

Estudo do Desempenho de Variações do Método de Cauchy para Minimização Irrestrita

S.A. SANTOS¹, L.O. XAVIER², Departamento de Matemática Aplicada, IMECC, UNICAMP, 13083-970 Campinas, SP, Brasil.

Resumo. No contexto de métodos para minimização irrestrita, são apresentadas uma nova proposta de direção e novas escolhas para o tamanho de passo no algoritmo de máxima descida. Estas propostas foram inspiradas nos trabalhos de Barzilai e Borwein (1988), Raydan e Svaiter (2002) e Dai, Yuan e Yuan (2002). Para a comparação entre os novos métodos e os clássicos, foram realizados testes com problemas quadráticos e de quadrados mínimos não lineares. A análise comparativa dos resultados foi feita pela técnica sugerida por Dolan e Moré (2002).

1. Introdução

O problema considerado em nosso estudo consiste em minimizar $f(x)$, com $f : \mathbb{R}^n \rightarrow \mathbb{R}$ suave. A minimização sem restrições, além da importância em si mesma dentro da programação não linear, pela vasta gama de problemas modelados desta maneira, pode também ser utilizada como subproblema na minimização com restrições via penalização ou Lagrangeano aumentado. O algoritmo para a resolução do problema de minimização irrestrita consiste basicamente de três etapas:

Modelo de Algoritmo Geral para Minimização Irrestrita

- **Passo 1.** Escolha da direção de descida:
Tomar $d^k \in \mathbb{R}^n$ tal que $\nabla f(x^k)^T d^k < 0$.
- **Passo 2.** Determinação do tamanho do passo: Calcular $\lambda_k > 0$.
- **Passo 3.** Tomar $x^{k+1} = x^k + \lambda_k d^k$.

Em conseqüência, diferentes escolhas para a direção e para a determinação do passo darão origem a métodos distintos. Por exemplo, optando-se pela estratégia de busca linear, calcula-se $\lambda_k > 0$ tal que haja decréscimo suficiente para f : $f(x^k + \lambda_k d^k) \ll f(x^k)$. Outra idéia é tomar passos que gerem seqüências não monótonas para os valores funcionais (cf. [1]), e que podem se enquadrar na busca não monótona de Grippo, Lampariello e Lucidi [6].

¹sandra@ime.unicamp.br; apoio parcial CNPq (300206/96-8) e PRONEX-Otimização.

²larissa@ime.unicamp.br; bolsista FAPESP (02/13486-4) no período 04/2003 a 12/2004.

Neste trabalho abordamos dois tipos de função objetivo. O primeiro consiste em funções quadráticas $f(x) = \frac{1}{2}x^T Hx + b^T x + c$, com $H = H^T \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$. O segundo tipo de problema considerado é o de quadrados mínimos não lineares, com função objetivo $f(x) = F(x)^T F(x)$, onde $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Para a resolução destes dois tipos de problemas, utilizamos conjuntos de métodos distintos que exploram as características específicas e a estrutura de cada uma destas famílias.

Organizamos nossa apresentação da seguinte maneira. Na Seção 2 introduzimos alguns métodos já conhecidos, para depois apresentarmos as novas idéias para a escolha de direções e tamanhos de passo nas Seções 3 e 4, comparando seus desempenhos na Seção 5. Considerações finais são feitas na Seção 6.

2. Métodos Conhecidos

Uma escolha clássica para o passo 1 do algoritmo é a direção de *máxima descida* ou de *Cauchy* [3]: $d^k = -g^k$, onde denotamos $g^k = \nabla f(x^k)$. Para esta escolha de direção e fazendo uma busca linear exata no passo 2 do algoritmo, caso a função objetivo seja quadrática, temos caracterizado o *método de Cauchy com busca linear exata* (CBLE), para o qual o passo é dado por

$$\lambda_k = \frac{(g^k)^T g^k}{(g^k)^T H g^k}. \quad (2.1)$$

Porém, a convergência deste método, na maioria dos problemas, pode ser lenta, devido ao seu comportamento em ‘zigue-zague’ (cf. [2, p. 290]). Para melhorar o desempenho do método de máxima descida, Barzilai e Borwein [1] propuseram dois novos tamanhos de passo, que podem ser usados também na minimização de funções mais gerais, originados de aproximações para a matriz Hessiana e sua inversa a cada iteração por matrizes escalares $\alpha_k I$ e $\beta_k I$ respectivamente, as quais satisfazem equações secantes e resultam em

