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-GC: uma generalização do método dos gradientes conjugados

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,

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 .

Resumo: *Recentemente, baseado na derivada de Jackson, foi proposta uma generalização do método da máxima descida, denominada método do q-gradiente (q-G), para problemas de otimização global contínua. Dentro desse contexto, este trabalho apresenta uma generalização do método dos gradientes conjugados (q-GC) com base no conceito do vetor q-gradiente. Para avaliar o desempenho do método q-GC foram considerados os resultados obtidos pelo método q-G e por três Algoritmos Genéticos (AGs) para um conjunto de seis funções teste de* 20 *variáveis e mesmo critério de parada. No geral, os resultados mostram que o q-GC é um método promissor para solução de problemas de otimização multimodais.*

Palavras-chave: *q-gradiente, q-derivada, método q-G, método q-GC*

1 Introdução

Recentemente, uma generalização do método da máxima descida, denominada método q-G, foi desenvolvida por [10] para problemas de otimização global contínua. Esse método utiliza como direção de busca, a direção contrária à direção do vetor q-gradiente, definido a partir da derivada clássica com o auxílio do parâmetro q. Dentro desse contexto, esse trabalho apresenta uma generalização do método dos gradientes conjugados, denominada método dos q-Gradientes Conjugados (q-GC), onde a primeira direção de busca é a direção contrária à direção do vetor q-gradiente e as outras direções são combinações lineares da direção contrária à direção do vetor q-gradiente do ponto atual com as direções anteriores. Tanto no método q-G quanto no método q-GC, q é um parâmetro chave e, à medida que q tende a 1, as q-versões retomam suas respectivas versões clássicas.

Para testar essa nova abordagem, o método q-GC foi aplicado em seis funções teste (três unimodais e três multimodais) comumente utilizadas na área de otimização contínua. Os resultados são comparados com os obtidos pelo método q-G, além dos Algoritmos Genéticos (AGs) G3-PCX (veja [2]), SPC-vSBX e SPC-PNX (veja [1]). Embora preliminares, os resultados mostram a capacidade do método de escapar de mínimos locais em problemas multimodais. Já para os problemas unimodais, o q-GC permite encontrar o mínimo da função em menos iterações ou com uma precisão melhor que o método q-G.

2 O método q-GC

A derivada de Jackson, ou q-derivada, de uma função $f(x)$ de uma única variável é dada por (veja [6])

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

onde q é um número real diferente de 1 e x diferente de 0. No limite, quando $q \to 1$ (ou $x \to 0$), a q-derivada retorna à derivada clássica. Para funções diferenciáveis de n variáveis, $f(\mathbf{x})$, o q-gradiente é o vetor das n q-derivadas parciais de primeira ordem de f dado por (veja [10])

$$D_{q_i,x_i} f(\mathbf{x}) = \begin{cases} \dfrac{f(x_1, ..., q_i x_i, ..., x_n) - f(x_1, ..., x_i, ..., x_n)}{q_i x_i - x_i}, & x_i \neq 0 \ \text{ e } \ q_i \neq 1 \\ \dfrac{\partial f(\mathbf{x})}{\partial x_i}, & x_i = 0 \ \text{ ou } \ q_i = 1 \end{cases}, \tag{2}$$

onde o parâmetro q é um vetor $\mathbf{q} = (q_1, \dots, q_i, \dots, q_n)$ com $q_i \neq 1$, $\forall i$. Note que, se $x_i = 0$ ou $q_i = 1$, a q-derivada parcial de primeira ordem retorna à derivada parcial clássica. A equação anterior define o vetor q-gradiente de f como [10]

$$\nabla_{\mathbf{q}} f(\mathbf{x})^T = [D_{q_1,x_1} f(\mathbf{x}) \dots D_{q_i,x_i} f(\mathbf{x}) \dots D_{q_n,x_n} f(\mathbf{x})], \tag{3}$$

em que no limite, $q_i \to 1$ ($\forall i = 1, \dots, n$), o vetor q-gradiente retorna ao vetor gradiente clássico.

