Tendências em Matemática Aplicada e Computacional, **1**, No. 2 (2000), 469-474.
Sociedade Brasileira de Matemática Aplicada e Computacional.

Métodos de Gradiente Conjugado com Busca Linear Inexata[1]

B.P.B. SILVA, C.H. dos SANTOS, J.J. ROSSETTO, M. CARDIA, N.M.P. VOLPI, Y. JIAHONG, Departamento de Matemática, UFPR, 81531-990 Curitiba, PR, Brasil.

Este trabalho é dedicado à memória da Professora Beatriz P. de Barros e Silva falecida em 3 de abril de 2000.

Resumo. Neste trabalho apresentamos uma família de métodos de gradiente conjugado. Definimos uma direção de busca que depende de dois parâmetros e nem sempre satisfaz o critério das direções conjugadas, mas que, de certa forma, é uma aproximação da direção de Newton. Com uma escolha conveniente dos parâmetros, obtemos alguns dos métodos conhecidos na literatura. Testes computacionais foram realizados para comparar os novos métodos obtidos.

1. Introdução

Os métodos de gradiente conjugado têm despertado muito interesse para a resolução de problemas de minimização sem restrições de grande porte, devido a sua grande vantagem de não precisar guardar qualquer matriz, obtendo com isso um baixo custo computacional. Recentemente, vários autores têm considerado direções de busca que não satisfazem o critério das direções conjugadas, optando por aproximá-las, de alguma forma, da direção de Newton e em seguida fazendo uma busca linear inexata. Uma vasta literatura pode ser encontrada a respeito desse assunto, muitas das quais serão mencionadas no decorrer deste trabalho.

Seja o problema

$$\underset{x \in \mathbb{R}^n}{\text{minimizar}}\, f(x), \tag{1.1}$$

onde $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ é uma função com derivadas parciais contínuas.

Os métodos considerados neste trabalho geram os iterados

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

$$d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k, \tag{1.3}$$

[1]Apoio financeiro da FUNPAR

onde g_k denota $g(x_k) = \nabla f(x_k)$ e

$$\beta_k = \frac{(\theta_k y_k - \frac{1}{s_k} p_k)^t g_{k+1}}{y_k^t d_k} \tag{1.4}$$

com $y_k = g_{k+1} - g_k$ e $p_k = x_{k+1} - x_k = \alpha_k d_k$. Observemos que com uma escolha conveniente dos escalares θ_k e s_k obtemos alguns dos métodos conhecidos na literatura. Seja $\theta_k = 1$. Se $s_k = 1$ então, substituindo (1.4) em (1.3), temos a direção apresentada por Perry em [8]. Se $s_k \longrightarrow \infty$, temos a direção de Hestenes-Stiefel [7]. Se a busca linear for exata, temos a direção de Polak-Ribière [9] e, se os sucessivos gradientes são ortogonais, temos a direção de Fletcher-Reeves [6]. Agora, se

$$\theta_k = \frac{p_k^t p_k}{p_k^t y_k} \tag{1.5}$$

e se $s_k = 1$ então, (1.4) é a fórmula de Birgin-Martinez [2], que foi fortemente motivada pelos trabalhos de Barzilai-Borwein [1] e Raydan [10] e deu bons resultados computacionais.

Motivados pelo sucesso de Birgin-Martinez [2], considerando os bons resultados obtidos por Sherali-Ulular [13], que exigem que a direção de Perry [8] satisfaça equação

$$s_k d_{k+1} = -H_{k+1}^{-1} g_{k+1},$$

onde H_{k+1} é a hessiana de f em x_{k+1}, gerando $\beta_k = (y_k - \frac{1}{s_k} p_k)^t g_{k+1} / y_k^t d_k$ e inspirados no trabalho de Dai-Yuan-Yuan [3] que é uma generalização, sob o ponto de vista de interpolação, dos trabalhos de Barzilai-Borwein [1] e Raydan [10], tomaremos outros valores para θ_k e s_k. Seja θ_k dado por (1.5) e consideremos