$$\alpha_k = \frac{(s^{k-1})^T y^{k-1}}{(s^{k-1})^T s^{k-1}} \quad (\text{BB1}) \quad \text{e} \quad \beta_k = \frac{(s^{k-1})^T y^{k-1}}{(y^{k-1})^T y^{k-1}} \quad (\text{BB2}),$$

onde $s^{k-1} = x^k - x^{k-1}$ e $y^{k-1} = g^k - g^{k-1}$. Assim, no passo 2 do algoritmo temos $\lambda_k = \frac{1}{\alpha_k}$ ou $\lambda_k = \beta_k$, conforme estejamos aproximando $H_k = \nabla^2 f(x^k)$ ou sua inversa H_k^{-1} . Estes métodos geram $\{f(x^k)\}$ não monótona.

Além destas escolhas, neste trabalho consideramos também os métodos propostos para problemas quadráticos por Raydan e Svaiter em [8]. O primeiro, o *método de Cauchy aleatório relaxado* (CAR), mantém a direção de máxima descida permitindo uma liberdade de escolha para o tamanho de passo em um intervalo que contém o passo determinado pela busca linear exata. Com este método, o passo 3 do algoritmo fica

$$x^{k+1} = x^k - \theta_k \lambda_k g^k$$

onde θ_k é escolhido aleatoriamente entre 0 e 2 e λ_k é dado por (2.1).

O segundo método apresentado em [8] é o *método de Cauchy-Barzilai-Borwein* (CBB), que utiliza o fato do primeiro passo proposto por Barzilai e Borwein ser

equivalente ao passo obtido fazendo-se uma busca linear exata na iteração anterior na minimização de uma quadrática. Na construção deste método, realiza-se uma iteração do método de máxima descida com busca linear exata seguida de uma iteração com BB1, utilizando o mesmo passo duas vezes, obtendo-se

$$x^{k+1} = x^k - 2\lambda_k g^k + \lambda_k^2 H g^k,$$

onde λ_k é dado por (2.1).

Para os testes de quadrados mínimos não lineares não utilizamos os métodos que envolvem a busca linear exata para a determinação do passo, pois esta busca se transformaria em um novo problema de minimização, cujo custo dispendido para se obter a solução poderia não justificar o benefício obtido.

Assim, além da direção de Cauchy com os passos propostos por Barzilai e Borwein, utilizamos o *método de Cauchy com bissecção* (Cbis):

$$x^{k+1} = x^k - \lambda g^k$$

onde $\lambda \in \{1, \frac{1}{2}, \frac{1}{4}, \dots\}$ é o maior número para o qual se verifica $f(x^k - \lambda g^k) < f(x^k)$.

3. Um Escalamento para a Direção de Cauchy

Nesta seção apresentamos novas idéias para a construção de direções e tamanhos de passo, originando novos métodos. Estudando os passos propostos por Barzilai e Borwein, tivemos a motivação de realizar uma modificação na escolha da matriz que aproxima a matriz Hessiana.

Fazendo a aproximação $L(x) = g^k + H_k(x - x^k)$ para o gradiente da função objetivo por um polinômio de Taylor de primeira ordem em torno de x^k e impondo $L(x^{k+1}) = 0$, resolvemos em cada iteração o sistema linear

$$H_k s^k = -g^k. \quad (3.1)$$

Para os passos propostos por Barzilai e Borwein, aproximamos as matrizes H_k ou H_k^{-1} por matrizes escalares. Com o objetivo de resolver um sistema que não fosse sobredeterminado, decidimos aproximar a matriz Hessiana por uma matriz diagonal $H_k \simeq M_k = \text{diag}(\alpha^k)$ em cada iteração, onde $(\alpha^k)^T = (\alpha_1^k, \alpha_2^k, \dots, \alpha_n^k)$.