Em geral, os métodos de otimização utilizam o procedimento iterativo $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$, em que $\mathbf{d}^k$ é a direção de busca e α^k é o tamanho do passo dado nesta direção na iteração k. Os métodos de otimização baseados em gradiente diferem entre si na forma em que a direção e o tamanho do passo são calculados. O método da máxima descida, por exemplo, utiliza $\mathbf{d}^k = -\nabla f(\mathbf{x}^k)$ como direção de busca e o tamanho do passo α^k é, em geral, determinado por uma técnica de busca linear que minimiza a função objetivo ao longo de $\mathbf{d}^k$. No método q-G, a direção de busca é a direção contrária à direção do vetor q-gradiente da função objetivo no ponto x^k, $-\nabla_{\mathbf{q}} f(\mathbf{x}^k)$, como definida na Equação (3). De forma similar, o método q-GC é uma generalização do método dos gradientes conjugados de Fletcher e Reeves [4] no qual a primeira direção de busca é a direção contrária à direção do vetor q-gradiente e as outras direções são combinações lineares da direção contrária à direção do vetor q-gradiente do ponto atual com as direções anteriores.

O parâmetro q utilizado no cálculo da q-derivada pode ser qualquer número real diferente de 1. Assim como em [10], os valores de $q_i^k x_i^k$, $i = 1, ..., n$, foram sorteados segundo uma distribuição gaussiana com média no ponto x_i^k e desvio padrão σ^k. O desvio padrão é inicialmente diferente de zero e tende a zero ao longo do procedimento iterativo por meio da expressão $\sigma^{k+1} = \beta\sigma^k$ em que $\beta \in [0, 1]$ é o fator de redução.

Geralmente, os métodos baseados em gradiente realizam busca linear a cada iteração para determinar o comprimento do passo a ser dado em uma direção que é de descida (veja [3]). Uma vez que a direção de busca do método q-GC pode ser tanto de descida quanto de subida, dependendo do valor do parâmetro $\mathbf{q}$, o cálculo para o comprimento do passo na iteração k é dado por $\alpha^{k+1} = \beta\alpha^k$. Por simplicidade, β é o mesmo fator de redução usado no cálculo do σ^k (veja [10]).

As principais etapas do algoritmo do método q-GC são descritas a seguir.

DOI: 10.5540/03.2015.003.01.0429

Algoritmo do Método q-GC

Dados $f(\mathbf{x})$ contínua e diferenciável, ponto inicial $\mathbf{x}^0$ e os parâmetros livres $\sigma^0 > 0$, $\alpha^0 > 0$ e $0 < \beta < 1$

1: **Faça** $k = 0$
2: **Faça** $\mathbf{x_{melhor}} = \mathbf{x}^k$
3: Gere $q_i^k (i = 1, ..., n)$ com distribuição gaussiana para σ^k e $\mu^k = \mathbf{x}^k$
4: Calcule $\nabla_{\mathbf{q}} f(\mathbf{x^k})$
5: $\mathbf{d}^k = -\nabla_{\mathbf{q}} f(\mathbf{x^k})$
6: $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$
7: $k = k + 1$
8: **Enquanto o critério de parada não for satisfeito, faça:**
9: Gere $q_i^k (i = 1, ..., n)$ com distribuição gaussiana para σ^k e $\mu^k = \mathbf{x}^k$
10: Calcule $\nabla_{\mathbf{q}} f(\mathbf{x^k})$
11: $\delta^{k-1} = \nabla_{\mathbf{q}} f(\mathbf{x^k})^T \nabla_{\mathbf{q}} f(\mathbf{x^k}) / \nabla_{\mathbf{q}} f(\mathbf{x^{k-1}})^T \nabla_{\mathbf{q}} f(\mathbf{x^{k-1}})$
12: $\mathbf{d}^k = -\nabla_{\mathbf{q}} f(\mathbf{x^k}) + \delta^{k-1} \mathbf{d}^{k-1}$
13: $m = \nabla_{\mathbf{q}} f(\mathbf{x^k})^T \mathbf{d}^k$
14: **Se** $m \geq 0$ e $\mathbf{q} = \mathbf{1}$, **então** $\mathbf{d}^k = -\nabla_{\mathbf{q}} f(\mathbf{x}^{k-1})$
15: $\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha^k \mathbf{d}^k$
16: **Se** $f(\mathbf{x}^{k+1}) < f(\mathbf{x_{melhor}})$ **então** $\mathbf{x_{melhor}} = \mathbf{x}^{k+1}$
17: $\sigma^{k+1} = \beta \sigma^k$
18: $\alpha^{k+1} = \beta \alpha^k$
19: $k = k + 1$
20: **Retorna** $\mathbf{x_{melhor}}$

