

**Estudo comparativo de
passos espectrais e
buscas lineares não monótonas**

Fernando Taietti Camargo

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS

Área de Concentração: Ciência da Computação

Orientador: Prof. Dr. Ernesto G. Birgin

São Paulo, Abril de 2008

Estudo comparativo de passos espectrais e buscas lineares não monótonas

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Fernando Taietti Camargo e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Prof. Dr. Ernesto G. Birgin (Orientador) - IME-USP.
- Prof. Dr. Julio Michael Stern - IME-USP.
- Profa. Dra. Celma de Oliveira Ribeiro - EP-USP.

Resumo

O método do Gradiente Espectral, introduzido por Barzilai e Borwein e analisado por Raydan, para minimização irrestrita, é um método simples cujo desempenho é comparável ao de métodos tradicionais como, por exemplo, gradientes conjugados. Desde a introdução do método, assim como da sua extensão para minimização em conjuntos convexos, foram introduzidas várias combinações de passos espectrais diferentes, assim como de buscas lineares não monótonas diferentes. Dos resultados numéricos apresentados em vários trabalhos não é possível inferir se existem diferenças significativas no desempenho dos diversos métodos. Além disso, também não fica clara a relevância das buscas não monótonas como uma ferramenta em si próprias ou se, na verdade, elas são úteis apenas para permitir que o método seja o mais parecido possível com o método original de Barzilai e Borwein. O objetivo deste trabalho é comparar os diversos métodos recentemente introduzidos como combinações de diferentes buscas lineares não monótonas e diferentes passos espectrais para encontrar a melhor combinação e, a partir daí, aferir o desempenho numérico do método.

Palavras chave: minimização irrestrita, minimização em conjuntos convexos, passo espectral, busca linear não monótona.

Abstract

The Spectral Gradient method, introduced by Barzilai and Borwein and analyzed by Raydan for unconstrained minimization, is a simple method whose performance is comparable to traditional methods, such as conjugate gradients. Since the introduction of method, as well as its extension to minimization of convex sets, there were introduced various combinations of different spectral steplengths, as well as different nonmonotone line searches. By the numerical results presented in many studies it is not possible to infer whether there are significant differences in the performance of various methods. It also is not sure the relevance of the nonmonotone line searches as a tool in themselves or whether, in fact, they are usefull only to allow the method to be as similar as possible with the original method of Barzilai e Borwein. The objective of this study is to compare the different methods recently introduced as different combinations of nonmonotone linear searches and different spectral steplengths to find the best combination and from there, evaluating the numerical performance of the method.

Keywords: unconstrained minimization, minimization in convex sets, spectral steplength, nonmonotone line search.

Sumário

1	Introdução	1
2	O método do Gradiente Espectral Projetado	3
2.1	O passo espectral	5
2.2	A busca linear	7
2.3	Algoritmo	8
3	Passos espectrais	10
3.1	Barzilai e Borwein, 1988 (BB1 e BB2)	10
3.2	Grippo e Sciandrone, 2002 (ABB-GS)	11
3.3	Dai e Fletcher, 2005 (ABB-DF)	12
3.4	Dai, 2003 e Dai, Hager, Schittkowski e Zhang, 2006 (CBB)	12
3.5	Dai e Fletcher, 2006 (Multi-ponto)	13
4	Buscas lineares não monótonas	14
4.1	Grippo, Lampariello e Lucidi, 1986 (GLL1)	14
4.2	Grippo, Lampariello e Lucidi, 1991 (GLL2)	16
4.3	Toint, 1996	19
4.4	Dai e Zhang, 2001	20

4.5	Zhang e Hager, 2004	23
4.6	La Cruz, Martínez e Raydan, 2005	23
5	Experimentos	25
6	Conclusões e pesquisas futuras	50
	Referências Bibliográficas	52

Capítulo 1

Introdução

Estamos interessados em resolver o problema

$$\text{Minimizar } f(x) \text{ sujeita a } x \in \Omega,$$

onde Ω é um conjunto fechado e convexo em \mathbb{R}^n . Supomos que f está definida e tem derivadas parciais contínuas num conjunto aberto que contém Ω .

O método do Gradiente Espectral (SGM) para minimização irrestrita foi introduzido por Barzilai e Borwein [5] e analisado por Raydan [52]. As duas versões de passos espectrais introduzidas por Barzilai e Borwein podem, de alguma forma, ser interpretadas como aproximações secantes da matriz Hessiana. As principais características deste método são que só direções de gradiente são utilizadas em cada busca linear e que uma estratégia não monótona garante convergência global. Surpreendentemente, este método supera algoritmos sofisticados de gradientes conjugados em muitos problemas.

Desde a introdução do método, assim como da sua extensão para minimização em conjuntos convexos, foram introduzidas várias combinações de passos espectrais diferentes [25–28, 35, 36, 43, 54], assim como de buscas lineares não monótonas diferentes [30, 41, 45, 57, 62]. Dos resultados numéricos apresentados não é possível inferir se existem diferenças significativas no desempenho dos diversos métodos. Além disso, também não fica clara a relevância das buscas não monótonas como uma ferramenta em si próprias ou se, na verdade, elas são úteis apenas para permitir que o método seja o mais parecido possível com o método original de Barzilai e

Borwein.

O objetivo deste trabalho é comparar os diversos métodos recentemente introduzidos como combinações de diferentes buscas lineares não monótonas e diferentes passos espectrais para encontrar a melhor combinação e, a partir daí, aferir o desempenho numérico do método e poder responder às seguintes perguntas:

- Existe algum passo spectral que seja melhor do que os outros em todas as situações?
- É verdade que métodos baseados em combinações de passos espectrais têm um desempenho melhor?
- Existe diferença entre as diferentes buscas não monótonas?
- As buscas não monótonas servem apenas para provar a convergência do método ou elas têm um papel fundamental no seu desempenho?

Os métodos espectrais têm sido utilizados com sucesso no contexto de métodos de restrições ativas quando uma face deve ser abandonada [2,9,10]. Responder às perguntas acima é importante não só para melhorar o desempenho dos métodos espectrais de minimização irrestrita como também para melhorar o desempenho de métodos mais sofisticados de programação não linear que utilizam os gradientes espectrais como sub-algoritmos.

No Capítulo 2, apresentamos o método do Gradiente Espectral Projetado. No Capítulo 3, apresentamos tipos diferentes de passos espectrais. Apresentamos diversas variantes de buscas lineares não monótonas no Capítulo 4. Finalmente, no Capítulo 5 mostramos os experimentos numéricos realizados e no Capítulo 6 mencionamos algumas conclusões obtidas neste trabalho.

Capítulo 2

O método do Gradiente Espectral Projetado

O conteúdo deste capítulo baseia-se fortemente em [13, 14, 16].

Primeiramente, não podemos deixar de mencionar a evolução do método do Gradiente Espectral (SGM) proposto por Barzilai e Borwein [5] em 1988 até ser utilizado pelo método do Gradiente Espectral Projetado. Para acompanhar o avanço do método SGM utilizamos o trabalho de Fletcher [33].

Em [5], Barzilai e Borwein utilizaram uma estratégia diferente para escolher o tamanho do passo. O resultado principal desse estudo foi mostrar que, para $n = 2$, o método apresenta convergência R-superlinear. Os autores de [5] também mostraram que o método deles é consideravelmente superior ao método de máxima descida clássico para uma instância de uma função quadrática com $n = 4$, porém nenhum outro resultado numérico foi apresentado. Em 1993, Raydan [52] demonstrou a convergência para o caso quadrático. Entretanto, como não apresentou resultados numéricos, o método atraiu pouca atenção até 1997 com a publicação de Raydan [53]. Em [53], Raydan introduziu uma estratégia de globalização com a técnica de busca linear de Grippo, Lampariello e Lucidi [41], que permite convergência global ao método para funções não quadráticas. Verificou-se grande experiência numérica quando o método SGM foi aplicado a problemas com até 10^4 variáveis, mostrando que o método compara-se razoavelmente bem contra as técnicas de CONMIN e de gradientes conjugados de Polak-Ribière (método PR-CG). A possibilidade de aplicar o método SGM a problemas de

otimização com restrições de caixa foi considerada por Friedlander, Martínez e Raydan [37] para funções quadráticas e por Birgin, Martínez e Raydan [13] em que consideraram o método no contexto de projeção sobre um conjunto convexo. O desenvolvimento teórico de Dai e Liao [29] apresentou o resultado de que o método sem modificação apresenta convergência R-linear no caso quadrático. Em [33], Fletcher considera que o método SGM converge e ainda apresenta um desempenho relativamente melhor do que o método de gradiente conjugado PR-CG nos casos em que a função objetivo é composta em grande parte por uma função quadrática acrescida de um termo não quadrático pequeno.

Nas últimas décadas, surgiram muitas diversificações do método do gradiente projetado que podem ser vistas como extensões restritas do método de máxima descida para minimização irrestrita.

O método do Gradiente Espectral Projetado (SPG) nasceu da combinação entre as idéias não monótonas de Barzilai-Borwein (passo espectral) com as estratégias de gradiente projetado clássicas [7, 40, 46]. Entretanto, algumas diferenças críticas fazem este método muito mais eficiente do que seus predecessores de gradiente projetado.

Dado que o método do gradiente projetado é muito simples de se implementar e eficiente, os autores de [13] combinaram o método do gradiente projetado a dois ingredientes de otimização. Primeiro, estenderam as estratégias de globalização típicas associadas a estes métodos para o esquema de busca linear não monótona desenvolvido por Grippo, Lampariello e Lucidi em [41] para o método de Newton o que representou uma grande melhoria em relação aos métodos de gradiente projetado tradicionais. Segundo, propuseram a associação do passo espectral, introduzido por Barzilai e Borwein [5] e analisado por Raydan [52]. O passo espectral é um quociente de Rayleigh relacionado com uma média da matriz Hessiana. Esta escolha do tamanho do passo requer pouco esforço computacional e aumenta a velocidade de convergência dos métodos de gradiente.

Desde o seu surgimento, o método SPG tem sido muito utilizado em aplicações de diversas áreas tais como física [4, 24, 51, 59, 60], geofísica [6, 31, 61], química [34], engenharia [23], entre outras [8, 11, 12, 17, 18, 44, 55]. Em [6], o SPG foi adaptado para resolver problemas de tomografia de reflexão sísmica e aplicado tanto a problemas com restrições de caixa quanto problemas sem restrições de um conjunto de testes baseado em dados sintéticos. Verificou-se nos experimentos que o método SPG é uma técnica robusta na solução desses problemas quando é feita uma definição adequada da região viável convexa. Em [61], o método SPG é

utilizado para resolver os problemas de inversão sísmica, que até então era resolvido utilizando-se diferentes estratégias de otimização clássicas que necessitam de decréscimo monótono da função objetivo. Os resultados dos testes apontam que a aplicação com SPG apresentou desempenho melhor do que com o método de gradiente clássico pois reduziu o número de avaliações de função e valores residuais.

Além disso, o método SPG também tem sido objeto de várias modificações do parâmetro espectral, estratégias não monótonas alternativas foram sugeridas, convergência e propriedades de estabilidade foram esclarecidas e tem sido combinado com outros algoritmos para problemas de otimização diferentes. Em [14], uma família de problemas de locação foi descrita com um número grande de variáveis e restrições. O método SPG foi capaz de resolver problemas desta família com até 96254 variáveis e 578648 restrições em poucos segundos de tempo de CPU. Em [10], um método de restrições ativas que utiliza o método SPG para deixar as faces foi introduzido e problemas restritos por caixas com 10^7 variáveis foram resolvidos. Em [15], foi introduzido o método do Gradiente Espectral Projetado Inexato (ISPG) que é uma generalização do método SPG e é utilizado quando as projeções são difíceis de calcular pois utiliza o método de projeção alternada de Dykstra para encontrar as projeções aproximadas.

2.1 O passo espectral

O método SPG utiliza a versão mais popular do passo espectral de Barzilai e Borwein. A seguir, apresentamos como as duas versões do passo espectral de Barzilai e Borwein são encontradas a partir de aproximações secantes da matriz Hessiana.

Suponhamos que estamos interessados em resolver o problema

$$\min_{x \in \mathbb{R}^n} f : \mathbb{R}^n \rightarrow \mathbb{R}$$

e consideremos o modelo quadrático

$$m_k(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2} (x - x_k)^T B_k (x - x_k),$$

onde B_k é uma aproximação de $\nabla^2 f(x_k)$. Se B_k satisfizer a equação

$$B_k s = y, \tag{2.1}$$

onde $s = x_k - x_{k-1}$ e $y = \nabla f(x_k) - \nabla f(x_{k-1})$, então o modelo quadrático satisfará $m_k(x_k) = f(x_k)$, $\nabla m_k(x_k) = \nabla f(x_k)$ e $\nabla m_k(x_{k-1}) = \nabla f(x_{k-1})$. Se por outro lado quisermos que a manipulação de B_k seja computacionalmente barata, poderíamos pensar em impor uma condição do tipo $B_k = \omega I$. Como muito provavelmente uma matriz desse tipo não irá satisfazer exatamente (2.1), podemos formular o problema de achar a matriz da forma ωI que satisfaça a equação (2.1) no sentido de quadrados mínimos. Assim, temos o seguinte problema

$$\min_{\omega} \|\omega I s - y\|_2^2 \equiv (\omega I s - y)^T (\omega I s - y) = q(\omega)$$

$$q(\omega) = \omega^2 s^T s - 2\omega s^T y + y^T y$$

$$q'(\omega) = 2\omega s^T s - 2s^T y = 0 \Rightarrow \omega^* = \frac{s^T y}{s^T s}$$

Como $q''(\omega) = 2s^T s > 0$ (pois $\|s\|_2^2 > 0$) então ω^* é minimizador. Assim, como $B_k = \omega I$, temos a versão mais popular do passo espectral de Barzilai e Borwein

$$\lambda_k^{(1)} = (\omega^*)^{-1} = \frac{s^T s}{s^T y}.$$

Por outro lado, denotando-se $H_k = B_k^{-1}$ e se H_k satisfizer a equação

$$s = H_k y, \tag{2.2}$$

então ao impor uma condição do tipo $H_k = \theta I$ e formular o problema de achar a matriz da forma θI que satisfaça a equação (2.2) no sentido de quadrados mínimos, temos o seguinte problema

$$\min_{\theta} \|s - \theta I y\|_2^2 \equiv (s - \theta I y)^T (s - \theta I y) = q(\theta)$$

$$q(\theta) = \theta^2 y^T y - 2\theta s^T y + s^T s$$

$$q'(\theta) = 2\theta y^T y - 2s^T y = 0 \Rightarrow \theta^* = \frac{s^T y}{y^T y}$$

Como $q''(\theta) = 2y^T y > 0$ (pois $\|y\|_2^2 > 0$) então θ^* é minimizador. Assim, como $H_k = B_k^{-1}$ e

$H_k = \theta I$, temos uma outra versão do passo espectral

$$\lambda_k^{(2)} = \theta^* = \frac{s^T y}{y^T y}.$$

Portanto, como o método SPG utiliza a versão $\lambda_k^{(1)}$ do passo espectral de Barzilai e Borwein e também utiliza salvaguardas $0 < \lambda_{\min} < \lambda_{\max} < \infty$, então em cada iteração temos

$$\lambda_{k+1} = \min \left(\lambda_{\max}, \max \left(\lambda_{\min}, \frac{s_k^T s_k}{s_k^T y_k} \right) \right).$$

2.2 A busca linear

O método SPG visa minimizar f sobre um conjunto fechado e convexo Ω . O método, assim como sua contraparte irrestrita [53] tem a forma

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

A direção de busca d_k é definida como

$$d_k = P_{\Omega}(x_k - \lambda_k \nabla f(x_k)) - x_k$$

onde P_{Ω} denota a projeção Euclideana sobre Ω . A direção d_k é uma direção de descida o que significa que, se $d_k \neq 0$, tem-se que $f(x_k + \alpha d_k) \ll f(x_k)$ para α suficientemente pequeno. Isto significa que, em princípio, poderia definir métodos convergentes impondo decréscimo suficiente a cada iteração. Entretanto, isto induz a resultados práticos desastrosos. Por esta razão, os métodos espectrais utilizam uma busca linear não monótona que não impõe decréscimo do valor da função a cada iteração. Em [13, 14], a busca linear de Grippo, Lampariello e Lucidi [41] é utilizada. Esta busca linear depende de um parâmetro inteiro $M \geq 1$ e impõe um decréscimo no valor da função a cada M iterações (se $M = 1$ então esta busca linear reduz-se a uma busca linear monótona).