$$t_k = \theta_k^{-1} = \frac{p_k^t y_k}{p_k^t p_k}. \tag{1.6}$$

O modelo quadrático

$$q_{k+1} = f_{k+1} + \alpha g_{k+1}^t p_k + \frac{1}{2} t_k \alpha^2 \|p_k\|^2, \tag{1.7}$$

onde $f_{k+1} = f(x_{k+1})$, é uma aproximação para $f(x_{k+1} + \alpha p_k)$. É fácil ver que t_k dado por (1.6) satisfaz a condição de interpolação

$$\nabla q_{k+1}(-1) = g_k^t p_k. \tag{1.8}$$

Portanto, θ_k dado por (1.5) é uma aproximação para a inversa da derivada direcional segunda de f em x_{k+1} ao longo de p_k.

Para qualquer $t_k \in I\!R$, o modelo quadrático (1.7) satisfaz as condições de interpolação

$$q_{k+1}(0) = f_{k+1} \tag{1.9}$$

e

$$\nabla q_{k+1}(0) = g_{k+1}^t p_k, \tag{1.10}$$

Se (1.8) é substituída por uma outra condição de interpolação

$$q_{k+1}(-1) = f_k, \tag{1.11}$$

então podemos obter de (1.7) que

$$t_k = \frac{2(f_k - f_{k+1} + g_{k+1}^t p_k)}{\|p_k\|^2} \tag{1.12}$$

donde

$$\theta_k = \frac{p_k^t p_k}{2(f_k - f_{k+1} + g_{k+1}^t p_k)}. \tag{1.13}$$

Além disso, podemos também construir um modelo cúbico

$$c_{k+1}(\alpha) = f_{k+1} + \alpha g_{k+1}^t p_k + \frac{1}{2} t_k \alpha^2 \|p_k\|^2 + \xi_k \alpha^3$$

tal que as condições (1.8)-(1.11) são todas satisfeitas, donde segue que

$$t_k = \frac{6(f_k - f_{k+1}) + 4g_{k+1}^t p_k + 2g_k^t p_k}{p_k^t p_k} \tag{1.14}$$

e, portanto,

$$\theta_k = \frac{p_k^t p_k}{6(f_K - f_{k+1}) + 4g_{k+1}^t p_k + 2g_k^t p_k}. \tag{1.15}$$

É fácil ver que as fórmulas (1.13) e (1.15) são idênticas a (1.5) se f é quadrática no segmento de reta x_k e x_{k+1}. Isto indica que se a função objetivo é quadrática (1.13) e (1.15) são exatamente os mesmos que (1.5). Então para uma função qualquer, como estamos usando as informações do valor da função e do valor do gradiente em x_k e x_{k+1} para derivar (1.13) e (1.15), é razoável esperar que estas duas fórmulas sejam melhores que (1.5).

Portanto, (1.3) e (1.4) nos dão uma estrutura unificada produzindo uma família de métodos de gradiente conjugado.

Observamos que as direções de Sherali-Ulular e de Birgin-Martinez serão paralelas se tomarmos o mesmo valor para os respectivos parâmetros s_k e θ_k.

2. O Algoritmo

Nesta seção descreveremos um método que é uma combinação das idéias dos algoritmos de gradiente conjugado escalado de Birgin-Martinez [2] e Sherali-Ulular [13].

Algoritmo 2.1: Dados $x_o \in I\!R^n$, $c = 10^{-4}$, $r = 0.9$ e $\epsilon > 0$

Inicialização
Faça

$k = 0$
$d_0 = -g_0$

Repita

$\alpha_k = 1$ ou $\left(\alpha_k = \begin{cases} 1, & \text{se } \; k = 0 \\ \alpha_{k-1}\frac{\|d_{k-1}\|}{\|d_k\|}, & \text{se } \; k \neq 0 \end{cases} \right)$

$\alpha = \alpha_k$

Enquanto $f(x_k + \alpha d_k) > f(x_k) + c\alpha g_k^t d_k$ ou $g(x_k + \alpha d_k)^t d_k < r g_k^t d_k$ escolha $\sigma \in [0.1, 0.5]$ e faça $\alpha = \sigma\alpha$.