O algoritmo termina quando um critério de parada apropriado é satisfeito. Em aplicações do mundo real, onde o mínimo global não é conhecido, o critério de parada poderá ser um número máximo de avaliações da função objetivo ou o valor do gradiente da função no ponto atual $||\nabla f(\mathbf{x}^k)||$, desde que o método q-GC retome sua versão clássica no final da busca. O algoritmo retorna o melhor ponto visitado durante o procedimento iterativo como mostra o passo 20 do algoritmo. Note que as direções são reinicializadas apenas se a direção for de subida e o parâmetro $\mathbf{q}$ for igual a $\mathbf{1}$ (passo 14). O método q-GC utiliza apenas três parâmetros de ajuste: σ^0, α^0 e β. Embora uma má escolha desses parâmetros causem uma deteriorização em seu desempenho, simulações mostraram que o algoritmo do método q-GC é suficientemente robusto sendo capaz de atingir a bacia de atração do mínimo global[1]. A seguir, o método q-GC é comparado com o método q-G e com AGs considerados eficientes na resolução de problemas de otimização global.

3 Experimentos Computacionais

O desempenho do método q-GC é avaliado em seis funções teste de 20 variáveis. As funções unimodais são: *Ellipsoidal*, Schwefel e Rosenbrock; e as funções multimodais são: Ackley, Rastrigin e Rastrigin Rotacionada. A descrição completa dessas funções pode ser encontrada em [10]. Todas as funções possuem mínimo global em $f(\mathbf{x}^*) = 0$ com $\mathbf{x}^* = \mathbf{0}$, com exceção da função Rosembrock onde $\mathbf{x}^* = \mathbf{1}$. Os critérios de parada são: máximo de 10^6 avaliações da função objetivo ou $f(\mathbf{x}) < 10^{-20}$, os mesmos adotados em [1], [2] e [10]. A Tabela 1 exibe os valores dos parâmetros de ajuste σ^0, α^0 e β, usados em cada função pelo método q-GC. Os valores de σ^0 e α^0 foram normalizados pelo maior comprimento linear do espaço de busca, $L = \sqrt{\sum_{i=1}^{n} (\mathbf{x}_{max_i} - \mathbf{x}_{min_i})^2}$, com $\mathbf{x}_{max_i}$, $\mathbf{x}_{min_i}$ definidos como em [1].

Os resultados são apresentados nas Tabelas 2 e 3. As colunas “Melhor”, “Mediano” e “Pior”

[1]Bacia de atração do mínimo global $\mathbf{x}^*$, $\mathbf{x}^* \in \mathbb{R}^n$, de uma função $f(\mathbf{x})$ é a região B, $B \subset \mathbb{R}^n$, que contém o ponto $\mathbf{x}^*$ e, para todo $\mathbf{x} \in B$, uma trajetória de descida da $f(\mathbf{x})$ converge para o ponto $\mathbf{x}^*$ (veja [5]).

DOI: 10.5540/03.2015.003.01.0429

Funções	$\sigma^{\mathbf{0}}/L$	$\alpha^{\mathbf{0}}/L$	β
Ellipsoidal	$0,0012$	$0,0750$	$0,9000$
Schwefel	$0,0025$	$0,0025$	$0,9900$
Rosenbrock	$0,0001$	$0,3052$	$0,9995$
Ackley	$0,0092$	$0,0083$	$0,9000$
Rastrigin	$0,1953$	$0,0004$	$0,9995$
Rastrigin Rotacionada	$0,1074$	$0,0005$	$0,9995$

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

referem-se, respectivamente, ao menor, ao mediano e ao maior número de avaliações da função objetivo necessárias para atingir a precisão desejada de 10^{-20}. A coluna "$f(\mathbf{x_{melhor}})$" exibe o melhor valor da função objetivo encontrado durante o procedimento iterativo. Note que, quando a precisão desejada não é atingida, o melhor valor encontrado pelo método é exibido. 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	Algoritmos	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-gradiente	**5.905**	**7.053**	**7.381**	$\mathbf{10^{-20}}$	**50/50**
	q-GC	8.914	9.260	9.486	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-gradiente	289.174	296.103	299.178	10^{-20}	50/50
	q-GC	81.913	83.117	84.311	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-gradiente	10^6	-	-	10^{-10}	50/50
	q-GC	10^6	-	-	10^{-10}	50/50

Tabela 2: Comparação entre os algoritmos sobre o conjunto de funções teste unimodais.