A busca linear baseia-se em uma interpolação quadrática com salvaguarda e visa satisfazer um critério do tipo Armijo com um parâmetro de decréscimo suficiente $\gamma \in (0, 1)$. A rotina de salvaguarda é executada quando o mínimo da quadrática unidimensional não pertence ao

intervalo $[\sigma_1, \sigma_2\alpha]$ (e não quando não pertence ao intervalo $[\sigma_1\alpha, \sigma_2\alpha]$, como geralmente é implementado). Isto significa que, quando a interpolação tende a rejeitar 90% (para $\sigma_1 = 0.1$) do intervalo de busca original ($[0, 1]$), conclui-se que o valor sugerido não é confiável então toma-se $\alpha/2$ como o novo valor. A rotina de busca linear completa está descrita na seção a seguir.

2.3 Algoritmo

Em [13], foram apresentadas duas versões diferentes do método do gradiente projetado: SPG1 e SPG2. De acordo com os experimentos numéricos realizados em [13], não existem diferenças significativas entre os desempenhos de SPG1 e SPG2. A única diferença entre SPG1 e SPG2 está no fato que SPG1 também calcula a projeção toda vez que o critério do tipo Armijo não é aceito (*backtracking*) enquanto que SPG2 faz uma única projeção por iteração. Então SPG2 é indicado nos casos em que projetar tem custo alto. Assim, o algoritmo descrito a seguir considera a versão SPG2 do método.

Dado $\hat{x} \in \mathbb{R}^n$ definimos $P_\Omega(\hat{x})$ como a projeção de \hat{x} no conjunto Ω , ou seja, $P_\Omega(\hat{x}) = \arg \min_{x \in \Omega} \|x - \hat{x}\|$. Denotamos $g(x) = \nabla f(x)$. O algoritmo inicia com $x_0 \in \mathbb{R}^n$ e usa um inteiro $M \geq 1$ relacionado à não monotonicidade da busca linear, salvaguardas $0 < \lambda_{\min} < \lambda_{\max}$ para o passo espectral, um parâmetro de decréscimo suficiente $\gamma \in (0, 1)$ e salvaguardas $0 < \sigma_1 < \sigma_2 < 1$. Inicialmente, $\lambda_0 \in [\lambda_{\min}, \lambda_{\max}]$ é arbitrário.

Algoritmo 1: SPG [13, 14]

Faça $k \leftarrow 0$. Se $x_0 \notin \Omega$, substitua x_0 por $P_\Omega(x_0)$.

Enquanto (o critério de parada não for satisfeito) faça

Calcule $d_k = P_\Omega(x_k - \lambda_k g(x_k)) - x_k$, α_k usando o algoritmo de busca linear descrito abaixo e $x_{k+1} = x_k + \alpha_k d_k$.

Calcule $s_k = x_{k+1} - x_k$, $y_k = g(x_{k+1}) - g(x_k)$ e $\beta_k = \langle s_k, y_k \rangle$.

Se $(\beta_k \leq 0)$ faça $\lambda_{k+1} = \lambda_{\max}$

senão calcule $\lambda_{k+1} = \min(\lambda_{\max}, \max(\lambda_{\min}, \langle s_k, s_k \rangle / \beta_k))$.

Faça $k \leftarrow k + 1$.

fim do enquanto

Faça $x_* \leftarrow x_k$.

Busca Linear

Calcule $f_{\max} = \max \{f(x_{k-j}) \mid 0 \leq j \leq \min(k, M-1)\}$, $x_+ \leftarrow x_k + d_k$,

$\delta \leftarrow \langle g(x_k), d_k \rangle$ e faça $\alpha \leftarrow 1$.

Enquanto $(f(x_+) > f_{\max} + \alpha\gamma\delta)$ faça

Calcule $\alpha_{temp} \leftarrow -\frac{1}{2}\alpha^2\delta/(f(x_+) - f(x_k) - \alpha\delta)$.

Se $(\alpha_{temp} \geq \sigma_1$ e $\alpha_{temp} \leq \sigma_2\alpha)$ faça $\alpha \leftarrow \alpha_{temp}$

Senão faça $\alpha \leftarrow \alpha/2$.

Calcule $x_+ \leftarrow x_k + \alpha d_k$.

fim do enquanto

Faça $\alpha_k \leftarrow \alpha$.

No caso de rejeição do ponto $f(x_k + \alpha d_k)$, os próximos pontos são calculados ao longo da mesma direção. Em particular, o próximo passo testado será o minimizador da quadrática unidimensional $q(\cdot)$, tal que $q(0) = f(x_k)$, $q(\alpha) = f(x_k + \alpha d_k)$ e $\nabla q(0) = \nabla f(x_k)^T d_k$. Caso o minimizador não pertença ao intervalo $[\sigma_1, \sigma_2\alpha]$, tomamos $\alpha/2$.

Capítulo 3

Passos espectrais

O passo espectral foi introduzido por Barzilai e Borwein em [5] e analisado por Raydan em [52]. O passo espectral é um quociente de Rayleigh relacionado com uma média da matriz Hessiana. As duas versões de passos espectrais introduzidas por Barzilai e Borwein podem, de alguma forma, ser interpretadas como aproximações secantes da matriz Hessiana. Desta forma, o método de gradiente espectral utiliza apenas direções de gradiente em cada busca linear.

Desde Barzilai e Borwein, surgiram muitas variações do passo espectral em que utilizam uma aproximação da Hessiana, ωI , para encontrar um λ_k e obter uma direção do tipo quasi-Newton $d_k = -\lambda_k g(x_k)$. No Algoritmo 1 utiliza-se o passo espectral BB1 de Barzilai e Borwein.

A seguir apresentamos, em ordem cronológica, alguns passos espectrais alternativos. Estes passos baseiam-se em variações do passo espectral introduzidas em [5].

3.1 Barzilai e Borwein, 1988 (BB1 e BB2)

Sejam $s_k = x_{k+1} - x_k$, $y_k = g(x_{k+1}) - g(x_k)$ e B_{k+1} uma aproximação de $\nabla^2 f(x_{k+1})$. Dada a equação secante $B_{k+1}s_k = y_k$, impõe-se $B_{k+1} = \omega I$ para encontrar a melhor constante ω que aproxima a equação secante no sentido de quadrados mínimos. Assim, temos que $\omega^* = s_k^T y_k / s_k^T s_k$ é a solução de $\min_{\omega} \|\omega I s_k - y_k\|_2^2$, o que resulta no passo BB1

$$\lambda_{k+1}^{BB1} = (\omega^*)^{-1} = \frac{s_k^T s_k}{s_k^T y_k}$$

Da mesma forma, considerando $H_{k+1} = B_{k+1}^{-1} = \theta I$ e a equação secante $s_k = H_{k+1}y_k$ temos o problema $\min_{\theta} \|s_k - \theta I y_k\|_2^2$ que tem como solução $\theta^* = s_k^T y_k / y_k^T y_k$. Portanto, resulta no passo BB2:

$$\lambda_{k+1}^{BB2} = \theta^* = \frac{s_k^T y_k}{y_k^T y_k}$$

Portanto, a escolha do passo introduzida por Barzilai e Borwein requer menos esforço computacional do que a escolha ótima feita no método de máxima descida de Cauchy [21] e apresenta rapidez na convergência do método para quadráticas. No entanto, somente a utilização do passo BB (BB1 ou BB2) não garante descida na função objetivo. Barzilai e Borwein analisaram a convergência apenas para quadráticas bidimensionais.

3.2 Grippo e Sciandrone, 2002 (ABB-GS)

Em [43], Grippo e Sciandrone propuseram uma nova estratégia de globalização para o método do gradiente de Barzilai e Borwein com a utilização da busca linear não monótona de [41], de técnicas de *watchdog* não monótonas e de uma escolha alternada do passo espectral como segue:

$$\lambda_{k+1} \in \{\lambda_k^{BB1}, \lambda_k^{BB2}\}$$

tal que

$$0 < \theta_l \leq \lambda_{k+1}^{-1} \leq \theta_u,$$

onde θ_l e θ_u são limitantes positivos e definidos por

$$\theta_l = 10^{-5} \max \left\{ 10^{-5}, \frac{\|g(x_k)\|}{1 + \|x_0\|} \right\} \text{ e } \theta_u = 10^{10} \frac{\|g(x_0)\|}{1 + \|x_0\|}$$

Quando em iterações sucessivas ambos λ_k^{BB1} e λ_k^{BB2} satisfazem os limitantes inferior e superior, então alterna-se entre λ_k^{BB1} e λ_k^{BB2} na escolha de λ_{k+1} . E se ambos não satisfazem, então faz-se $\lambda_{k+1} = 1 / \|g(x_k)\|$.

Realizaram testes com problemas irrestritos descritos em [53] e com 95 problemas da coleção CUTE [19].

3.3 Dai e Fletcher, 2005 (ABB-DF)

Interessados em métodos de gradiente projetado com passo de Barzilai e Borwein para resolver problemas de programação quadrática com restrições de caixa, Dai e Fletcher em [26] compararam o método do gradiente projetado utilizando o passo BB1 ao método utilizando alternadamente os passos BB1 e BB2 (ABB) da seguinte forma:

$$\lambda_{k+1}^{ABB} = \begin{cases} \lambda_{k+1}^{BB1}, & \text{quando } k+1 \text{ é ímpar} \\ \lambda_{k+1}^{BB2}, & \text{caso contrário} \end{cases}$$

Para garantir a convergência global, foram utilizadas duas versões de buscas lineares não monótonas: GLL1 e uma versão modificada da busca de Dai e Zhang [30] (modificação: considera $+\infty$ como o valor de referência inicial da função). Os resultados numéricos apontaram que o método PABB com a busca de Dai e Zhang modificada foi o que apresentou melhor desempenho. Também mostraram que as versões que utilizaram GLL1 foram superadas pelas que fizeram uso da busca de Dai e Zhang modificada.

3.4 Dai, 2003 e Dai, Hager, Schittkowski e Zhang, 2006 (CBB)

O passo BB cíclico é caracterizado pela reutilização do passo BB1 para as m iterações consecutivas, como descrito a seguir:

$$\lambda_{mk+i} = \lambda_{mk+1}^{BB1}, \text{ para } i = 1, \dots, m$$

onde $m \geq 1$ é o tamanho do ciclo.

Em [25], Dai estabeleceu a convergência R-linear de CBB para quadráticas fortemente convexas. Nos testes realizados com diferentes valores de m , verificou que o método CBB com $m \in [4, 8]$ obteve performance melhor do que o método que utiliza o passo BB1. Foram utilizados os mesmos problemas de minimização irrestrita descritos em [53].

Em [28], Dai, Hager, Schittkowski e Zhang provaram a convergência R-linear local para o método CBB em um minimizador local de uma função não linear geral (qualquer). Examinaram a velocidade de convergência do método CBB para valores diferentes de $m \in [1, 7]$

aplicado a problemas de programação quadrática. Os resultados mostraram que o método CBB com $m \geq 3$ apresentou melhor eficiência em relação ao método com passo BB1.

3.5 Dai e Fletcher, 2006 (Multi-ponto)

Em [27], Dai e Fletcher propuseram um novo algoritmo de gradiente projetado eficiente com uma nova maneira de calcular o tamanho do passo como segue:

Substituíram o par (s_k, y_k) da fórmula do Passo BB1 por (s_{k-i}, y_{k-i}) com $i \geq 0$. Definiram os vetores $S^{(k)} = (s_k^T, \dots, s_{k-m+1}^T)^T$ e $Y^{(k)} = (y_k^T, \dots, y_{k-m+1}^T)^T$ com o objetivo de calcular a média dos últimos $m \geq 1$ pares diferentes, onde m é um inteiro. Portanto, λ_{k+1} é determinado pela solução do problema $\min_{\omega} \|\omega IS^{(k)} - Y^{(k)}\|_2^2$ e obtém-se

$$\lambda_{k+1} = \frac{S^{(k)T} S^{(k)}}{S^{(k)T} Y^{(k)}} = \frac{\sum_{i=0}^{m-1} s_{k-i}^T s_{k-i}}{\sum_{i=0}^{m-1} s_{k-i}^T y_{k-i}}$$

onde $m = \min(m, \bar{m})$ e \bar{m} é o maior inteiro tal que $s_{k-i}^T y_{k-i} > 0$ para todo $0 \leq i \leq \bar{m}$.

Dados os resultados que obtiveram em [26], utilizaram a mesma busca linear não monótona de Dai e Zhang [30] modificada (difere apenas no valor de referência inicial da função, q é $+\infty$) e realizaram testes com problemas gerados aleatoriamente de dimensão razoavelmente grande para matrizes Hessianas indefinidas e definidas positiva. Os resultados apontaram que o novo passo com $m = 2$ é superior ao passo BB1.

Capítulo 4

Buscas lineares não monótonas

Um dos ingredientes principais dos métodos espectrais é a busca linear não monótona. Em [41], Grippo, Lampariello e Lucidi propuseram uma estratégia de busca linear que permite que a função objetivo aumente em algumas iterações desde que satisfaça o critério de Armijo enquanto mantém a convergência global do método. Desta forma, introduziram uma técnica de busca linear que possibilitou um aumento da velocidade de convergência em relação a técnicas que necessitam de decréscimo monótono da função objetivo para garantir a convergência global. No Algoritmo 1 utiliza-se a busca linear não monótona de [41].

A seguir apresentamos, em ordem cronológica, algumas buscas lineares alternativas. Estas buscas lineares baseiam-se em variações não monótonas do critério de Armijo e ainda garantem convergência global.

4.1 Grippo, Lampariello e Lucidi, 1986 (GLL1)

O esquema de busca linear não monótona descrito a seguir foi desenvolvido por Grippo, Lampariello e Lucidi [41] combinado ao método de Newton. O algoritmo introduzido em [41] segue:

Denota-se $g_k = \nabla f(x_k)$ e $H_k = \nabla^2 f(x_k)$.

Dados x_0 , $M \geq 1$, $c_1 > 0$, $c_2 > 0$, $\gamma \in (0, 1)$ e $\sigma \in (0, 1)$.

Algoritmo: GLL1

Faça $k \leftarrow 0$ e $m_0 \leftarrow 0$.

Calcule g_k .

Enquanto ($g_k \neq 0$) faça

 Calcule H_k .

 Se H_k é não singular e $d = -H_k^{-1}g_k$ satisfaz $|g_k^T d| \geq c_1 \|g_k\|^2$ e $\|d\| \leq c_2 \|g_k\|$ então faça

 Se $g_k^T d < 0$ então faça $d_k \leftarrow d$

 Senão $d_k \leftarrow -d$.

 Senão faça $d_k \leftarrow -g_k$ e $m_k \leftarrow 0$.

 Calcule α_k usando o algoritmo de busca linear descrito abaixo e $x_{k+1} = x_k + \alpha_k d_k$.

 Faça $k \leftarrow k + 1$ e $m_k \leq \min(m_{k-1} + 1, M - 1)$.

fim do enquanto

Faça $x_* \leftarrow x_k$.

Busca Linear

Calcule $f_{\max} = \max \{f(x_{k-j}) \mid 0 \leq j \leq m_k\}$, $x_+ \leftarrow x_k + d_k$, $\delta \leftarrow \langle g_k, d_k \rangle$ e faça $\alpha \leftarrow 1$.

Enquanto ($f(x_+) > f_{\max} + \alpha\gamma\delta$) faça

$\alpha \leftarrow \sigma\alpha$.

 Calcule $x_+ \leftarrow x_k + \alpha d_k$.

fim do enquanto

Faça $\alpha_k \leftarrow \alpha$.

Assim, a busca linear não monótona descrita acima consiste em calcular um passo α_k tal que

$$f(x_k + \alpha_k d_k) \leq \max_{0 \leq j \leq m_k} \{f(x_{k-j})\} + \gamma \alpha_k g_k^T d_k. \quad (4.1)$$

Em [41], recomendam-se os seguintes valores para os parâmetros: $c_1 = 10^{-5}$, $c_2 = 10^5$, $\gamma = 10^{-3}$ e $\sigma = 0.5$. Note que nesta versão original da busca GLL1 o valor de m_k que regula a não monotonicidade varia de iteração para iteração (entre 0 e $M - 1$). Os autores realizaram testes com 5 problemas de minimização irrestrita (Extended Rosenbrock, Wood, Powel singular, Cube, Trigonometric e Helical valley) e observaram que, na maioria destes problemas, o número de buscas lineares e o número de avaliações da função para $M = 10$ são

consideravelmente menores do que os que são necessários para $M = 1$ (busca linear monótona). Assim, concluíram que o uso da busca linear não monótona com o método de Newton apresenta benefícios especialmente nas fases intermediária e final do processo de minimização.