Fim

$\alpha_k = \alpha$

$x_{k+1} = x_k + \alpha_k d_k$

$\beta_k = \frac{(\theta_k y_k - \frac{1}{s_k} p_k)^t g_{k+1}}{y_k^t d_k}$, com $\theta_k = 1$ (ou dado por (1.5) ou (1.13) ou (1.15)) e $s_k = 1$ (ou $s_k = \alpha_k$)

$d = -\theta_k g_{k+1} + \beta_k d_k$

Se $d^t g_{k+1} \leq -10^{-3}\|d\|\|g_{k+1}\|$ então $d_{k+1} = d$.

Senão, $d_{k+1} = -\theta_k g_{k+1}$

$k = k + 1$

Até que $\|g_k\| < \epsilon$

Comentários:

- A direção $d = -\theta_k g_{k+1} + \beta_k d_k$ não é necessariamente uma direção de descida. E esta é a razão que motivou as várias modificações da fórmula de Perry [8] feitas por Shanno [11].
- Estamos usando como reinicialização o mesmo critério ingênuo de Birgin-Martinez [2]. Com isso estamos garantindo que d é uma direção de descida.
- Em Dennis-Schnabel [4] e Fletcher [5] é mostrado que um comprimento de passo α satisfazendo

  $$f(x_k + \alpha d_k) \leq f(x_k) + c\alpha g_k^t d_k \qquad \text{e} \qquad g(x_k + \alpha d_k)^t d_k \geq r g_k^t d_k$$

  sempre existe se f é limitada inferiormente na direção d_k.
- Nas implementações também realizamos busca linear como em Shanno-Phua [12].

3. Resultados Numéricos e Conclusões

Os resultados numéricos foram obtidos usando os problemas testes de Sherali-Ulular [13]. Consideramos oito métodos diferentes com dois critérios para escolha do tamanho de passo inicial na busca linear. Os métodos foram identificados como na Tabela 1 abaixo:

Tabela 1

MÉTODO	θ_k	s_k
M_1	$p_k^t p_k / p_k^t y_k$	α_k
M_2	$p_k^t p_k / 2(f_k - f_{k+1} + g_{k+1}^t p_k)$	α_k
M_3	1	α_k
M_4	$p_k^t p_k / [6(f_K - f_{k+1}) + 4g_{k+1}^t p_k + 2g_k^t p_k]$	α_k
M_5	$p_k^t p_k / p_k^t y_k$	1
M_6	$p_k^t p_k / 2(f_k - f_{k+1} + g_{k+1}^t p_k)$	1
M_7	1	1
M_8	$p_k^t p_k / [6(f_K - f_{k+1}) + 4g_{k+1}^t p_k + 2g_k^t p_k]$	1

Os resultados da Tabela 2 correspondem a escolha do passo inicial na busca linear de $\alpha_k = 1$ e os da Tabela 3 para $\alpha_k = \begin{cases} 1 & \text{se} \;\; k = 0, \\ \alpha_{k-1} \frac{\|d_{k-1}\|}{\|d_k\|} & \text{se} \;\; k \neq 0. \end{cases}$

Os valores representados nas colunas das Tabelas 2 e 3 são o número de iterações e o número de chamadas da função, exceto na primeira coluna, onde P é o número do problema e n é o número de variáveis.

Tabela 2

P/n	M1	M2	M3	M4	M5	M6	M7	M8
1/2	4/30	4/30	9/42	5/36	6/40	6/40	14/71	16/79
2/2	28/184	27/202	35/113	24/116	43/237	37/208	33/180	14/102
3/4	130/461	102/366	216/647	118/357	164/643	254/857	53/242	250/836
4/4	100/388	169/630	115/355	116/436	105/485	78/307	52/209	108/420
5/2	30/93	11/44	16/66	9/55	35/102	9/39	12/48	12/71
6/5	126/469	137/470	172/454	203/597	143/556	131/500	153/466	97/374
6/10	356/929	343/885	277/647	271/759	225/702	252/779	347/881	227/713
7/20	22/61	25/64	18/61	25/56	24/56	17/51	35/101	34/87
7/40	34/93	36/92	28/81	38/101	27/66	36/93	64/149	40/108
7/200	87/209	96/255	77/170	110/261	93/246	79/220	96/221	90/235
8/3	51/267	94/333	84/243	56/218	57/231	83/279	80/260	70/275
9/5	304/937	266/874	201/535	209/657	235/684	271/886	368/1024	248/690
10/10	8/60	8/60	7/54	9/62	9/65	8/61	9/87	5/50