Impondo a equação secante $L(x^{k-1}) = g^{k-1}$ para a aproximação $H_k \simeq M_k$ obtemos $M_k s^{k-1} = y^{k-1}$, com $s^{k-1} = x^k - x^{k-1}$ e $y^{k-1} = g^k - g^{k-1}$, ou seja,

$$\alpha_i^k = \frac{y_i^{k-1}}{s_i^{k-1}}, \quad \text{para todo } i = 1, \dots, n. \quad (3.2)$$

Assim, de $M_k s^k = -g^k$ e (3.2) obtemos a nova direção, que pode ser interpretada como uma direção de *Cauchy escalada* (CE), e é definida componente a componente por

$$s_i^k = -\frac{1}{\alpha_i^k} g_i^k = -\frac{s_i^{k-1}}{y_i^{k-1}} g_i^k,$$

com a qual temos o esquema iterativo $x^{k+1} = x^k + s^k$.

Para que esta direção esteja bem definida impusemos as seguintes salvaguardas. Substituímos as componentes y_i^{k-1} por 1 caso $|y_i^{k-1}| < \varepsilon$ e s_i^{k-1} por $\|s^{k-1}\|$ caso $|s_i^{k-1}| < \varepsilon$, onde ε é o valor utilizado no critério de parada $\|g^k\| < \varepsilon$, e $\|\cdot\|$ é a norma euclidiana.

Ampliando nosso estudo, percebemos que esta direção não está necessariamente definida como uma direção de descida. Para que isto ocorra s^k deve ser tal que $(g^k)^T s^k < 0$. Para assegurar este resultado basta impor que M_k seja positiva definida, tomando os elementos de sua diagonal (α_i^k) em valor absoluto, já que $(g^k)^T s^k = -(g^k)^T M_k^{-1} g^k$. Assim, temos a *direção de Cauchy escalada de descida* (CEdesc), cujas componentes são dadas por

$$s_i^k = - \left| \frac{s_i^{k-1}}{y_i^{k-1}} \right| g_i^k, \quad i = 1, \dots, n.$$

4. Modificações nos Passos de Barzilai e Borwein

Com o objetivo de acelerar a convergência na resolução de problemas mais gerais, ainda utilizando a direção de Cauchy, propusemos uma modificação para os passos de Barzilai e Borwein.

Uma nova idéia é utilizar a informação de um ponto intermediário x^a entre x^{k-1} e x^k , impondo que $g(x^a) = L(x^a)$ ao aproximar a matriz Hessiana H_k ou sua inversa H_k^{-1} por matrizes escalares $\bar{\alpha}_k I$ e $\bar{\beta}_k I$, respectivamente.

Tomando $x^a = x^k - c s^{k-1}$, com $c \in (0, 1)$ (para os testes computacionais adotamos $c = 0.1$), e denotando $y^a = \frac{1}{c}(g(x^k) - g(x^a))$ temos os dois novos passos $\frac{1}{\bar{\alpha}_k}$ e $\bar{\beta}_k$ com

$$\bar{\alpha}_k = \frac{(s^{k-1})^T y^a}{(s^{k-1})^T s^{k-1}} \quad (\text{BB1mod}) \quad \text{e} \quad \bar{\beta}_k = \frac{(s^{k-1})^T y^a}{(y^a)^T y^a} \quad (\text{BB2mod}).$$

Em [4], Dai, Yuan e Yuan interpretaram a escolha do primeiro passo proposto por Barzilai e Borwein usando o conceito de interpolação e propuseram um novo passo para ser utilizado com a direção de Cauchy. Inicialmente, estes autores aproximaram a função objetivo f por uma quadrática (Taylor de segunda ordem) em pontos da forma $x^k + \theta s^{k-1}$, com $\theta \in \mathbb{R}$, obtendo:

$$f(x^k + \theta s^{k-1}) \simeq f(x^k) + \theta (g^k)^T s^{k-1} + \frac{\theta^2}{2} (s^{k-1})^T H_k s^{k-1} = q(\theta).$$