Raydan [53] usou a busca linear não monótona de GLL1 para provar a convergência global do método do Gradiente Espectral de Barzilai e Borwein para minimização irrestrita. Nesse caso, a direção de busca é dada por $d_k = -g_k$. O SPG, para minimização em conjuntos convexos, também usa a busca GLL1 combinada com a direção de busca $d_k = P_\Omega(x_k - \lambda_k g(x_k)) - x_k$.

4.2 Grippo, Lampariello e Lucidi, 1991 (GLL2)

Em [42], Grippo, Lampariello e Lucidi desenvolveram um esquema geral que combina a busca linear não monótona de [41] (que chamamos de GLL1) com a técnica *watchdog* de [22]. Neste esquema, utilizam o relaxamento da propriedade de monotonicidade da GLL1 e permitem que alguns passos possam ser automaticamente aceitos sem que alguma condição do tipo Armijo seja testada, desde que estes passos sejam suficientemente pequenos. Se um decréscimo razoável tiver sido obtido desde a última verificação então o algoritmo é reiniciado a partir da iteração correspondente ao último valor aceitável. O algoritmo introduzido em [42] segue:

Dados x_0 , $M \geq 1$, $c_1 > 0$, $c_2 > 0$, $c_3 > 0$, $N \geq 1$, $\Delta_0 > 0$, $\eta \in (0, 1)$, $\mu > 0$, $\gamma \in (0, 1)$ e $0 < \sigma_1 < \sigma_2 < 1$.

Faça $k \leftarrow 0$, $l \leftarrow 0$, $f_0 \leftarrow f(x_0)$ e $r_0 \leftarrow f(x_0)$.

Calcule g_k .

Enquanto ($\|g_k\| \neq 0$) faça

Encontre d_k tal que $\|g_k\| \geq c_1 \|d_k\|$, $\|d_k\| \geq c_2 \|g_k\|$ e $g_k^T d_k \leq -c_3 \|g_k\|^2$.

Se $k = 0$ ou x_k foi obtido pela utilização de busca linear então faça $d_l \leftarrow d_k$.

Se $k = l + N$ então

Calcule $f(x_k)$ se necessário.

Se $f(x_k) \geq r_k$ então

Faça $x_k \leftarrow x_l$, $d_k \leftarrow d_l$, $\Delta_{k+1} \leftarrow \Delta_k$.

Calcule α_k usando o algoritmo de busca linear descrito abaixo.

Faça $x_{k+1} \leftarrow x_k + \alpha_k d_k$, $l \leftarrow l + 1$, $x_l \leftarrow x_{k+1}$, $f_l \leftarrow f(x_{k+1})$, $r_{k+1} \leftarrow \max_{0 \leq j \leq \min(l, M-1)} \{f_{l-j}\}$.

Senão

Faça $l \leftarrow l + 1$, $x_l \leftarrow x_k$, $d_l \leftarrow d_k$, $f_l \leftarrow f(x_k)$ e $r_k \leftarrow \max_{0 \leq j \leq \min(l, M-1)} \{f_{l-j}\}$.

Se $\|d_k\| \leq \Delta_k$ então

Faça $x_{k+1} \leftarrow x_k + d_k$, $\Delta_{k+1} \leftarrow \eta \Delta_k$, $r_{k+1} \leftarrow r_k$

Senão

Calcule α_k usando o algoritmo de busca linear descrito abaixo.

Faça $x_{k+1} \leftarrow x_k + \alpha_k d_k$, $l \leftarrow l + 1$, $x_l \leftarrow x_{k+1}$, $f_l \leftarrow f(x_{k+1})$

e $r_{k+1} \leftarrow \max_{0 \leq j \leq \min(l, M-1)} \{f_{l-j}\}$.

Senão

Se $\|d_k\| \leq \Delta_k$ então

Se $\|d_k\| > \|d_{k-1}\|$ então

Calcule $f(x_k + d_k)$ e $f(x_k)$ se necessário.

Se $f(x_k + d_k) - f(x_0) \geq \mu (f(x_0) - f(x_k))$ então

faça $x_k \leftarrow x_l$, $d_k \leftarrow d_l$, $\Delta_{k+1} \leftarrow \eta \Delta_k$

Senão

faça $x_{k+1} \leftarrow x_k + d_k$, $\Delta_{k+1} \leftarrow \eta \Delta_k$, $r_{k+1} \leftarrow r_k$.

Senão

faça $x_{k+1} \leftarrow x_k + d_k$, $\Delta_{k+1} \leftarrow \eta \Delta_k$, $r_{k+1} \leftarrow r_k$.

Senão

Calcule $f(x_k)$ se necessário.

Se $f(x_k) \geq r_k$ então

faça $x_k \leftarrow x_l$, $d_k \leftarrow d_l$, $\Delta_{k+1} \leftarrow \Delta_k$

Senão

$l \leftarrow l + 1$, $x_l \leftarrow x_k$, $d_l \leftarrow d_k$, $f_l \leftarrow f(x_k)$, $r_k \leftarrow \max_{0 \leq j \leq \min(l, M-1)} \{f_{l-j}\}$.

Calcule g_{k+1} e faça $k \leftarrow k + 1$.

fim do enquanto

Busca linear

Calcule $x_+ \leftarrow x_k + d_k$, $\delta \leftarrow \langle g_k, d_k \rangle$ e faça $\alpha \leftarrow 1$.

Enquanto $(f(x_+) > r_k + \alpha \gamma \delta)$ faça

Escolha $\sigma \in [\sigma_1, \sigma_2]$.

Faça $\alpha \leftarrow \sigma\alpha$ e calcule $x_+ \leftarrow x_k + \alpha d_k$.

fim do enquanto

Faça $\alpha_k \leftarrow \alpha$.

Sobre o algoritmo:

- (i) O passo $x_{k+1} = x_k + d_k$ (com $\alpha_k = 1$) é automaticamente aceito se $\|d_k\| \leq \Delta_k$ e $\|d_k\| \leq \|d_{k-1}\|$.
- (ii) O índice l representa a última iteração na qual o valor da função objetivo foi avaliado e satisfaz o critério de decréscimo suficiente. Como citado acima, o algoritmo volta a x_l caso a função objetivo não tenha diminuído razoavelmente depois de N aceitações automáticas do passo.
- (iii) No Passo 5, quando $\|d_k\| \leq \Delta_k$ e $\|d_k\| > \|d_{k-1}\|$, parece ser conveniente forçar um controle cuidadoso da monotonicidade dos valores da função objetivo se estes tornam-se significativamente maiores do que o valor inicial $f(x_0)$ – este comportamento combinado com passos grandes podem realmente indicar que os iterados estão deixando a região de confiança.
- (iv) O mecanismo de atualização do valor de referência r_k usa os últimos M valores da função que foram calculados e que não excederam seus valores de referência correspondentes. É importante ressaltar que alguns valores de função podem não ter sido calculados quando passos pequenos foram automaticamente aceitos.

Em [42], utiliza-se tanto a direção de busca usada em [41] quanto uma determinada por uma fatoração de Cholesky modificada da Hessiana. Recomendam-se os seguintes valores para os parâmetros: $\mu = 10^6$, $\eta = 10^{-3}$, $M = 10$, $\Delta_0 = 1$, $N = 10$ e $\gamma = 10^{-3}$. Os autores realizaram testes com 18 problemas de minimização irrestrita sugeridos em [49] e com mais 15 problemas irrestritos descritos em [42] e considerados “difíceis” (mal escalados ou com não linearidades difíceis). Concluíram que a estratégia proposta em [42] agregada a um método de Newton modificado, se comparada com uma técnica de busca linear padrão, pode permitir ganhos computacionais consideráveis, especialmente na solução de problemas “difíceis”.

4.3 Toint, 1996

Toint [57] comparou as buscas GLL1 [41] e GLL2 [42] com um método clássico de busca linear monótona, que nomeou de BLS, em um conjunto de 327 problemas irrestritos da coleção CUTE [19] e concluiu que o melhor desempenho foi observado para o método GLL2. Em seguida, propôs uma variante da busca GLL2. A motivação da mudança está relacionada ao desejo de controlar a não monotonicidade de acordo com quanto a função objetivo se assemelha do modelo quadrático.

A variação da GLL2 proposta por Toint [57] está relacionada, basicamente, com a forma em que r_k é calculado em cada iteração do algoritmo GLL2. Assim, na variante proposta por Toint, o valor de r_k é redefinido como segue:

- Nas primeiras M iterações, $r_k = f(x_k)$
- Para todas as iterações seguintes, r_k é dado por

$$r_k = f(x_k) + \left(\prod_{j=1}^i \alpha_{l-j} \right)^{\frac{1}{i}} [f(x_{l-i}) - f(x_k)]$$

onde

$$i = \arg \max_{1 \leq j \leq M} \{f(x_{l-j})\}.$$

A direção de busca d_k é determinada pela resolução de um sistema de Newton “perturbado” na forma

$$(H_k + M_k)d_k = -g_k + h_k$$

onde H_k denota $\nabla^2 f(x_k)$, M_k é uma matriz de perturbação simétrica semidefinida positiva tal que os autovalores de $H_k + M_k$ estão no intervalo $[k_1, k_2]$ (para constantes k_1 e k_2 independentes de k) e h_k satisfaz as condições

$$d_k^T h_k = 0$$

e

$$\|h_k\| \leq \|g_k\| \min(0.1, \|g_k\|^{0.5}).$$

A matriz M_k foi escolhida pelo método de gradientes conjugados modificado proposto em [1]. Nos testes realizados em [57] pareceu vantajoso forçar a monotonicidade nas primeiras

iterações. Em um número significativo de problemas testados por Toint, o algoritmo com a sua variação necessitou de menos iterações e avaliações de função se comparado com GLL2. Além disso, também notou ser vantajoso usar a variação ao invés de GLL2 quando a avaliação da função objetivo é cara.

4.4 Dai e Zhang, 2001

Dai e Zhang [30] desenvolveram uma busca linear não monótona adaptativa cuja idéia inicial partiu de [58], em que Toint apresentou uma técnica não monótona adaptativa para o método de regiões de confiança. Em [30] a busca linear foi combinada ao método do Gradiente Espectral para minimização irrestrita no qual a direção de busca é dada por $d_k = -g_k$.

Segue o algoritmo de busca linear não monótona adaptativa:

Dados x_0 , $0 < \alpha_{\min} < \alpha_{\max}$, $0 < \sigma_1 < \sigma_2 < 1$, $\epsilon \geq 0$, $P > M > L > 0$, $\gamma_1 \geq 1$ e $\gamma_2 \geq 1$.

Algoritmo

Faça $k \leftarrow 0$, $l \leftarrow 0$, $p \leftarrow 0$.

Faça $f_{\min} \leftarrow f(x_0)$, $f_r \leftarrow f(x_0)$, $f_c \leftarrow f(x_0)$, $f_{\max} \leftarrow f(x_0)$ e $d_0 \leftarrow -g_0$.

Escolha $\alpha \in [\alpha_{\min}, \alpha_{\max}]$.

Enquanto ($\|g_k\| > \epsilon$) faça

Calcule α_k e atualize f_r , f_{\min} , f_c e f_{\max} usando o algoritmo de busca linear descrito abaixo.

Calcule $x_{k+1} \leftarrow x_k + \alpha_k d_k$ e $d_{k+1} \leftarrow -g_{k+1}$.

Calcule $s_k \leftarrow x_{k+1} - x_k$, $y_k \leftarrow g_{k+1} - g_k$ e $\beta_k \leftarrow \langle s_k, y_k \rangle$.

Se ($\beta_k \leq 0$) faça $\alpha \leftarrow \alpha_{\max}$

Senão $\alpha \leftarrow \max(\alpha_{\min}, \min(s_k^T s_k / s_k^T y_k, \alpha_{\max}))$.

Faça $k \leftarrow k + 1$.

fim enquanto

Faça $x_* \leftarrow x_k$.

Busca Linear Adaptativa

Se $l = L$ então

Se $(f_{\max} - f_{\min}) \geq \gamma_1(f_c - f_{\min})$ faça $f_r \leftarrow f_c$

Senão faça $f_r \leftarrow f_{\max}$.

Faça $l \leftarrow 0$.

Se $p > P$ então

Se $f_{\max} > f(x_k)$ e $(f_r - f(x_k)) \geq \gamma_2(f_{\max} - f(x_k))$ faça $f_r \leftarrow f_{\max}$.

Se $f(x_k + \alpha d_k) > f_r + \gamma \alpha \nabla f(x_k) d_k$ então

Faça $p \leftarrow 0$ e $\alpha_{old} \leftarrow \alpha$.

Calcular $\alpha_{new} \in [\sigma_1 \alpha_{old}, \sigma_2 \alpha_{old}]$.

Enquanto $f(x_k + \alpha_{new} d_k) > \min(f_{\max}, f_r) + \gamma \alpha_{new} \nabla f(x_k) d_k$ faça

Faça $\alpha_{old} \leftarrow \alpha_{new}$.

Calcular $\alpha_{new} \in [\sigma_1 \alpha_{old}, \sigma_2 \alpha_{old}]$.

fim do enquanto

Faça $\alpha \leftarrow \alpha_{new}$.

Senão faça $p \leftarrow p + 1$.

Seja $f_{k+1} \leftarrow f(x_k + \alpha d_k)$.

Se $f_{k+1} < f_{\min}$ faça $f_c \leftarrow f_{\min} \leftarrow f_{k+1}$ e $l \leftarrow 0$

Senão faça $l \leftarrow l + 1$.

Se $f_{k+1} > f_c$ faça $f_c \leftarrow f_{k+1}$.

Faça $f_{\max} = \max\{f(x_{k-j}) \mid 0 \leq j \leq \min(k, M-1)\}$ e $\alpha_k \leftarrow \alpha$.

Algumas definições referentes à busca linear de Dai e Zhang descrita acima:

- (i) f_{\min} : é o melhor valor da função objetivo obtido até o momento, ou seja, na k -ésima iteração

$$f_{\min} = \min_{0 \leq j \leq k} \{f(x_j)\}$$

- (ii) l : representa o número de iterações transcorridas desde que o valor de f_{\min} foi encontrado. Mais precisamente, na k -ésima iteração l é dado por:

$$l = \begin{cases} k - i, & \text{se } k - i < L \\ 0, & \text{caso contrário} \end{cases}$$

onde

$$i = \arg \min_{0 \leq j \leq k} \{f(x_j)\} \quad (4.2)$$

(iii) f_c : é o maior valor da função objetivo desde que f_{\min} foi encontrado. Assim,

$$f_c = \max_{i \leq j \leq k} \{f(x_j)\},$$

onde i é como definido em (4.2).

(iv) f_{\max} : é o maior valor da função nas últimas M iterações, ou seja, na k -ésima iteração

$$f_{\max} = \max_{0 \leq j \leq \min(k, M-1)} \{f(x_{k-j})\}$$

(v) f_r : é o valor de referência usado na condição não monótona de Armijo. De acordo com as condições de teste do algoritmo acima, f_r poderá receber o valor de f_{\max} , de f_c ou permanecer com o mesmo valor.

(vi) p : é o número de vezes consecutivas em que o primeiro passo tentativo foi aceito (ou seja, não houve “backtracking”).

Pelo Passo 3 do algoritmo podemos observar que se o primeiro passo tentativo não é aceito então f_r é trocado por $\min(f_{\max}, f_r)$. Esta troca representa que a busca linear foi desenvolvida para que o primeiro passo tentativo seja aceito mais facilmente do que os outros passos tentativos. Vale ressaltar que o método do gradiente espectral, combinado à busca linear de Dai e Zhang, apresenta convergência global.

Dai e Zhang [30] sugerem que os inteiros P , M e L sejam escolhidos de tal forma que $L \leq 5$ e $P \geq 4M \geq 8L$ e que as constantes γ_1 e γ_2 sejam, respectivamente, $\gamma_1 = M/L$ e $\gamma_2 = P/M$.

4.5 Zhang e Hager, 2004

A condição da busca linear não monótona de Zhang e Hager [62] tem a mesma forma de (4.1) apresentada em [41], exceto que o “max” é trocado por uma média de valores da função. A busca linear de Zhang e Hager foi combinada ao método quasi-Newton L-BFGS [48, 50]. Foram feitas comparações usando 80 problemas de otimização irrestrita da coleção CUTE [19]. Mais precisamente, segue abaixo o algoritmo de busca linear não monótona dos autores para determinar α_k .

