20}	10/10
	SPC-PNX	236.342	283.321	299.301	10^{-20}	10/10
	q-G	289.174	296.103	299.178	10^{-20}	50/50
Rosenbrock	**G3-PCX**	**16.508**	**21.452**	**25.520**	$\mathbf{10^{-20}}$	**36/50**
	SPC-vSBX	10^6	-	-	10^{-4}	48/50
	SPC-PNX	10^6	-	-	10^{-10}	38/50
	q-G	10^6	-	-	10^{-10}	50/50

Tabela 2: Desempenho comparativo, em termos no número de avaliações da função objetivo (NAFO), entre o método q-G e os AGs G3-PCX, SPC-vSBX e SPC-PNX para as funções unimodais *Ellipsoidal* e Schwefel, e a função Rosenbrock.

Na Tabela 2, para a função *Ellipsoidal* todos os algoritmos atingiram a precisão desejada (10^{-20}) em todas as execuções independentes. Porém, o método q-G e o G3-PCX satisfazem esse critério de parada para os menores números de avaliações da função objetivo e com desempenho similar. Para a função Schwefel, embora o método q-G alcance a precisão desejada em todas as execuções, o G3-PCX apresenta o melhor desempenho. Por fim, para a função Rosenbrock, conhecida por ser um exemplo de convergência prematura para métodos de otimização, o único algoritmo que atinge a precisão desejada é o G3-PCX. O método q-G fica apenas atrás do G3-PCX, atingindo a segunda melhor precisão (10^{-10}) em mais execuções independentes em comparação com os demais algoritmos.

Na Tabela 3, para a função Ackley a precisão desejada é 10^{-15} para o método q-G e 10^{-10} para os demais algoritmos[1]. Todos os algoritmos, exceto o G3-PCX, atingiram a bacia de atração do mínimo global e a precisão desejada de acordo com seus limites computacionais. No entanto, o método q-G satisfaz essa condição com o menor número de avaliações da função objetivo. Para a função Rastrigin, novamente o G3-PCX e agora o SPC-PNX não alcançaram a bacia de

[1] A função Ackley avaliada em $\mathbf{x}^* = \mathbf{0}$ com precisão dupla não é zero. Para o compilador utilizado neste trabalho, tem-se $f(\mathbf{x}^*) = -0.444089209850062\text{E} - 15$.

Função	Algoritmo	Melhor	Mediano	Pior	$f(\mathbf{x}_{melhor})$	Sucesso
Ackley	G3-PCX	10^6	-	-	3.959	0
	SPC-vSBX	57.463	63.899	65.902	10^{-10}	10/10
	SPC-PNX	45.736	48.095	49.392	10^{-10}	10/10
	q-G	**11.850**	**12.465**	**13.039**	$\mathbf{10^{-15}}$	**50/50**
Rastrigin	G3-PCX	10^6	-	-	15.936	0
	SPC-vSBX	260.685	306.819	418.482	10^{-20}	6/10
	SPC-PNX	10^6	-	-	4.975	0
	q-G	**676.050**	**692.450**	**705.037**	$\mathbf{10^{-20}}$	**48/50**
	G3-PCX	10^6	-	-	$309,429$	0
Rastrigin	SPC-vSBX	10^6	-	-	$8,955$	0
Rotacionada	SPC-PNX	10^6	-	-	$3,980$	0
	q-G	**541.857**	**545.957**	**549.114**	$\mathbf{10^{-20}}$	**20/50**

Tabela 3: Desempenho comparativo, em termos no número de avaliações da função objetivo (NAFO), entre o método q-G e os AGs G3-PCX, SPC-vSBX e SPC-PNX para as funções multimodais Ackley, Rastrigin e Rastrigin Rotacionada.

atração do mínimo global em nenhuma execução. O método q-G alcançou a bacia de atração do mínimo global em 96% das execuções. Finalmente, para a função Rastrigin Rotacionada o método q-G foi o único a atingir a bacia de atração do mínimo global, e fez isso em 40% das execuções.

Comparando os resultados para as funções unimodais ou quase unimodais (*Ellipsoidal*, Schwefel e Rosenbrock) o escore é favorável ao G3-PCX, com duas vitórias e um empate. Se o algoritmo G3-PCX não estivesse na competição, o método q-G seria o vencedor no caso unimodal com duas vitórias (funções *Ellipsoidal* e Rosenbrock) e um empate (com os métodos SPC-vSBX e SPC-PNX para a função Schwefel). Já para as funções multimodais (Ackley, Rastrigin e Rastrigin Rotacionada) o G3-PCX é claramente inferior. Para essas funções, o escore é favorável para o método q-G com duas vitórias (funções Ackley e Rastrigin Rotacionada).

4 Conclusões

Este trabalho apresenta uma generalização do método da máxima descida, denominada método q-G. No método q-G a direção de busca é dada pela direção contrária à direção do vetor q-gradiente. A estratégia utilizada para calcular o parâmetro $\mathbf{q}$, e consequentemente, o vetor q-gradiente, admite que a direção de busca possa ser tanto de descida quanto de subida, característica que permite ao método escapar de mínimos locais.

O método q-G foi comparado com três algoritmos genéticos G3-PCX, SPC-vSBX e SPC-PNX para um conjunto de seis funções teste. Os resultados obtidos mostram que o método q-G é competitivo em relação aos três algoritmos genéticos, sobretudo para as funções teste multimodais.

Referências

[1] P. J. Ballester and J. N. Carter, An effective real-parameter genetic algorithm with parent centric normal crossover for multimodal optimisation, em "Genetic and Evolutionary Computation Conference" pp. 901-913,Springer-Verlag, Seatle, WA, 2004.

[2] K. Deb and A. Anand and D. Joshi, A computationally efficient evolutionary algorithm for real-parameter optimization, *Evolutionary Computation*, 10 (2002) 345-369.

[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 and R. L. Galski and F. M. Ramos, The q-gradient method for global optimization, *arXiv:1209.2084*, math.OC (2012).