Para interpolar a função f por esta quadrática, devem ser impostas três condições, que são estabelecidas fazendo-se

$$f(x^k + \theta s^{k-1}) = q(\theta) \quad (4.1)$$

$$\text{ou} \quad \frac{df(x^k + \theta s^{k-1})}{d\theta} = \frac{dq(\theta)}{d\theta} \quad (4.2)$$

para algum valor de θ . Com a aproximação $H_k \simeq tI$, vemos que para $\theta = 0$ as condições (4.1) e (4.2) valem para todo $t \in \mathbb{R}$. A terceira condição obtida

com $\theta = -1$ em (4.2) corresponde ao primeiro passo de Barzilai e Borwein (BB1). Tomando $\theta = -c$, com $c \in (0, 1)$, em (4.2), temos a primeira modificação (BB1mod). Fazendo $\theta = -1$ em (4.1) temos o passo proposto em [4]:

$$\frac{1}{t} = \frac{(s^{k-1})^T s^{k-1}}{2(f(x^{k-1}) - f(x^k) + (g^k)^T s^{k-1})} \quad (\text{DYY}).$$

Finalmente, com $\theta = -c$ em (4.1), obtivemos uma nova proposta de passo, utilizando informações no ponto intermediário x^a definido acima:

$$\frac{1}{t} = \frac{c^2 (s^{k-1})^T s^{k-1}}{2(f(x^a) - f(x^k) + c(g^k)^T s^{k-1})} \quad (\text{DYYmod}).$$

A desvantagem desta escolha é a avaliação extra da função no ponto x^a em cada iteração. Porém este cálculo tem, em geral, um custo menor do que o cálculo de $g(x^a)$ em cada iteração, que era necessário para as modificações propostas para os passos de Barzilai e Borwein, feitas anteriormente.

Utilizando a condição (4.2) com $\theta = -1$ e aproximando a matriz Hessiana H_k por $M_k = \text{diag}(\alpha^k)$ temos $(g^{k-1})^T s^{k-1} = (g^k)^T s^{k-1} - (s^{k-1})^T M_k s^{k-1}$ ou

$$(M_k s^{k-1} - y^{k-1})^T s^{k-1} = 0. \quad (4.3)$$

Podemos assim interpretar a nova direção proposta (CE) via interpolação, e notar que a escolha para M_k feita em (3.2) satisfaz (4.3), mas não é a única possível.

5. Testes Computacionais

Os algoritmos foram implementados em *Matlab* (6.1 e 6.5) e os testes efetuados em um Pentium 4 com processador de 2199MHz e 512 Mb de memória Ram.

O primeiro conjunto de testes é composto de 18 quadráticas bidimensionais estritamente convexas. Como critérios de parada impusemos $\|g^k\| \leq 10^{-7}$ ou que o número máximo de 50 iterações fosse atingido. Como os métodos com os passos de Barzilai e Borwein (BB1 e BB2) e a direção de Cauchy escalada (CE) exigem dois pontos para o início do esquema iterativo, para uniformizar a comparação obtivemos o segundo ponto aplicando CBLE ao ponto inicial. Os resultados são apresentados na Tabela 1. Para os 18 testes realizados notamos a rápida convergência de CE, ressaltando que as salvaguardas para esta direção nunca foram ativadas.

Notamos que nestes problemas quadráticos bidimensionais inicializados com CBLE, para $k = 2$, o sistema constituído pelas equações (3.1) e por $M_2 s^2 = -g^2$, nas componentes de s^2 , tem solução única. Em conseqüência, $x^3 = x^2 + s^2$ é a solução exata do problema. Neste caso também observamos que existe $p \in \mathbb{R}$ tal que $s^2 = ps^1$, e que este escalar depende do elemento $h_{12} = h_{21}$ de tal forma que, caso H seja diagonal, $p = 0$ e a convergência se dá em duas iterações. Estas conclusões foram obtidas combinando-se gráficos ilustrativos do comportamento do método, no *Matlab*, e resultados algébricos do *Mathematica*.

No conjunto seguinte de testes tomamos $f(x) = x^T H x$, onde H é diagonal com elementos aleatoriamente gerados entre 1 e 1000. O ponto inicial adotado foi o

Tabela 1: Resultados dos problemas quadráticos convexos bidimensionais.