Dados $x_k, d_k, 0 \leq \eta_{\min} \leq \eta_{\max} \leq 1, \gamma \in (0, 1), 0 < \sigma_1 < \sigma_2 < 1, C_0 = f(x_0)$ e $Q_0 = 1$.

Busca Linear

Calcule $x_+ \leftarrow x_k + d_k, \delta \leftarrow \langle g(x_k), d_k \rangle$ e faça $\alpha \leftarrow 1$.

Enquanto $(f(x_+) > C_k + \alpha\gamma\delta)$ faça

Faça $\alpha_{old} \leftarrow \alpha$.

Calcule $\alpha \in [\sigma_1\alpha_{old}, \sigma_2\alpha_{old}]$.

Calcule $x_+ \leftarrow x_k + \alpha d_k$.

fim do enquanto

Calcule $\eta_k \in [\eta_{\min}, \eta_{\max}]$.

Faça $Q_{k+1} \leftarrow \eta_k Q_k + 1$ e $C_{k+1} \leftarrow (\eta_k Q_k C_k + f(x_{k+1}))/Q_{k+1}$.

Faça $\alpha_k \leftarrow \alpha$.

Podemos observar que C_{k+1} é uma combinação convexa de C_k e $f(x_{k+1})$ e que a escolha de η_k controla o grau de não monotonicidade. Assim, se $\eta_k = 0$ para todo k , temos uma busca linear monótona. Já, se $\eta_k = 1$ para todo k , então C_k é a média da soma dos valores da função.

A convergência global do método foi provada pelos autores. De acordo com os experimentos numéricos realizados em [62], os melhores resultados foram encontrados pela variação dinâmica de η_k . Porém, também foram obtidos resultados razoavelmente bons para uma classe ampla de problemas utilizando um valor fixo $\eta_k = 0.85$.

4.6 La Cruz, Martínez e Raydan, 2005

Em [45], La Cruz, Martínez e Raydan combinaram e estenderam as condições GLL [41] e de Li e Fukushima [47] para produzir uma busca linear não monótona robusta que de algum

modo leva em conta as vantagens de ambos esquemas. A busca linear proposta por La Cruz, Martínez e Raydan foi incorporada a um método quasi-Newton sem derivadas para resolver sistemas não lineares. A aproximação do Jacobiano tem a forma

$$B_k = \sigma_k I$$

com

$$\sigma_k = \frac{s_k^T s_k}{y_k^T s_k}, \quad s_k = x_{k+1} - x_k \text{ e } y_k = F(x_{k+1}) - F(x_k).$$

Resumidamente, a nova condição de descida pode ser escrita como:

Dados $M \geq 1$, $\gamma \in (0, 1)$ e $0 < \sigma_1 < \sigma_2 < 1$.

$$f(x_k + \alpha_k d_k) \leq \max_{0 \leq j \leq \min(k, M-1)} \{f(x_{k-j})\} + \zeta_k - \gamma \alpha_k^2 f(x_k)$$

onde

$$\alpha_k = \sigma \alpha_k \text{ com } \sigma \in [\sigma_1, \sigma_2],$$

$$\sum_{k=0}^{\infty} \zeta_k < \infty$$

e supondo que $\zeta_k > 0$ para todo $k \in \mathbb{N}$.

O termo GLL $\max_{0 \leq j \leq \min(k, M-1)} \{f(x_{k-j})\}$ é responsável pelo comportamento suficientemente não monótono de $f(x_k)$ quando k é grande. Por outro lado, a presença de $\zeta_k > 0$ garante que todas as iterações estejam bem definidas e o termo $-\gamma \alpha_k^2 f(x_k)$ produz os argumentos para provar convergência global.

Capítulo 5

Experimentos

Neste capítulo apresentamos os resultados numéricos dos experimentos comparando as combinações dos diferentes passos espectrais apresentados no Capítulo 3 com as buscas lineares não monótonas descritas no Capítulo 4 no contexto do método do Gradiente Espectral Projetado.

As versões de passos espectrais utilizados foram:

- BB1 (Barzilai e Borwein, 1988),
- ABB-DF (Dai e Fletcher, 2005),
- ABB-GS (Grippo e Sciandrone, 2002),
- CBB (Dai, 2003 e Dai, Hager, Schittkowski e Zhang, 2006) com tamanho de ciclo $m = 4$,
- Multi-ponto (Dai e Fletcher, 2006) com $m = 2$.

Seguem abaixo as buscas lineares não monótonas utilizadas e seus respectivos parâmetros:

- GLL1 (Grippo, Lampariello e Lucidi, 1986) com $M = 10$,
- GLL2 (Grippo, Lampariello e Lucidi, 1991) com $M = 10$,
- TOINT (Toint, 1996) com $M = 10$,
- DZ (Dai e Zhang, 2001) com $L = 5$, $M = 10$ e $P = 40$,

- ZH1 (Zhang e Hager, 2004) com $\eta_{\min} = 0.1$, $\eta_{\max} = 0.95$ e η_k calculado dinamicamente da seguinte forma:

$$\rho = (\max(\epsilon, \min(\|\nabla f(x_k)\|_{\infty}, \|\nabla f(x_0)\|_{\infty})) - \epsilon) / (\|\nabla f(x_0)\|_{\infty} - \epsilon)$$

$$\eta_k = \rho \eta_{\min} + (1 - \rho) \eta_{\max},$$

- ZH2 (Zhang e Hager, 2004) com valor fixo $\eta_k = 0.85$,
- LMR (La Cruz, Martínez e Raydan, 2005) com $M = 10$ e $\zeta_k = \max(|f(x_0)|, 1) / (k + 1)^2$.

Assim, foram definidas 35 versões diferentes do método como segue:

1. GLL1-1: método utilizando a busca linear GLL1 e o passo BB1
2. GLL1-2: método utilizando a busca linear GLL1 e o passo ABB-DF
3. GLL1-3: método utilizando a busca linear GLL1 e o passo ABB-GS
4. GLL1-4: método utilizando a busca linear GLL1 e o passo CBB
5. GLL1-5: método utilizando a busca linear GLL1 e o passo Multi-ponto
6. GLL2-1: método utilizando a busca linear GLL2 e o passo BB1
7. GLL2-2: método utilizando a busca linear GLL2 e o passo ABB-DF
8. GLL2-3: método utilizando a busca linear GLL2 e o passo ABB-GS
9. GLL2-4: método utilizando a busca linear GLL2 e o passo CBB
10. GLL2-5: método utilizando a busca linear GLL2 e o passo Multi-ponto
11. TOINT-1: método utilizando a busca linear TOINT e o passo BB1
12. TOINT-2: método utilizando a busca linear TOINT e o passo ABB-DF
13. TOINT-3: método utilizando a busca linear TOINT e o passo ABB-GS
14. TOINT-4: método utilizando a busca linear TOINT e o passo CBB

15. TOINT-5: método utilizando a busca linear TOINT e o passo Multi-ponto
16. DZ-1: método utilizando a busca linear DZ e o passo BB1
17. DZ-2: método utilizando a busca linear DZ e o passo ABB-DF
18. DZ-3: método utilizando a busca linear DZ e o passo ABB-GS
19. DZ-4: método utilizando a busca linear DZ e o passo CBB
20. DZ-5: método utilizando a busca linear DZ e o passo Multi-ponto
21. ZH1-1: método utilizando a busca linear ZH1 e o passo BB1
22. ZH1-2: método utilizando a busca linear ZH1 e o passo ABB-DF
23. ZH1-3: método utilizando a busca linear ZH1 e o passo ABB-GS
24. ZH1-4: método utilizando a busca linear ZH1 e o passo CBB
25. ZH1-5: método utilizando a busca linear ZH1 e o passo Multi-ponto
26. ZH2-1: método utilizando a busca linear ZH2 e o passo BB1
27. ZH2-2: método utilizando a busca linear ZH2 e o passo ABB-DF
28. ZH2-3: método utilizando a busca linear ZH2 e o passo ABB-GS
29. ZH2-4: método utilizando a busca linear ZH2 e o passo CBB
30. ZH2-5: método utilizando a busca linear ZH2 e o passo Multi-ponto
31. LMR-1: método utilizando a busca linear LMR e o passo BB1
32. LMR-2: método utilizando a busca linear LMR e o passo ABB-DF
33. LMR-3: método utilizando a busca linear LMR e o passo ABB-GS
34. LMR-4: método utilizando a busca linear LMR e o passo CBB
35. LMR-5: método utilizando a busca linear LMR e o passo Multi-ponto

Todos os experimentos foram realizados em um computador pessoal com processador AMD Athlon(tm) 64 Processor 3200+, 2.15-GHz e 1GB de memória RAM. Os algoritmos foram programados em Fortran77 e compilados com a opção -O3 para otimizar o código.

O conjunto de problemas utilizados em nossos experimentos é composto por todos os problemas irrestritos e com restrições de caixa da coleção CUTE [19, 39] e com as dimensões (número de variáveis) *default*. São 268 problemas (com dimensão média de aproximadamente 2417) sendo que 158 são irrestritos e 110 com restrições de caixa e estão listados nas Tabelas 5.1 e 5.2, respectivamente.

Após a realização de experimentos preliminares, observamos que para alguns dos problemas do conjunto de testes a rotina de avaliação do gradiente da função objetivo (rotina *ugr* que acompanha a coleção CUTEr) apresentou uma certa instabilidade numérica.

Usamos o critério de convergência dado por

$$\|g(x_k)\|_\infty \leq \epsilon.$$

O algoritmo também pára quando o tempo de CPU é maior do que 300 segundos.

Nos experimentos que realizamos, escolhemos $\lambda_{\min} = 10^{-10}$, $\lambda_{\max} = 10^{10}$, $\sigma_1 = 0.1$, $\sigma_2 = 0.9$, $\epsilon = 10^{-6}$ e $\gamma = 10^{-4}$.

O critério a seguir foi utilizado para comparar os 35 métodos compostos pelas combinações de passos espectrais e buscas lineares não monótonas:

Considere um determinado problema e seja $x_{\text{final}}^{(M)}$, $M = 1, \dots, 35$, o ponto final do método M aplicado a este problema. Definimos

$$f_{\text{best}} = \min_M \left\{ f \left(x_{\text{final}}^{(M)} \right) \right\}.$$

Dizemos que o método M encontrou uma solução do problema se $x_{\text{final}}^{(M)}$ é viável e

$$f \left(x_{\text{final}}^{(M)} \right) \leq f_{\text{best}} + 10^{-3} |f_{\text{best}}| + 10^{-6}.$$

Seja $t^{(M)}$, $M = 1, \dots, 35$, o tempo de CPU (em segundos) que o método M usou para

Problema	n	Problema	n	Problema	n	Problema	n
3PK	30	DIXMAANG	3000	HIMMELBB	2	PFIT1LS	3
AKIVA	2	DIXMAANH	3000	HIMMELBF	4	PFIT2LS	3
ALLINITU	4	DIXMAANI	3000	HIMMELBG	2	PFIT3LS	3
ARGLINA	200	DIXMAANJ	3000	HIMMELBH	2	PFIT4LS	3
ARGLINB	200	DIXMAANK	15	HUMPS	2	POWELLSG	5000
ARGLINC	200	DIXMAANL	3000	HYDC20LS	99	POWER	10000
ARWHEAD	5000	DIXON3DQ	10000	INDEF	5000	QUARTC	5000
BARD	3	DJTL	2	JENSMP	2	ROSENBR	2
BDQRTIC	5000	DQDRTIC	5000	JIMACK	3549	S308	2
BEALE	2	DQRTIC	5000	KOWOSB	4	SBRYBND	5000
BIGGS6	6	EDENSCH	2000	LIARWHD	5000	SCHMVETT	5000
BOX3	3	EG2	1000	LOGHAIRY	2	SCOSINE	5000
BRKMCC	2	EIGENALS	2550	MANCINO	100	SCURLY10	10000
BROWNAL	200	EIGENBLS	2550	MARATOSB	2	SCURLY20	10000
BROWNBS	2	EIGENCLS	2652	MEXHAT	2	SCURLY30	10000
BROWNDEN	4	ENGVAL1	5000	MEYER3	3	SENSORS	100
BROYDN7D	5000	ENGVAL2	3	MODBEALE	20000	SINEVAL	2
BRYBND	5000	ERRINROS	50	MOREBV	5000	SINQUAD	5000
CHAINWOO	4000	EXPFIT	2	NCB20	5010	SISSER	2
CHNROSNB	50	EXTROSNB	1000	NCB20B	5000	SNAIL	2
CLIFF	2	FLETCBV2	5000	NONCVXU2	5000	SPARSINE	5000
COSINE	10000	FLETCBV3	5000	NONCVXUN	5000	SPARSQR	10000
CRAGGLVY	5000	FLETCHBV	5000	NONDIA	5000	SROSENBR	5000
CUBE	2	FLETCHCR	1000	NONDQUAR	5000	STRATEC	10
CURLY10	10000	FMINSRF2	5625	NONMSQRT	4900	TESTQUAD	5000
CURLY20	10000	FMINSURF	5625	OSBORNEA	5	TOINTGOR	50
CURLY30	10000	FREUROTH	5000	OSBORNEB	11	TOINTGSS	5000
DECONVU	61	GENHUMPS	5000	PALMER1C	8	TOINTPSP	50
DENSCHNA	2	GENROSE	500	PALMER1D	7	TOINTQOR	50
DENSCHNB	2	GROWTHLS	3	PALMER2C	8	TQUARTIC	5000
DENSCHNC	2	GULF	3	PALMER3C	8	TRIDIA	5000
DENSCHND	3	HAIRY	2	PALMER4C	8	VARDIM	200
DENSCHNE	3	HATFLDD	3	PALMER5C	6	VAREIGVL	50
DENSCHNF	2	HATFLDE	3	PALMER6C	8	VIBRBEAM	8
DIXMAANA	3000	HEART6LS	6	PALMER7C	8	WATSON	12
DIXMAANB	3000	HEART8LS	8	PALMER8C	8	WOODS	4000
DIXMAANC	3000	HELIX	3	PARKCH	15	YFITU	3
DIXMAAND	3000	HIELOW	3	PENALTY1	1000	ZANGWIL2	2
DIXMAANE	3000	HILBERTA	2	PENALTY2	200		
DIXMAANF	3000	HILBERTB	10	PENALTY3	200		

Tabela 5.1: Problemas irrestritos - CUTE

encontrar uma solução. Se o método não encontrou uma solução então definimos $t^{(M)} = \infty$.

Problema	n	Problema	n	Problema	n	Problema	n
ALLINIT	4	HS2	2	ODNAMUR	11130	PENTDI	5000
BDEXP	5000	HS38	4	OSLBQP	8	PROBPENL	500
BIGGSB1	5000	HS3MOD	2	PALMER1A	6	PSPDOC	4
BLEACHNG	17	HS3	2	PALMER1B	4	QR3DLS	610
BQP1VAR	1	HS45	5	PALMER1E	8	QRTQUAD	5000
BQPGABIM	50	HS4	2	PALMER1	4	QUDLIN	5000
BQPGASIM	50	HS5	2	PALMER2A	6	S368	8
BQPGAUSS	2003	JNLBRNG1	10000	PALMER2B	4	SCOND1LS	5002
CAMEL6	2	JNLBRNG2	10000	PALMER2E	8	SIM2BQP	2
CHARDIS0	2000	JNLBRNGA	10000	PALMER2	4	SIMBQP	2
CHEBYQAD	100	JNLBRNGB	10000	PALMER3A	6	SINEALI	1000
CHENHARK	5000	KOEBHEL	3	PALMER3B	4	SPECAN	9
CVXBQP1	10000	LINVERSE	1999	PALMER3E	8	TORSION1	5476
DECONVB	61	LOGROS	2	PALMER3	4	TORSION2	5476
EG1	3	MAXLIKA	8	PALMER4A	6	TORSION3	5476
EXPLIN2	1200	MCCORMCK	5000	PALMER4B	4	TORSION4	5476
EXPLIN	1200	MDHOLE	2	PALMER4E	8	TORSION5	5476
EXPQUAD	1200	MINSURFO	5306	PALMER4	4	TORSION6	5476
HADAMALS	400	NCVXBQP1	10000	PALMER5A	8	TORSIONA	5476
HARKERP2	5000	NCVXBQP2	10000	PALMER5B	9	TORSIONB	5476
HART6	6	NCVXBQP3	10000	PALMER5D	8	TORSIONC	5476
HATFLDA	4	NOBNDTOR	5476	PALMER5E	8	TORSIOND	5476
HATFLDB	4	NONSCOMP	5000	PALMER6A	6	TORSIONE	5476
HATFLDC	25	OBSTCLAE	10000	PALMER6E	8	TORSIONF	5476
HIMMELP1	2	OBSTCLAL	10000	PALMER7A	6	WEEDS	3
HS110	10	OBSTCLBL	10000	PALMER7E	8	YFIT	3
HS1	2	OBSTCLBM	10000	PALMER8A	6		
HS25	3	OBSTCLBU	10000	PALMER8E	8		

Tabela 5.2: Problemas com restrições de caixa - CUTE

Assim definimos

$$t_{\text{best}} = \min_M \{t^{(M)} \mid \text{método } M \text{ encontrou uma solução}\},$$

e dizemos que o método M é *um dos métodos mais rápidos* para o problema quando

$$t^{(M)} \leq t_{\text{best}} < \infty.$$

Nenhum dos 35 métodos gerados pelas combinações de passos espectrais e buscas lineares não monótonas diferentes conseguiu atingir o critério de convergência para todos os 268 problemas de teste. A Tabela 5.3 mostra para cada método o número de problemas em que o critério

	BB1	ABB-DF	ABB-GS	CBB	Multi-ponto
DZ	225	229	228	197	229
GLL1	224	229	228	199	232
LMR	222	226	223	183	227
ZH1	225	230	229	88	232
ZH2	224	231	230	105	231
GLL2	211	230	228	200	232
TOINT	213	228	229	202	232

Tabela 5.3: Total de problemas em que o método satisfaz o critério de convergência.

de convergência foi satisfeito. Cada linha representa uma busca linear e cada coluna um passo espectral e cada combinação de linha e coluna representa um método diferente.