Tabela 3

P/n	M1	M2	M3	M4	M5	M6	M7	M8
1/2	7/37	7/37	4/21	11/54	6/16	6/16	7/107	7/20
2/2	23/215	28/190	26/185	22/153	34/209	28/137	28/196	21/124
3/4	150/608	434/1801	160/1181	115/450	111/420	54/198	88/593	234/868
4/4	69/368	97/547	42/224	82/377	179/1043	35/165	105/511	60/251
5/2	17/59	19/57	14/41	16/46	22/70	15/40	15/43	20/51
6/5	299/1343	355/1513	222/1664	112/390	197/711	173/632	67/477	121/390
6/10	437/1845	437/1916	454/3487	440/2938	224/804	255/868	131/958	193/791
7/20	24/52	27/61	158/204	25/57	21/32	24/35	156/222	26/62
7/40	50/124	35/85	184/251	38/104	47/94	35/58	169/270	29/69
7/200	72/240	71/241	130/352	70/206	71/167	66/163	120/360	64/160
8/3	38/145	68/231	44/296	65/215	57/190	59/212	64/418	65/185
9/5	172/640	223/859	158/785	175/296	176/711	156/636	144/733	157/568
10/10	7/25	7/25	6/84	7/25	7/15	7/15	7/76	7/15

Dos resultados numéricos obtidos concluímos que os métodos são competitivos, sendo que seu desempenho varia de acordo com o problema testado, com a escolha de um determinado parâmetro usado na busca linear (nossas tabelas apresentam os resultados apenas com valor 1/3 para tal parâmetro) e com a escolha inicial para o tamanho do passo α_k na busca linear.

Os autores deste trabalho agradecem ao Prof. Dr. Jin Yun Yuan (UFPR) pelos comentários e sugestões recebidas.

Referências

[1] J. Barzilai, J.M. Borwein, Two-point stepsize gradient methods, *IMA J. Num. Anal.*, **8** (1988), 141-148.

[2] E.G. Birgin, J.M. Martinez, A spectral conjugate gradiente method for unconstrained optimization, Technical Report, IMECC, Universidade de Campinas, 1998.

[3] Y.H. Dai, Y.J. Yuan, Y. Yuan, Modified two-point stepsize gradient methods for unconstrained optimization, Relatório Técnico, Departamento de Matemática, Universidade Federal do Paraná, 1998.

[4] J.E. Dennis Jr, R.B. Schnabel, "Numerical methods for unconstrained optimization and nonlinear equations", Prentice-Hall, Englewood Cliffs, 1983.

[5] R. Fletcher, "Practical methods of optimization", John-Wiley and Sons, New York, 1987.

[6] R. Fletcher, C.M. Reeves, Function minimization by conjugate gradient, *Comput. J.*, **7** (1964), 149-154.

[7] M.R. Hestenes, E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Bur. Standards Sec.*, **48** (1952), 429-436.

[8] A. Perry, A modified conjugate gradient algorithm, *Operations Research*, **26** (1978), 1073-1078.

[9] E. Polak, G. Ribière, Note sur la convergence de méthodes de directions conjuguées, *Revue Française d'Informatique et de Recherche Opérationnelle*, **16** (1969), 35-43.

[10] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, *SIAM J. Optim.*, **7** (1997), 26-33.

[11] D.F. Shanno, Conjugate gradiente method with inexact searches, *Mathematics of Operations Research*, **3** (1978), 244-256.

[12] D.F. Shanno, K.H. Phua, Minimization of unconstrained multivariate functions, *ACM Transactions on Mathematical Software*, **2** (1976), 87-94.

[13] H.D. Sherali, O. Ulular, Conjugate gradient methods using quasi-Newton updates with inexact line searches, *Journal of Mathematical Analysis and Applications*, **150** (1990), 359-377.