H	Problema			Número de iterações					
	b	c	x^0	CBLE	BB1	BB2	CE	CBB	CAR
[2 0; 0 200]	[0; 0]	0	[100; 1]	50	50	10	2	38	50
[2 -1; -1 2]	[0; 0]	0	[3; 1]	24	11	9	3	5	26
[2 0; 0 4]	[0; 0]	0	[10; 1]	10	6	8	2	3	21
[2 0; 0 4]	[0; 0]	0	[10; 5]	18	18	12	2	9	21
[2 1; 1 4]	[1; -1]	1	[5; 1]	23	11	9	3	5	26
[4 0; 0 6]	[3; 1]	2	[0; 0]	9	6	7	2	3	20
[2 1; 1 2]	[2; 1]	5	[0; 0]	15	8	9	3	4	24
[2 1; 1 20]	[1; 1]	0	[0; 0]	50	16	10	3	6	46
[2 -1; -1 2]	[0; 0]	0	[2; 5]	26	11	9	3	5	28
[2 0; 0 20]	[0; 0]	0	[1; 5]	7	6	6	2	3	50
[2 0; 0 20]	[0; 0]	0	[5; 1]	50	11	9	2	4	50
[2 0; 0 2]	[-5; 5]	0	[0; 0]	1	1	1	1	1	20
[2 0; 0 200]	[-10; -1]	0	[0; 0]	50	8	8	2	4	50
[2 0; 0 20]	[-10; -1]	0	[0; 0]	15	6	8	2	3	50
[2 0; 0 4]	[-10; -1]	0	[0; 0]	8	6	6	2	3	21
[2 0; 0 20]	[-5; 5]	0	[0; 0]	50	50	10	2	37	36
[2 0; 0 200]	[-5; 5]	0	[0; 0]	50	50	9	2	37	50
[2 0; 0 2]	[-10; -1]	0	[0; 0]	1	1	1	1	1	20

vetor com todas as componentes iguais a 10^6 , com a dimensão adequada. Como critério de parada impusemos que o processo se repita até que $\|x^k\| \leq 10^{-12}$, já que neste caso sabemos que a solução é sempre a origem, ou que o número máximo de 500 iterações seja atingido. Os resultados estão apresentados na Tabela 2, onde Dim é a dimensão do problema, a e A são o menor e o maior autovalores de H , respectivamente, exibidos para verificarmos o condicionamento desta matriz.

Para os quatro problemas testados o método CBB proposto por Raydan e Svaiter obteve melhor desempenho quando comparado aos métodos CBLE, BB1, BB2 e CAR. A direção proposta CE continua superando os demais métodos para problemas quadráticos de maior dimensão com a matriz hessiana diagonal e positiva definida, resolvendo os problemas com apenas 3 iterações. Tal comportamento se justifica pelo fato dos problemas serem diagonais, o que torna o método com a direção CE idêntico ao método de Newton³ a partir do segundo ponto, que é obtido do primeiro fazendo-se uma iteração de CBLE. Para a direção CE temos $M_k s^{k-1} = y^{k-1}$ pela relação (3.2). Por outro lado, $y^{k-1} = g^k - g^{k-1} = H s^{k-1}$ para problemas quadráticos. Assim, se H é diagonal, temos que a aproximação da matriz Hessiana pela matriz M_k é exata. A terceira iteração calculada pela direção CE foi necessária para atingir a precisão exigida pelo critério de parada. Como exemplo, para o problema de dimensão 100, tivemos $\|x^0\| = 10^7$, $\|x^1\| = 5.0652 * 10^6$, $\|x^2\| = 4.2070 * 10^{-10}$ e $\|x^3\| = 3.7404 * 10^{-26}$.

Realizamos ainda testes para problemas quadráticos $\min x^T H x$ com matriz Hessiana densa e positiva definida de dimensões maiores. Construímos matrizes da

³O método de Newton converge em uma iteração para problemas quadráticos estritamente convexos, e neste caso a direção de Newton é $d^k = -H^{-1}g^k$.

Tabela 2: Resultados dos problemas quadráticos com H diagonal positiva definida com elementos gerados aleatoriamente entre 1 e 1000.