Analisamos o desempenho dos métodos utilizando o perfil de desempenho de Dolan e Moré [32] da seguinte maneira: construímos um gráfico considerando a fração p de problemas (eixo p) em que algum método obteve um fator t (eixo t) do melhor tempo. Em um gráfico de perfil de desempenho, o topo da curva representa o método que resolveu a maioria dos problemas em um tempo com um fator t do melhor tempo. O percentual de problemas em que um método é mais rápido (Eficiência) é dado pelo lado esquerdo do gráfico. O lado direito do gráfico informa o percentual dos problemas que foram resolvidos com sucesso por cada um dos métodos. Assim, o lado direito é uma medida de Robustez de um método.

A seguir, comparamos as 35 versões dos métodos agrupadas como segue:

1. Comparação entre todos os métodos com o mesmo passo espectral
2. Comparação entre todos os métodos com a mesma busca linear
3. Comparação entre todos os métodos
4. Comparação entre os métodos que apresentaram desempenho entre os 12 melhores métodos tanto no critério Eficiência quanto em Robustez
5. Comparação entre o melhor método (da comparação do item 4) e o método SPG.

As Figuras 5.1-5.5 apresentam a comparação entre os métodos utilizando o tempo de CPU para analisar o perfil de desempenho. As Tabelas 5.4-5.8 mostram o resumo do desempenho comparativo dos 35 métodos divididos em grupos com a mesma versão de passo espectral combinada às diferentes buscas lineares não monótonas.

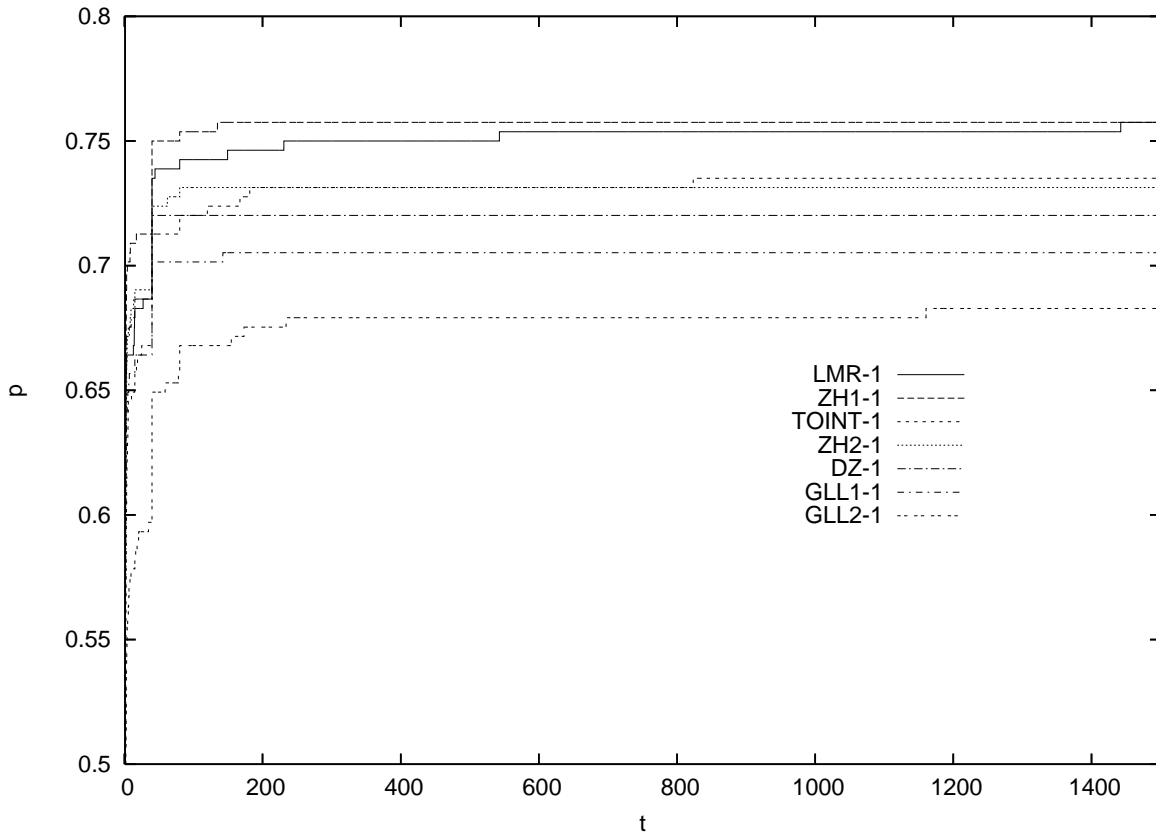


Figura 5.1: Curva do perfil de desempenho comparando os 7 métodos que utilizam o passo BB1.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
LMR-1	75.75	43.28	LMR-1	75.75	43.28
ZH1-1	75.75	40.30	DZ-1	72.02	43.28
TOINT-1	73.51	28.73	ZH1-1	75.75	40.30
ZH2-1	73.13	30.22	GLL1-1	70.52	37.31
DZ-1	72.02	43.28	ZH2-1	73.13	30.22
GLL1-1	70.52	37.31	TOINT-1	73.51	28.73
GLL2-1	68.28	26.49	GLL2-1	68.28	26.49

Tabela 5.4: Desempenho dos 7 métodos que utilizam o passo BB1 - Robustez x Eficiência.

Comparando-se os 7 métodos que utilizam o passo BB1, os métodos *LMR-1* e *ZH1-1* resolveram a maioria dos problemas. Enquanto que os métodos *LMR-1* e *DZ-1* foram os mais eficientes. Assim, os resultados numéricos leva-nos a concluir que o passo espectral BB1 combinado à busca linear LMR é melhor do que se for combinado às outras busca lineares.

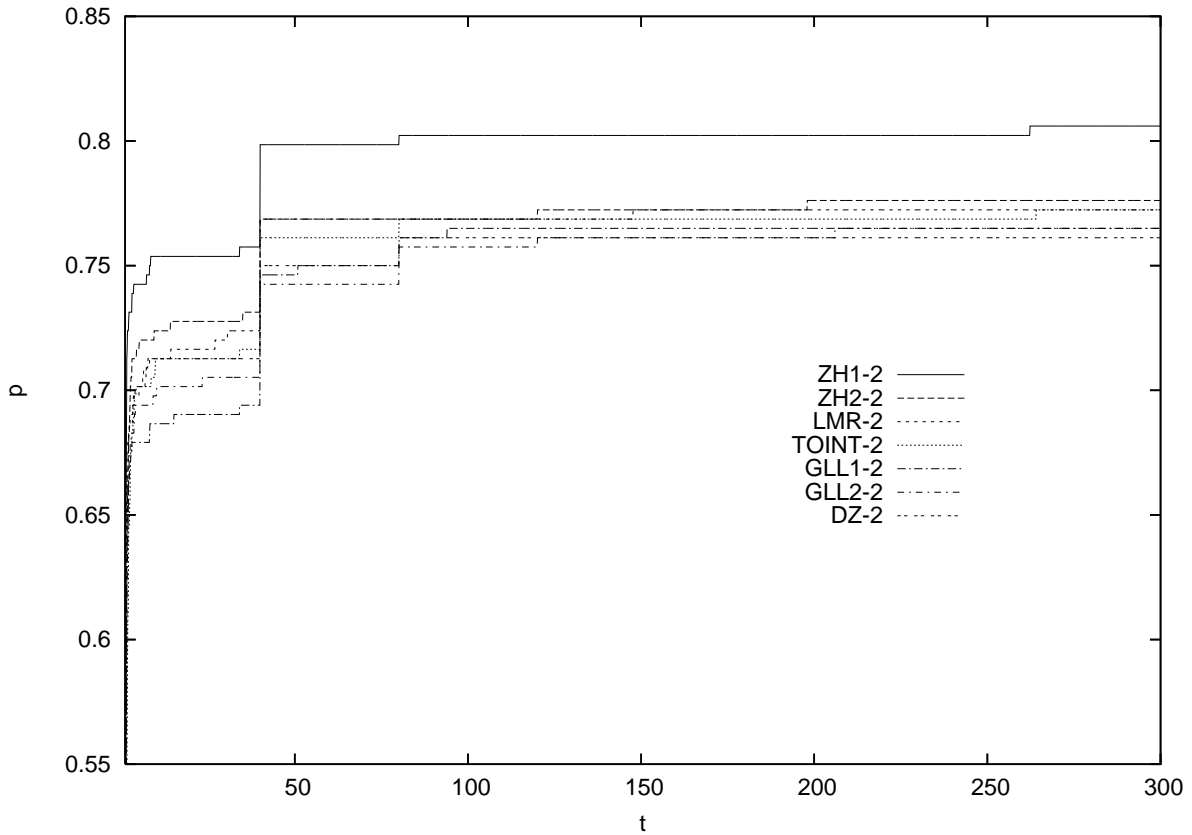


Figura 5.2: Curva do perfil de desempenho comparando os 7 métodos que utilizam o passo ABB-DF.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
ZH1-2	80.60	44.03	LMR-2	77.24	45.52
ZH2-2	77.99	32.84	ZH1-2	80.60	44.03
LMR-2	77.24	45.52	DZ-2	76.12	43.66
TOINT-2	77.24	31.34	GLL1-2	76.49	38.43
GLL1-2	76.49	38.43	ZH2-2	77.99	32.84
GLL2-2	76.49	29.85	TOINT-2	77.24	31.34
DZ-2	76.12	43.66	GLL2-2	76.49	29.85

Tabela 5.5: Desempenho dos 7 métodos que utilizam o passo ABB-DF - Robustez x Eficiência.

Comparando-se os 7 métodos que utilizam o passo ABB-DF, o método *ZH1-2* resolveu a maior parte dos problemas e perdeu no critério de eficiência apenas para o método *LMR-2*. Considerando que a diferença na eficiência entre eles é pequena, podemos concluir que o passo espectral ABB-DF combinado à busca linear *ZH1* obteve, no geral, um desempenho melhor.

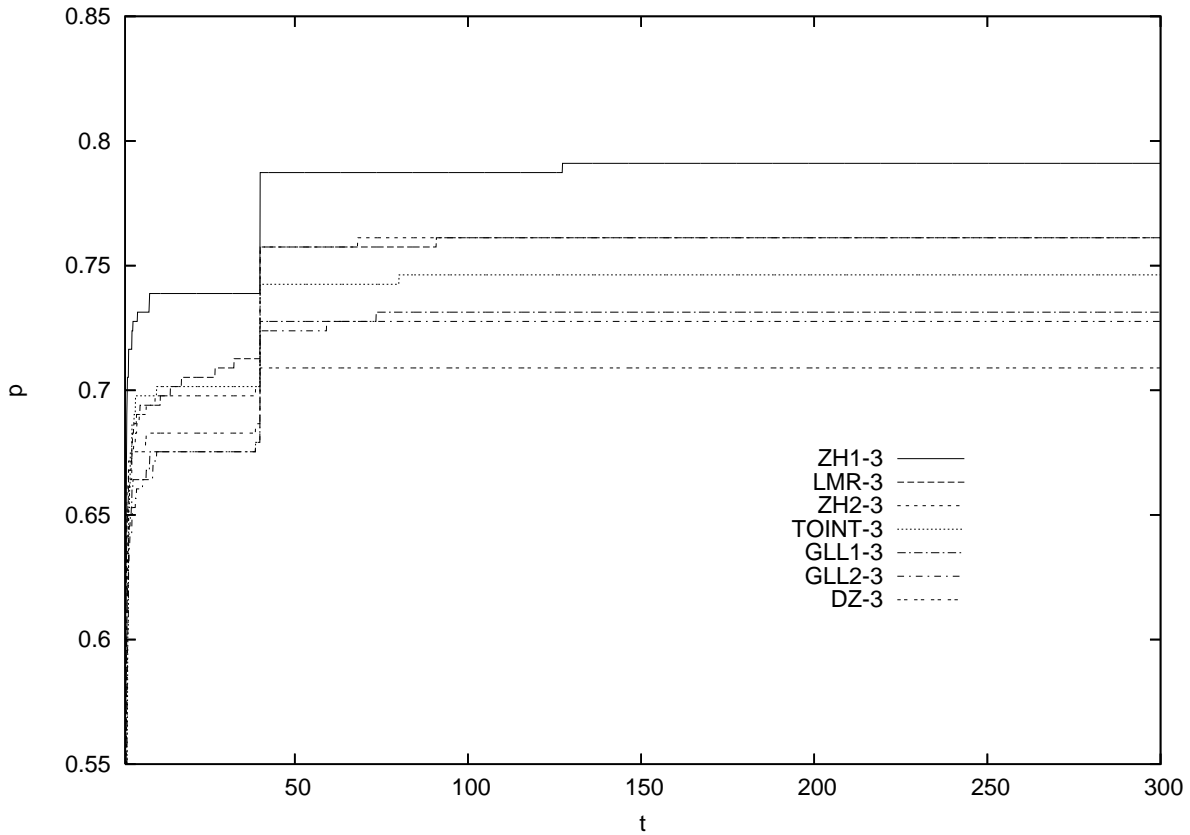


Figura 5.3: Curva do perfil de desempenho comparando os 7 métodos que utilizam o passo ABB-GS.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
ZH1-3	79.10	43.66	LMR-3	76.12	45.90
LMR-3	76.12	45.90	ZH1-3	79.10	43.66
ZH2-3	76.12	30.97	DZ-3	71.27	41.79
TOINT-3	75.37	29.85	GLL1-3	73.51	39.18
GLL1-3	73.51	39.18	ZH2-3	76.12	30.97
GLL2-3	73.51	26.49	TOINT-3	75.37	29.85
DZ-3	71.27	41.79	GLL2-3	73.51	26.49

Tabela 5.6: Desempenho dos 7 métodos que utilizam o passo ABB-GS - Robustez x Eficiência.

Comparando-se os 7 métodos que utilizam o passo ABB-GS, o método *ZH1-3* resolveu a maior parte dos problemas e perdeu no critério de eficiência apenas para o método *LMR-3*. Usando a Robustez como desempate, temos que o passo espectral ABB-GS combinado à busca linear ZH1 apresentou um desempenho melhor.