Para as funções unimodais (Tabela 2), o método q-GC apresenta um desempenho superior ao método q-G e dos AGs SPC-vSBX e SPC-PNX, mas é superado pelo AG G3-PCX. Note que o método q-GC atinge a precisão desejada em todas as 50 execuções. Para as funções multimodais (Tabela 3), tanto o método q-GC quanto o método q-G são claramente superiores aos AGs. Porém, o método q-GC é melhor que o método q-G pois atinge a precisão desejada com um maior número de sucessos ou um menor número de avaliações da função objetivo. Para a função Rastrigin Rotacionada, por exemplo, enquanto os três AGs foram incapazes de encontrar a bacia do mínimo global, o método q-GC conseguiu atingir essa bacia em 84% das execuções, contra

DOI: 10.5540/03.2015.003.01.0429

Função	Algoritmos	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-gradiente	11.850	12.465	13.039	10^{-15}	50/50
	q-GC	**11.295**	**11.424**	**11.522**	$\mathbf{10^{-15}}$	**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-gradiente	676.050	692.450	705.037	10^{-20}	48/50
	q-GC	**778.824**	**796.662**	**814.698**	$\mathbf{10^{-20}}$	**50/50**
Rastrigin Rotacionada	G3-PCX	10^6	-	-	309.429	0
	SPC-vSBX	10^6	-	-	8.955	0
	SPC-PNX	10^6	-	-	3.980	0
	q-gradiente	541.857	545.957	549.114	10^{-20}	20/50
	q-GC	**803.485**	**824.482**	**857.909**	$\mathbf{10^{-20}}$	**42/50**

Tabela 3: Comparação entre os algoritmos sobre o conjunto de funções teste multimodais.

apenas 40% apresentado pelo q-G.

4 Conclusões

Este trabalho apresenta uma q-versão do método dos gradientes conjugados, denominada método q-GC, baseado no conceito do vetor q-gradiente desenvolvido em [10]. A principal ideia por trás do método é a transição entre busca global, no início, e busca local, no final, do procedimento iterativo, com a presença de mecanismos que permitem que o método escape de mínimos locais e caminhe cada vez mais na direção do mínimo global. O método q-GC foi comparado com o método q-G (veja [10]) e com três AGs (veja [1, 2]), para seis funções teste da literatura. Embora preliminares, o desempenho do método q-GC é promissor, especialmente quando aplicado em funções multimodais.

Referências

[1] P. J. Ballester e J. N. Carter, An effective real-parameter genetic algorithm with parent centric normal crossover for multimodal optimisation, em “Genetic and Evolutionary Computation Conference - GECCO 2004” (K. Deb, ed.) pp. 901-913, Springer-Verlag, Seatle, WA, 2004.

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

[3] G. Di Pillo e L. Palagi, Nonlinear programming: Introduction, em “Handbook of applied optimization” (P. M. Pardalos e M. G. C. Resende) pp. 263-268, Oxford University Press, New York, 2002.

[4] R. Fletcher e C. M. Reeves, Function minimization by conjugate gradients, *Computer Journal*, 07 (1964) 149-154.

[5] H. Huang, Global Optimization: Filled Function Methods, em “Encyclopedia of Optimization, Second Edition”(C. A. Floudas e P. M. Pardalos, eds.) pp. 1316-1323, Springer, New York, 2009.

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

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

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

[9] J. Nocedal e S. J. Wright, “Numerical Optimization”, Springer-Verlag, New York, 1999.

[10] A. C. Soterroni, R. L. Galski e F. M. Ramos, The q-gradient method for global optimization, *arXiv:1209.2084*, math.OC (2012).

[11] G. N. Vanderplaats, “Numerical optimization techniques for engineering design”, McGraw-Hill, New York, 1984.