Problema			Número de iterações					
<i>Dim</i>	<i>a</i>	<i>A</i>	CBLE	CAR	BB1	BB2	CBB	CE
5	194.3338	664.9781	44	66	25	22	12	3
10	8.2362	757.0777	500	500	218	233	86	3
100	4.3308	997.6388	500	500	408	373	186	3
1000	1.6218	999.6068	500	500	500	500	404	3

forma $H = QDQ^T$, onde $D = \text{diag}(d)$ é a matriz que armazena os autovalores d_i de H e Q é uma matriz ortogonal construída via transformação de Householder: $Q = I - 2\frac{uu^T}{u^T u}$. As componentes dos vetores d e u foram geradas aleatoriamente nos intervalos $(1, 100)$ e $(-1, 1)$, respectivamente.

Os métodos foram programados explorando-se a estrutura para os produtos matriz-vetor, sem o armazenamento efetivo da matriz H . Como critério de parada impusemos que o processo se repita até que $\|x^k\| \leq 10^{-7}$ ou que o número máximo de 500 iterações seja atingido. O ponto inicial adotado foi o vetor com todas as componentes iguais a 1. Na Tabela 3 apresentamos os resultados, seguindo a mesma estrutura da Tabela 2.

Notamos um bom comportamento da direção CE para problemas bidimensionais, superando os demais métodos. Porém, com o aumento da dimensão do problema, a direção CE perde seu bom desempenho. Para estes problemas CBB apresenta os melhores resultados, seguido por BB1 e BB2, que competem entre si. Comparando CE e CEdesc, percebemos um melhor desempenho de CEdesc, superando também CBLE para os problemas de dimensão 10 e 15 com 38 e 157 iterações, respectivamente, contra 44 e 198 iterações de CBLE.

Tabela 3: Resultados dos problemas quadráticos com H positiva definida.

Problema			Número de iterações						
<i>Dim</i>	<i>a</i>	<i>A</i>	CBLE	CAR	BB1	BB2	CBB	CE	CEdesc
2	20.1592	63.2568	14	25	8	6	4	3	3
3	20.1592	75.9266	27	29	16	13	6	77	59
4	8.0685	75.9266	66	51	20	22	14	291	109
5	1.7171	75.9266	194	168	46	41	18	500	500
10	17.1781	92.2973	44	38	18	18	10	49	38
15	3.5285	86.6780	198	102	53	55	29	335	157

O primeiro conjunto de problemas de quadrados mínimos não lineares consiste na família de 18 problemas sugerida em [7], trocando-se a função de Jennrich e Sampson pela função mal escalada de Powell. Como critério de parada impusemos

$\|g_k\| \leq 10^{-8}$, $\|x^k - x^{k-1}\| \leq 10^{-16}$ ou número máximo de 1000 iterações. Para compararmos os resultados e colocarmos as novas propostas em um contexto, utilizamos gráficos de *performance profile*, cujo estudo foi baseado em Dolan e Moré [5]. Para a construção dos gráficos consideramos dois conjuntos de métodos. O primeiro com os métodos conhecidos Cbis, BB1, BB2 e DYY e a nova direção CE, e o segundo com os passos BB1, BB1mod, BB2, BB2mod, DYY e DYYmod para a direção de Cauchy. Estes gráficos compõem a Figura 1.

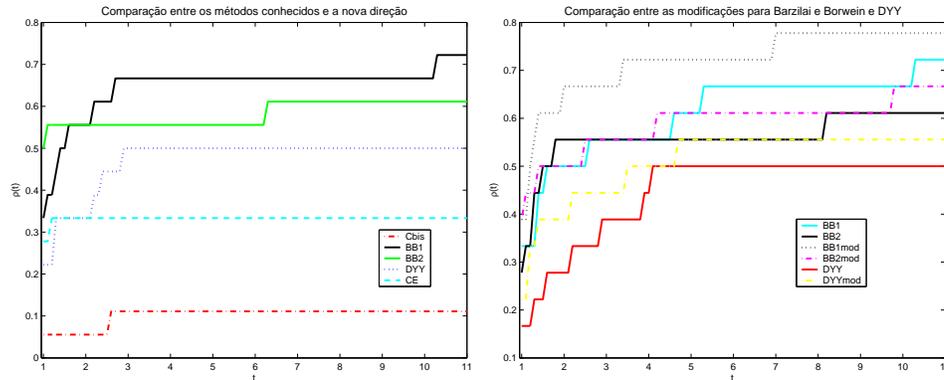


Figura 1: Gráficos de *performance profile* para 18 problemas de [7].