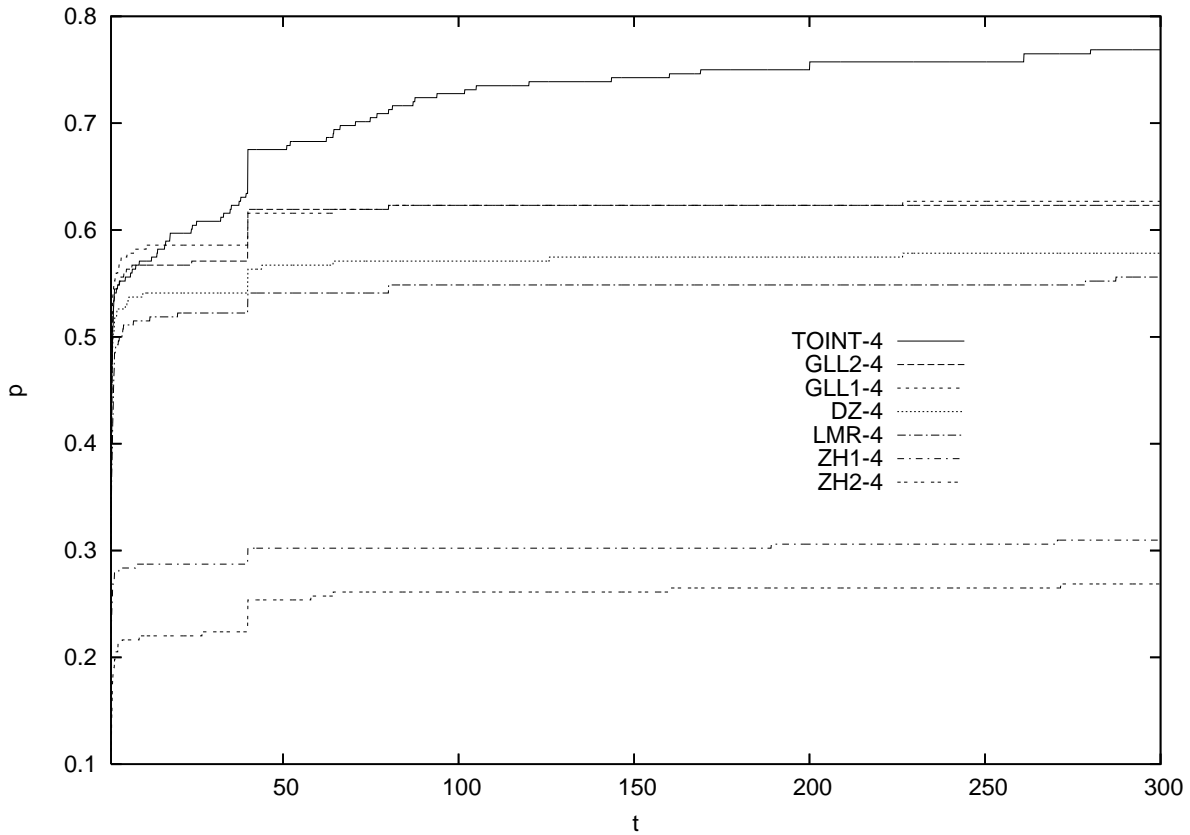


Figura 5.4: Curva do perfil de desempenho comparando os 7 métodos que utilizam o passo CBB.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
TOINT-4	81.72	39.93	GLL1-4	63.06	40.30
GLL2-4	63.43	30.60	TOINT-4	81.72	39.93
GLL1-4	63.06	40.30	DZ-4	57.84	32.46
DZ-4	57.84	32.46	GLL2-4	63.43	30.60
LMR-4	55.97	30.60	LMR-4	55.97	30.60
ZH1-4	31.34	20.52	ZH1-4	31.34	20.52
ZH2-4	27.99	11.19	ZH2-4	27.99	11.19

Tabela 5.7: Desempenho dos 7 métodos que utilizam o passo CBB - Robustez x Eficiência.

Comparando-se os 7 métodos que utilizam o passo CBB, o método *TOINT-4* resolveu a maior parte dos problemas e perdeu no critério de eficiência apenas para o método *GLL1-4*. Usando a Robustez como desempate, temos que o passo espectral CBB combinado à busca linear TOINT apresentou um desempenho melhor.

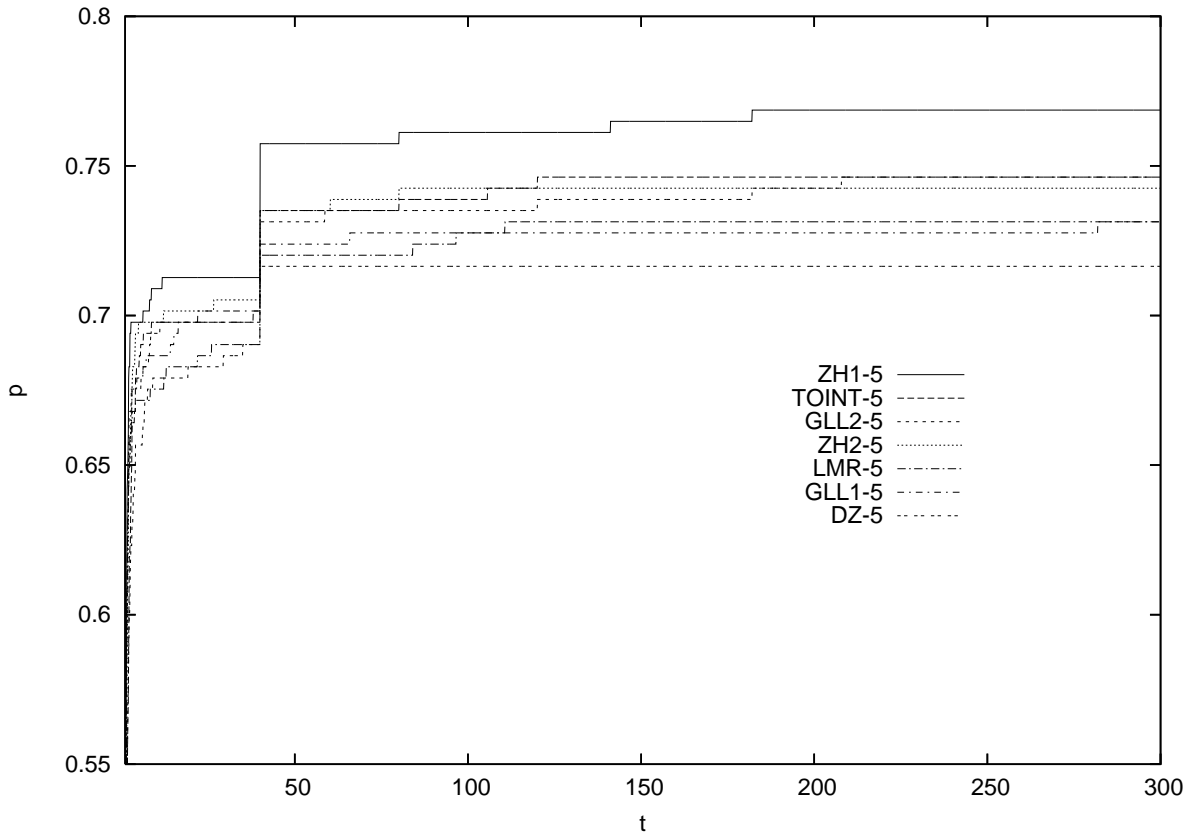


Figura 5.5: Curva do perfil de desempenho comparando os 7 métodos que utilizam o passo Multi-ponto.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
ZH1-5	77.24	41.42	LMR-5	73.51	44.40
TOINT-5	75.37	31.72	ZH1-5	77.24	41.42
GLL2-5	75.00	29.48	GLL1-5	73.51	39.93
ZH2-5	74.63	32.84	DZ-5	72.02	38.81
LMR-5	73.51	44.40	ZH2-5	74.63	32.84
GLL1-5	73.51	39.93	TOINT-5	75.37	31.72
DZ-5	72.02	38.81	GLL2-5	75.00	29.48

Tabela 5.8: Desempenho dos 7 métodos que utilizam o passo Multi-ponto - Robustez x Eficiência.

Comparando-se os 7 métodos que utilizam o passo Multi-ponto, o método *ZH1-5* resolveu a maior parte dos problemas e perdeu no critério de eficiência apenas para o método *LMR-5*. Usando a Robustez como desempate, temos que o passo espectral Multi-ponto combinado à busca linear ZH1 apresentou um desempenho melhor.

As Figuras 5.6-5.12 apresentam a comparação entre os métodos utilizando o tempo de CPU para analisar o perfil de desempenho. As Tabelas 5.9–5.15 mostram o resumo do desempenho comparativo dos 35 métodos divididos em grupos com a mesma técnica de busca linear não monótona combinada aos diferentes passos espectrais.

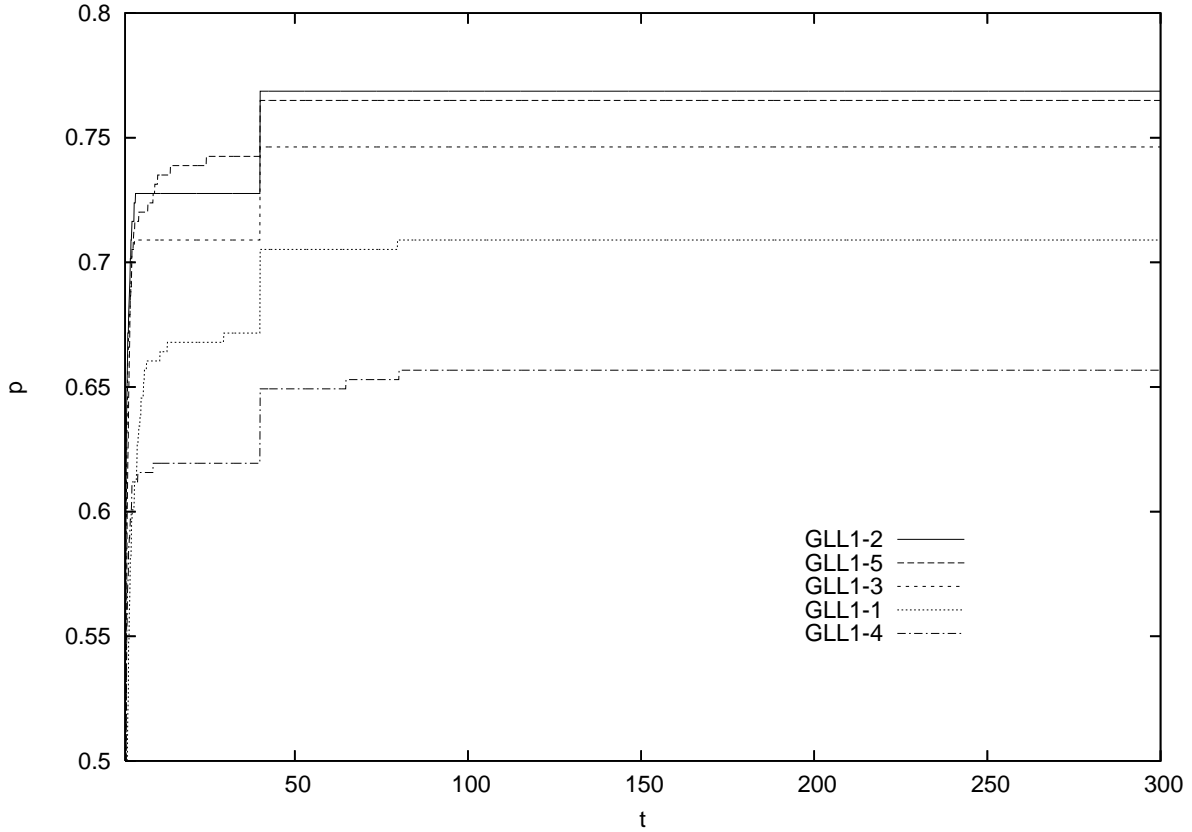


Figura 5.6: Curva do perfil de desempenho comparando os 5 métodos que utilizam a busca GLL1.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
GLL1-2	77.24	43.28	GLL1-3	75.00	45.52
GLL1-5	77.24	38.81	GLL1-2	77.24	43.28
GLL1-3	75.00	45.52	GLL1-4	66.42	39.93
GLL1-1	70.90	38.06	GLL1-5	77.24	38.81
GLL1-4	66.42	39.93	GLL1-1	70.90	38.06

Tabela 5.9: Desempenho dos 5 métodos que utilizam a busca GLL1 - Robustez x Eficiência.

Comparando-se os 5 métodos que utilizam a busca linear GLL1, os métodos *GLL1-2* e

GLL1-5 resolveram a maioria dos problemas. Enquanto que o método *GLL1-3* foi o mais eficiente, seguido pelo método *GLL1-2*. Usando a Robustez como desempate, temos que a busca linear GLL1 combinada ao passo espectral ABB-DF (método *GLL1-2*) apresentou um desempenho melhor.

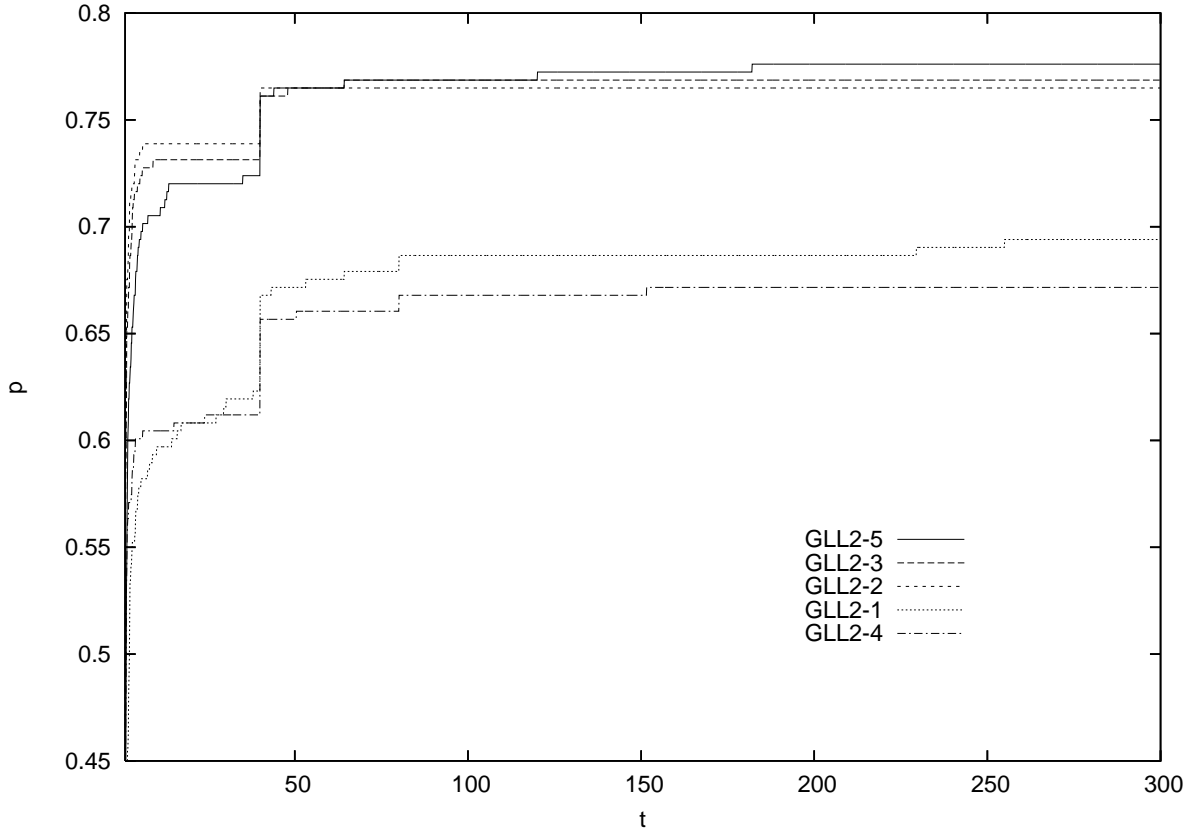


Figura 5.7: Curva do perfil de desempenho comparando os 5 métodos que utilizam a busca GLL2.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
GLL2-5	77.61	36.94	GLL2-2	76.49	46.27
GLL2-3	76.87	43.66	GLL2-3	76.87	43.66
GLL2-2	76.49	46.27	GLL2-4	67.16	39.55
GLL2-1	70.15	35.08	GLL2-5	77.61	36.94
GLL2-4	67.16	39.55	GLL2-1	70.15	35.08

Tabela 5.10: Desempenho dos 5 métodos que utilizam a busca GLL2 - Robustez x Eficiência.

Comparando-se os 5 métodos que utilizam a busca linear GLL2, o método *GLL2-5* resolveu

a maioria dos problemas. Enquanto que o método *GLL2-2* foi o mais eficiente. No entanto, se considerarmos tanto a Eficiência quanto a Robustez, então o método *GLL2-3* foi o que obteve um desempenho melhor.

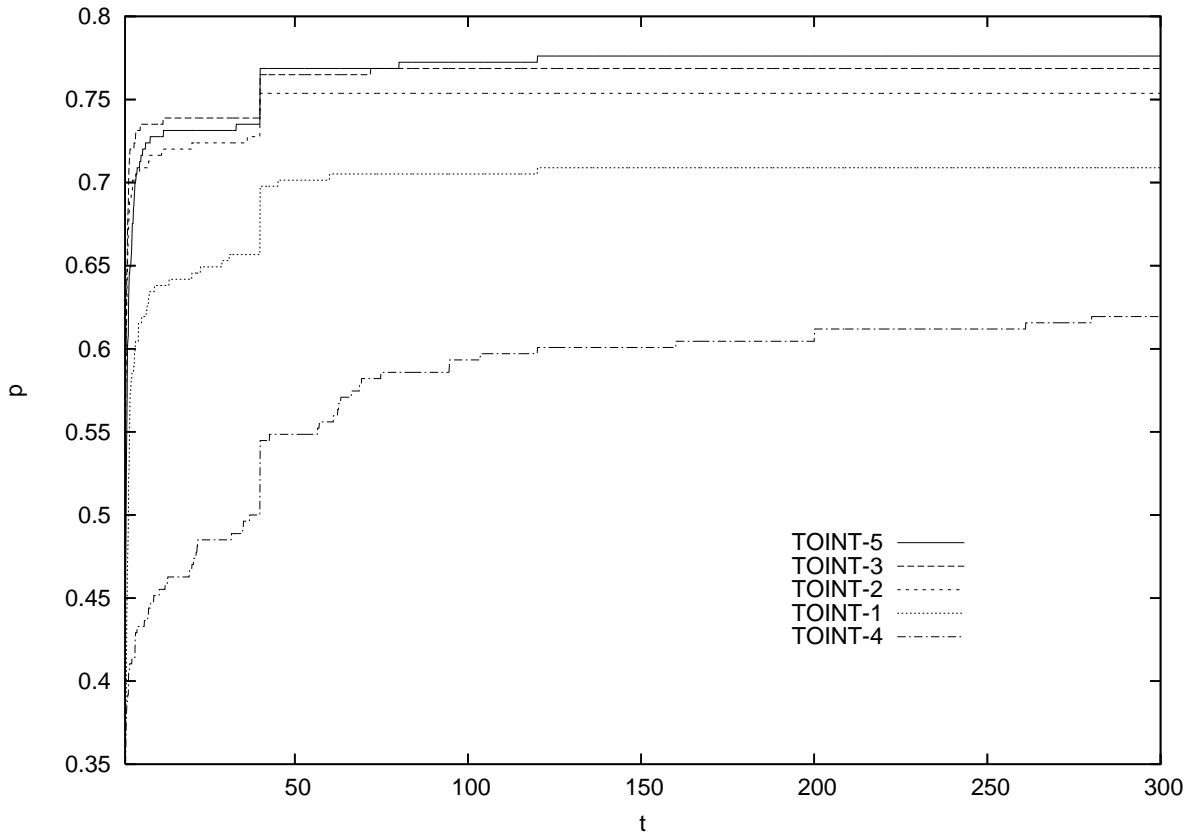


Figura 5.8: Curva do perfil de desempenho comparando os 5 métodos que utilizam a busca TOINT.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
TOINT-5	77.61	38.81	TOINT-3	76.87	44.40
TOINT-3	76.87	44.40	TOINT-2	75.37	44.40
TOINT-2	75.37	44.40	TOINT-5	77.61	38.81
TOINT-1	71.27	32.84	TOINT-1	71.27	32.84
TOINT-4	69.03	31.72	TOINT-4	69.03	31.72

Tabela 5.11: Desempenho dos 5 métodos que utilizam a busca TOINT - Robustez x Eficiência.

Comparando-se os 5 métodos que utilizam a busca linear TOINT, o método *TOINT-5* resolveu a maioria dos problemas. Enquanto que os métodos *TOINT-3* e *TOINT-2* foram os

mais eficientes. No entanto, se considerarmos tanto a Eficiência quanto a Robustez, então o método *TOINT-3* foi o que obteve um desempenho melhor.

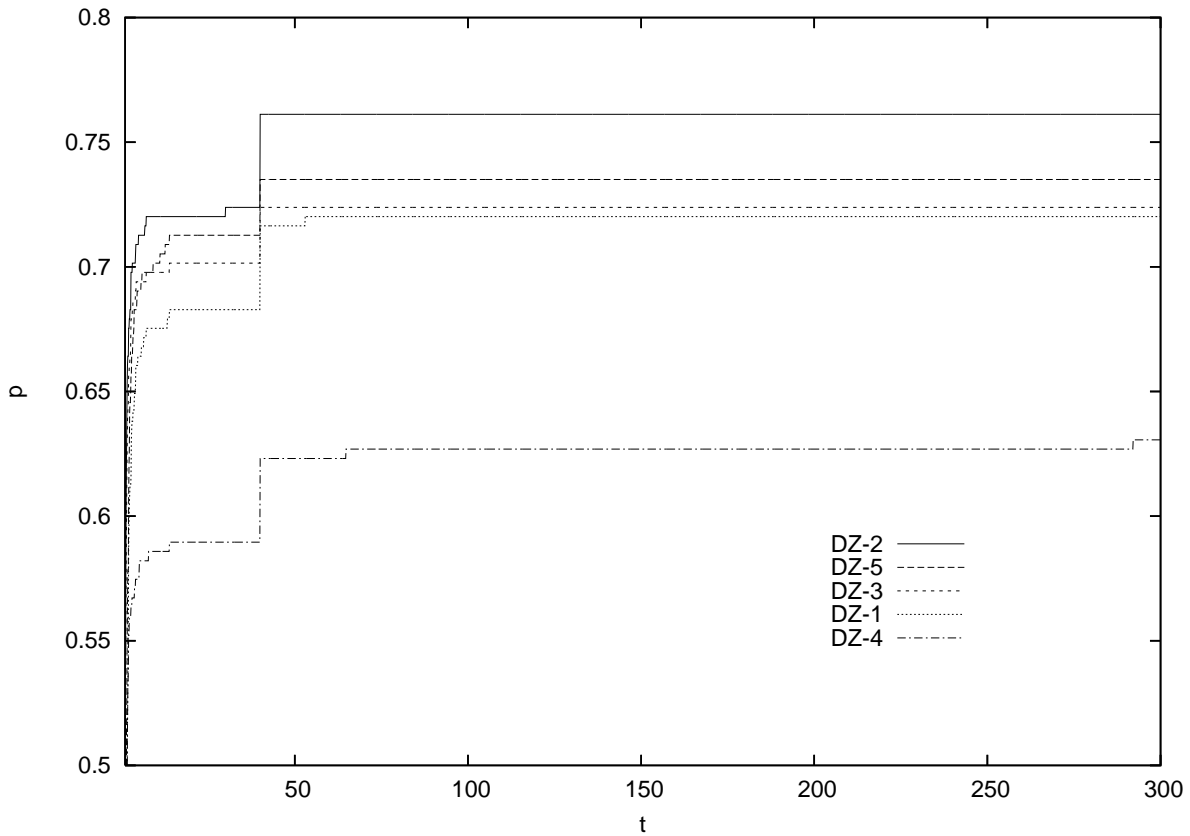


Figura 5.9: Curva do perfil de desempenho comparando os 5 métodos que utilizam a busca DZ.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
DZ-2	76.12	44.78	DZ-2	76.12	44.78
DZ-5	73.88	38.43	DZ-3	72.39	44.78
DZ-3	72.39	44.78	DZ-4	63.81	38.81
DZ-1	72.02	38.06	DZ-5	73.88	38.43
DZ-4	63.81	38.81	DZ-1	72.02	38.06

Tabela 5.12: Desempenho dos 5 métodos que utilizam a busca DZ - Robustez x Eficiência.

Comparando-se os 5 métodos que utilizam a busca linear DZ, o método *DZ-2* resolveu a maioria dos problemas e também foi um dos mais eficientes juntamente com o método *DZ-3*.

Portanto, temos que a busca linear DZ combinada ao passo espectral ABB-DF (método *DZ-2*) obteve o melhor desempenho.

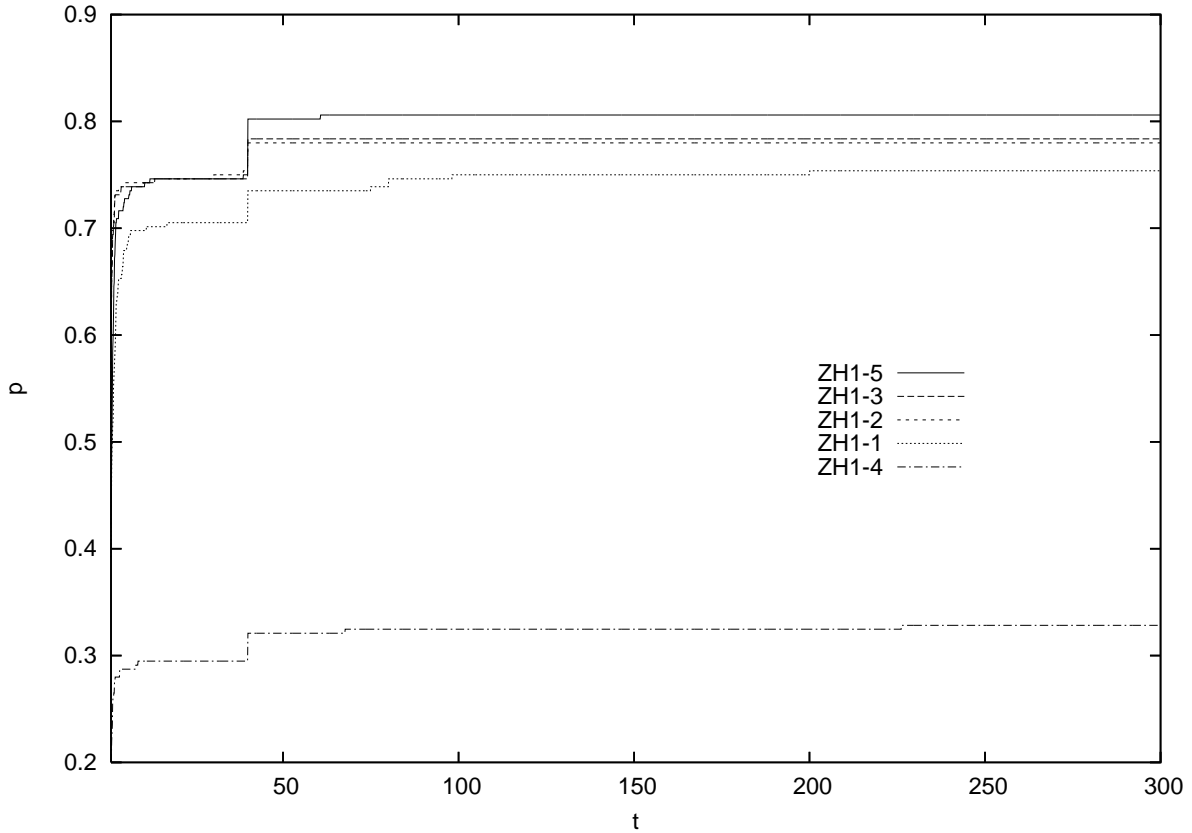


Figura 5.10: Curva do perfil de desempenho comparando os 5 métodos que utilizam a busca ZH1.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
ZH1-5	80.97	42.16	ZH1-3	78.36	52.24
ZH1-3	78.36	52.24	ZH1-2	77.99	48.51
ZH1-2	77.99	48.51	ZH1-5	80.97	42.16
ZH1-1	75.37	39.18	ZH1-1	75.37	39.18
ZH1-4	33.21	19.40	ZH1-4	33.21	19.40

Tabela 5.13: Desempenho dos 5 métodos que utilizam a busca ZH1 - Robustez x Eficiência.

Comparando-se os 5 métodos que utilizam a busca linear ZH1, o método *ZH1-5* resolveu a maioria dos problemas. Enquanto que o método *ZH1-3* foi o mais eficiente. No entanto, se

considerarmos tanto a Eficiência quanto a Robustez, então o método *ZH1-3* foi o que obteve um desempenho melhor.

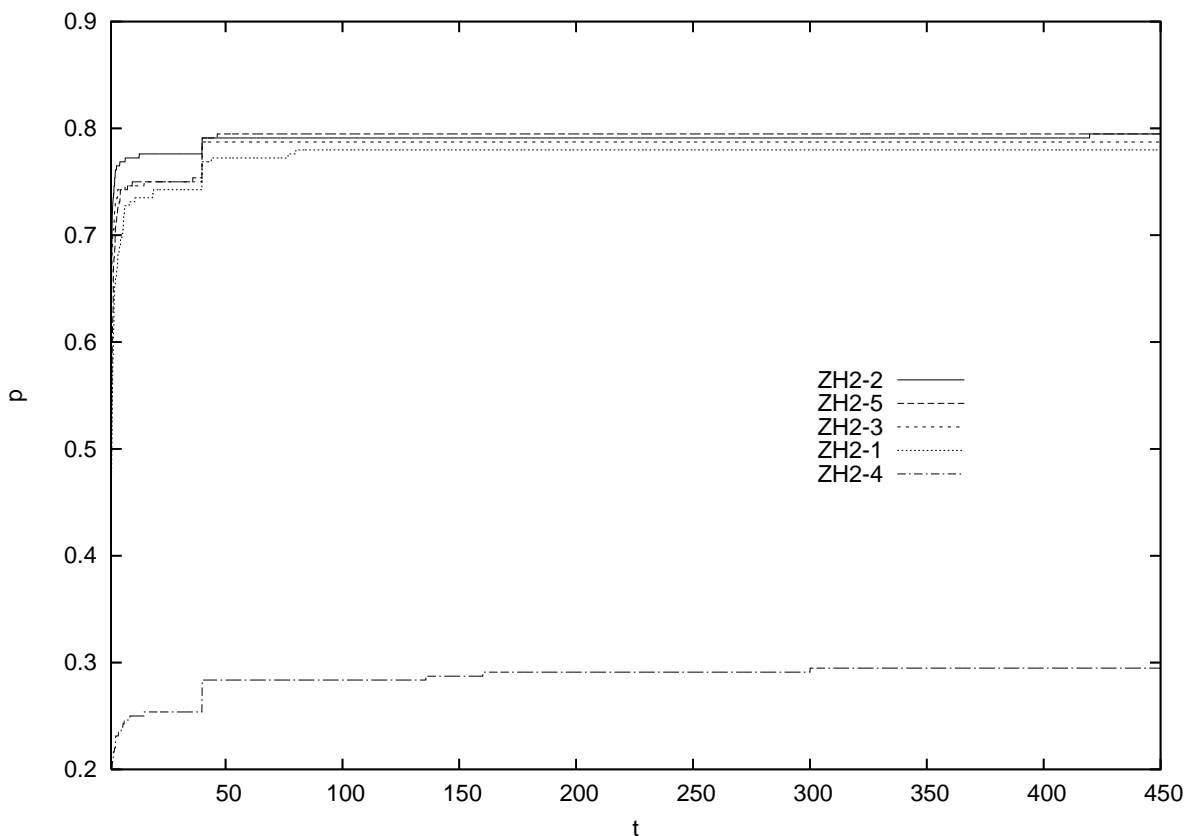


Figura 5.11: Curva do perfil de desempenho comparando os 5 métodos que utilizam a busca ZH2.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
ZH2-2	79.48	51.87	ZH2-2	79.48	51.87
ZH2-5	79.48	41.04	ZH2-3	78.73	50.00
ZH2-3	78.73	50.00	ZH2-5	79.48	41.04
ZH2-1	77.99	40.30	ZH2-1	77.99	40.30
ZH2-4	30.60	15.30	ZH2-4	30.60	15.30

Tabela 5.14: Desempenho dos 5 métodos que utilizam a busca ZH2 - Robustez x Eficiência.

Comparando-se os 5 métodos que utilizam a busca linear ZH2, os métodos *ZH2-2* e *ZH2-5* resolveram a maioria dos problemas. Enquanto que o método *ZH2-2* foi o mais eficiente.

Portanto, temos que a busca linear ZH2 combinada ao passo espectral ABB-DF (método *ZH2-2*) obteve o melhor desempenho.