De uma forma geral, no primeiro conjunto os passos propostos por Barzilai e Borwein são comparáveis entre si, o método de máxima descida com bissecção é o que resolve menos problemas com mais iterações, seguido da direção CE. No segundo conjunto de métodos destacam-se as modificações sugeridas, cujo desempenho supera o dos métodos originais.

Para verificarmos o comportamento da direção CE para problemas com Hessiana diagonal de maior dimensão, o próximo problema testado foi $\min F(x)^T F(x)$ tal que

$$F_i(x) = \begin{cases} x_i(1 - \text{sen}(x_i)) , & \text{se } i \text{ é ímpar} \\ 1 - x_i \text{sen}(x_i) \cos(x_i) , & \text{se } i \text{ é par,} \end{cases} \quad \text{com } i = 1, \dots, n. \quad (5.1)$$

Os resultados estão apresentados na Tabela 4, com o número de iterações executado com cada método, onde *C* indica convergência, *N* indica não convergência pelo método, * indica que a norma do gradiente no ponto final está entre 10^{-6} e 10^{-4} e ** indica que a norma do gradiente no ponto final entre 10^{-8} e 10^{-6} . O ponto inicial utilizado foi o vetor de dimensão *n* com todas as componentes iguais a 1. A Figura 2 ilustra a Tabela 4.

Para este conjunto de testes notamos o melhor desempenho da direção CE, superando os demais métodos. A direção CE foi de descida em todas as iterações, o que a fez coincidir com CEdesc. Vale dizer que a direção CE para problemas separáveis,

como (5.1), coincide com a direção de Newton discreta, na qual as derivadas segundas são aproximadas via diferenças finitas. Em decorrência da separabilidade, o número de iterações dispendido não cresce com a dimensão do problema. Comparando o método de máxima descida Cbis com os passos de Barzilai e Borwein, de Dai, Yuan e Yuan, e com as modificações propostas, temos o melhor desempenho para BB1mod e BB2mod.

Tabela 4: Resultados para o problema $\min F(x)^T F(x)$ para F dada por (5.1).

Método	Dimensão				
	2	5	10	20	100
Cbis	1000 *	1000 *	1000 *	1000 *	1000 N
DYY	65 C	51 C	80 **	72 **	78 **
DYYmod	50 **	55 **	48 **	52 **	57 *
BB1	54 C	60 C	56 C	56 C	67 C
BB1mod	40 C	47 C	44 C	44 C	44 C
BB2	50 C	53 C	55 C	55 C	55 C
BB2mod	41 C	45 C	44 C	44 C	50 C
CE	18 C	19 C	19 C	21 C	21 C
CEdesc	18 C	19 C	19 C	21 C	21 C

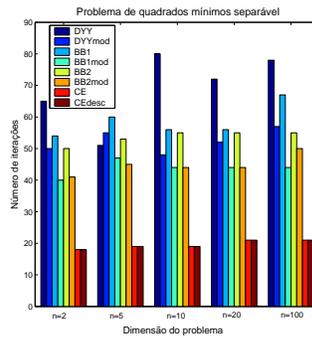


Figura 2: Ilustração para os resultados da Tabela 4.

6. Considerações Finais

Apresentamos, neste trabalho, métodos clássicos e novas propostas para a direção e o tamanho do passo na minimização irrestrita. Trabalhamos com problemas quadráticos e de quadrados mínimos não lineares. Para estas duas famílias de problemas fizemos vários conjuntos de experimentos numéricos comparativos para situar o desempenho das propostas apresentadas no contexto dos métodos já conhecidos, apresentados resumidamente aqui, e de forma completa em [9].

Comparando o desempenho dos métodos para problemas quadráticos temos que a direção proposta CE supera os demais métodos para problemas bidimensionais e para problemas de maior dimensão com matriz Hessiana diagonal. Para os outros problemas quadráticos o melhor desempenho é obtido pelo método CBB.

Para os problemas de quadrados mínimos não lineares com matriz Hessiana diagonal, a direção CE também supera os demais métodos. Para os problemas mais gerais, o melhor desempenho é alcançado pelo método de máxima descida com dois novos passos originados de modificações feitas para os passos de Barzilai e Borwein. Os passos de Barzilai e Borwein são comparáveis entre si, superando o método de máxima descida com busca linear exata para problemas quadráticos e com bissecção para problemas não lineares.

De uma maneira geral, a direção proposta neste trabalho, cuja definição incorpora naturalmente o tamanho do passo, e que pode ser vista como um escalamento da direção de Cauchy, se destacou para problemas quadráticos bidimensionais e para problemas com variáveis separáveis (Hessiana diagonal), tanto quadráticos quanto não lineares. O desempenho desta nova direção para tais problemas superou o dos demais métodos. Particularmente para problemas quadráticos bidimensionais constatamos que a direção CE inicializada com CBLE possui convergência finita, em no máximo 3 iterações. Vale dizer que nossos esforços no sentido de globalizá-la para problemas mais gerais não se reverteram em um algoritmo eficiente do ponto de vista prático. Nem mesmo a busca não monótona de Grippo, Lampariello e Lucidi [6] favoreceu o desempenho do algoritmo globalizado proposto. Provavelmente, a maneira com que construímos as salvaguardas tenha sido muito conservadora, o que provocou um comportamento análogo ao método de máxima descida globalizado: embora a convergência a um ponto estacionário esteja assegurada para qualquer inicialização, nada se garante com relação à velocidade dessa convergência. Esta análise é feita em [9].

Com inspiração nos artigos de Barzilai e Borwein [1], Raydan e Svaiter [8] e Dai, Yuan e Yuan [4], propusemos novas escolhas para o tamanho do passo, para serem usadas com a direção de máxima descida, originadas de modificações para os passos de Barzilai e Borwein. O desempenho dos métodos originados destas novas escolhas, em termos do número de iterações necessárias para a convergência, superou o dos demais métodos para os problemas mais gerais considerados neste trabalho.

Abstract. A new direction and new choices for the step length with the maximum descent direction are introduced in the context of unconstrained minimization. The ideas of Barzilai and Borwein, Raydan and Svaiter and Dai, Yuan and Yuan provided inspiration for the proposed development. Tests with quadratic and nonlinear least squares problems were performed to put the new methods in perspective, having the classical ones as benchmarks. The performance profile technique suggested by Dolan and Moré was used to compare the results.

Referências

- [1] J. Barzilai e J. Borwein, Two-point step size gradient methods, *IMA J. Numer. Anal.*, **8** (1988), 141-148.

- [2] M.S. Bazaraa, H.D. Sherali e C.M. Shetty, “Nonlinear Programming: Theory and Algorithms”, 2 ed., John Wiley, New York, 1993.
- [3] A. Cauchy, Méthode générale pour la resolution des systems d'équations simultanées, *Comp. Rend. Sci Paris*, **25** (1847), 536-538.
- [4] Y. Dai, J. Yuan e Y. Yuan, Modified Two-point stepsize gradient methods for unconstrained optimization, *Comput. Optim. Appl.*, **22**, No. 1 (2002), 103-109.
- [5] E.D. Dolan e J.J. Moré, Benchmarking optimization software with performance profiles, *Math. Programming*, **91** (2002), 201-213.
- [6] L. Grippo, F. Lampariello e S. Lucidi, A nonmonotone line search technique for Newton's method, *SIAM J. Numer. Anal.*, **23** (1986), 707-716.
- [7] J.J. Moré, B.S. Garbow e K.E. Hillstrom, Testing unconstrained optimization software, *ACM Trans. Math. Software*, **7** (1981), 17-41.
- [8] M. Raydan e B.F. Svaiter, Relaxed steepest descent and Cauchy-Barzilai-Borwein method, *Comput. Optim. Appl.*, **21** (2002), 155-167.
- [9] S.A. Santos e L.O. Xavier, “Estudo do Desempenho de Métodos para Minimização Irrestrita com Controle de Passo”, Relatório de Pesquisa RP18/04, IMECC, Unicamp, Campinas, SP, abril 2004. Disponível em http://www.ime.unicamp.br/rel_pesq/2004/rp18-04.html.