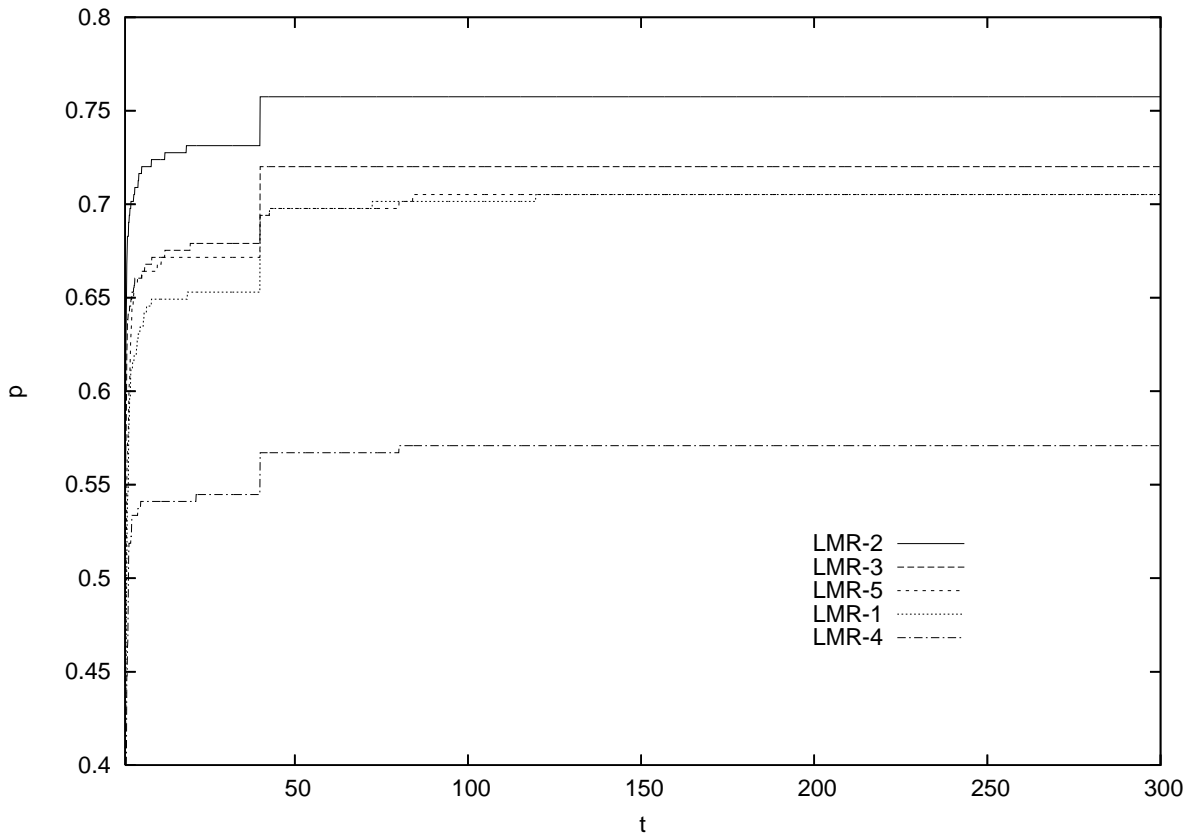


Figura 5.12: Curva do perfil de desempenho comparando os 5 métodos que utilizam a busca LMR.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
LMR-2	75.75	49.63	LMR-2	75.75	49.63
LMR-3	72.39	44.40	LMR-3	72.39	44.40
LMR-5	71.27	39.18	LMR-5	71.27	39.18
LMR-1	70.90	38.81	LMR-1	70.90	38.81
LMR-4	58.58	33.58	LMR-4	58.58	33.58

Tabela 5.15: Desempenho dos 5 métodos que utilizam a busca LMR - Robustez x Eficiência.

Comparando-se os 5 métodos que utilizam a busca linear LMR, o método *LMR-2* resolveu a maioria dos problemas e foi o mais eficiente. Portanto, temos que a busca linear LMR combinada ao passo espectral ABB-DF (método *LMR-2*) obteve o melhor desempenho.

Portanto, podemos concluir a partir da comparação entre todos os métodos com o mesmo passo espectral (primeiro grupo) que as buscas lineares não monótonas ZH1 e LMR ficaram entre as melhores neste grupo. A partir da comparação entre todos os métodos com a mesma busca linear não monótona (segundo grupo), podemos observar que para as buscas GLL2, TOINT e ZH1 os passos espectrais ABB-DF, ABB-GS e Multi-ponto ficaram entre os melhores. Para a busca GLL1, os melhores foram os passos espectrais ABB-DF e ABB-GS. Por fim, as buscas DZ, ZH2 e LMR alcançaram um melhor desempenho quando combinadas ao passo espectral ABB-DF.

A Tabela 5.16 mostra o resumo do desempenho comparativo dos 35 métodos aplicados aos 268 problemas do conjunto de testes. Assim, a coluna **Eficiência** representa o percentual de problemas em que o método foi mais rápido em encontrar uma solução e a coluna **Robustez** representa o percentual de problemas em que o método encontrou uma solução.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
ZH1-5	71.64	25.75	LMR-2	67.91	30.22
GLL2-5	70.15	23.13	ZH1-2	69.40	29.85
ZH1-2	69.40	29.85	DZ-2	65.30	28.73
ZH2-2	69.40	26.87	DZ-3	63.43	28.73
ZH2-3	69.40	23.88	ZH1-3	69.03	28.36
ZH1-3	69.03	28.36	GLL1-3	67.54	27.24
GLL1-2	69.03	26.87	ZH2-2	69.40	26.87
ZH2-5	68.66	23.88	GLL1-2	69.03	26.87
TOINT-5	68.66	23.13	LMR-3	66.42	26.87
TOINT-2	68.28	25.37	LMR-1	64.93	26.87
TOINT-3	68.28	24.25	LMR-5	65.30	26.49
LMR-2	67.91	30.22	GLL1-5	67.54	26.12
GLL2-2	67.91	24.63	ZH1-1	67.16	26.12
GLL1-3	67.54	27.24	ZH1-5	71.64	25.75
GLL1-5	67.54	26.12	TOINT-2	68.28	25.37
ZH1-1	67.16	26.12	GLL1-1	64.18	25.37
GLL2-3	67.16	22.76	DZ-5	64.55	25.00
ZH2-1	66.79	23.88	DZ-1	61.57	25.00
LMR-3	66.42	26.87	GLL1-4	59.33	25.00
DZ-2	65.30	28.73	GLL2-2	67.91	24.63
LMR-5	65.30	26.49	TOINT-3	68.28	24.25
LMR-1	64.93	26.87	DZ-4	55.22	24.25
TOINT-1	64.93	21.27	ZH2-3	69.40	23.88
DZ-5	64.55	25.00	ZH2-5	68.66	23.88
GLL1-1	64.18	25.37	ZH2-1	66.79	23.88
DZ-3	63.43	28.73	LMR-4	52.61	23.51
DZ-1	61.57	25.00	GLL2-5	70.15	23.13
GLL2-1	60.82	19.78	TOINT-5	68.66	23.13
GLL2-4	60.08	21.27	GLL2-3	67.16	22.76
TOINT-4	60.08	16.79	TOINT-1	64.93	21.27
GLL1-4	59.33	25.00	GLL2-4	60.08	21.27
DZ-4	55.22	24.25	GLL2-1	60.82	19.78
LMR-4	52.61	23.51	TOINT-4	60.08	16.79
ZH1-4	29.48	12.69	ZH1-4	29.48	12.69
ZH2-4	26.87	8.58	ZH2-4	26.87	8.58

Tabela 5.16: Desempenho dos 35 métodos - Robustez x Eficiência.

Na Tabela 5.16 pode ser visto que os métodos *ZH1-2*, *ZH2-2*, *ZH1-3*, *GLL1-2* e *LMR-2* estão entre os doze primeiros tanto em Robustez quanto em Eficiência. A Figura 5.13 apresenta a comparação entre os métodos utilizando perfil de desempenho descrito em [32]. A Tabela 5.17 mostra o desempenho destes cinco métodos no conjunto de problemas de teste.

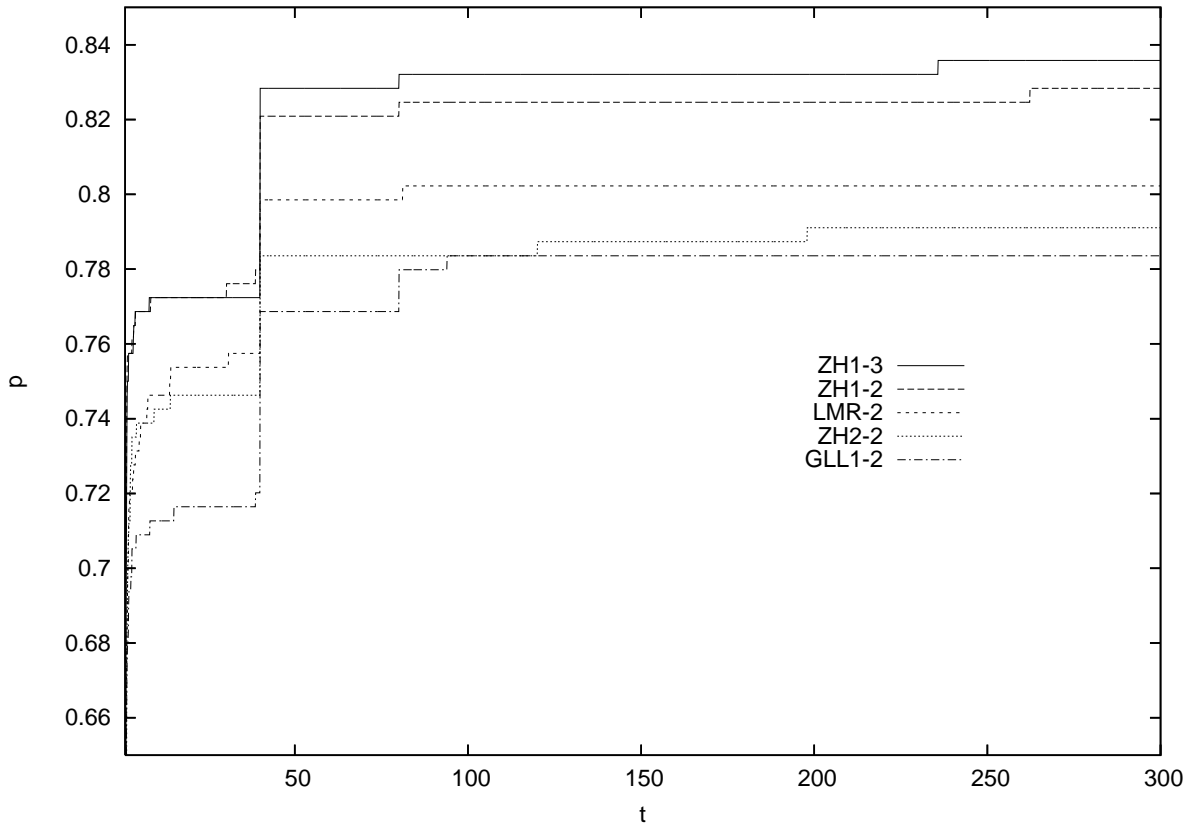


Figura 5.13: Curva do perfil de desempenho comparando os 5 métodos que estão entre os 12 primeiros.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
ZH1-3	83.58	48.88	ZH1-3	83.58	48.88
ZH1-2	82.84	44.03	LMR-2	80.22	48.13
LMR-2	80.22	48.13	ZH1-2	82.84	44.03
ZH2-2	79.85	37.31	GLL1-2	78.73	43.28
GLL1-2	78.73	43.28	ZH2-2	79.85	37.31

Tabela 5.17: Desempenho dos 5 métodos que estão entre os 12 primeiros na duas categorias: Robustez e Eficiência. Como o número de métodos comparados é menor, então os valores nas colunas “Robustez” e “Eficiência” tendem a aumentar.

A Figura 5.14 apresenta a mesma comparação feita na Figura 5.13 com a diferença que considera a Eficiência em detalhe.

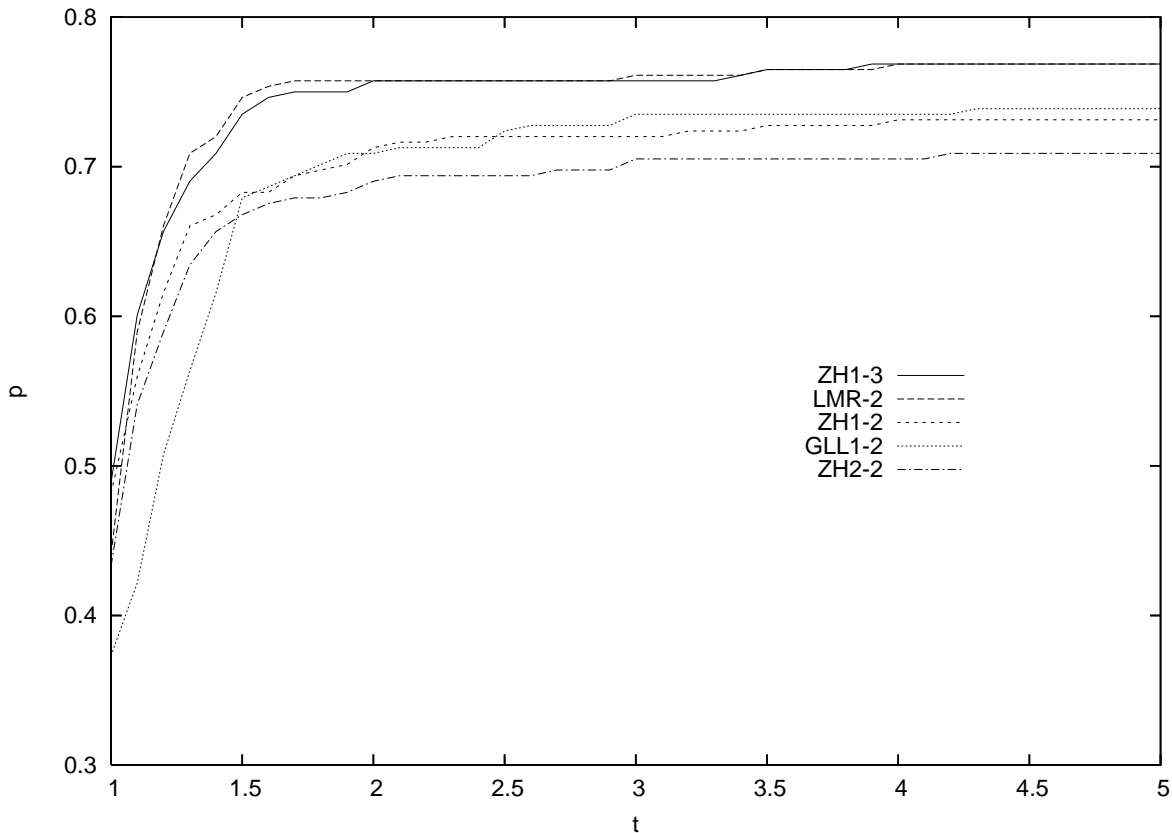


Figura 5.14: Curva do perfil de desempenho comparando os métodos (descritos na Tabela 5.17) - Eficiência em detalhe.

Dos resultados apresentados na Tabela 5.16, podemos inferir que os métodos com as versões de passos espectrais *ABB-DF*, *ABB-GS* e *Multi-ponto* combinados às buscas lineares não monótonas *ZH1*, *GLL1*, *GLL2* e *TOINT* alcançaram uma solução para a maioria dos problemas. Enquanto que todas as buscas testadas combinadas ao passo *CBB* apresentaram os piores desempenhos na categoria Robustez. Já, considerando a categoria Eficiência, os melhores resultados foram encontrados pelos passos *ABB-DF* e *ABB-GS* associados às buscas lineares *LMR*, *ZH1*, *ZH2*, *DZ* e *GLL1*. E os piores resultados nessa categoria foram obtidos pelos métodos com passos *CBB* e *BB1* combinados às buscas lineares *TOINT* e *GLL2* e pelos métodos com passo *CBB* e buscas lineares *ZH1* e *ZH2*.

Os piores desempenhos tanto em Robustez quanto em Eficiência são dados pelos métodos

que utilizaram o passo espectral *CBB* e as buscas lineares *ZH1* e *ZH2*. Assim, podemos concluir que a maioria dos métodos testados com o passo *CBB* não apresentaram desempenho razoável.

Considerando os cinco métodos (descritos na Tabela 5.17) que ficaram posicionados entre os doze melhores nas categorias Robustez e Eficiência, observamos que a busca linear não monótona *ZH1* combinada ao passo *ABB-GS* (ou seja, o método *ZH1-3*) obteve o melhor resultado e que os outros quatro métodos na sequência apresentam uma característica comum: a utilização do passo *ABB-DF*.

Finalmente, a partir da Tabela 5.16, podemos observar que o método *ZH1-2* obteve o melhor desempenho entre os 35 métodos testados considerando-se tanto “Robustez” quanto “Eficiência”. Desta forma, realizamos a comparação entre o método *ZH1-2* e o método SPG. Vale ressaltar que o método *GLL1-1* utilizado nos experimentos numéricos é exatamente o método SPG que utiliza o passo espectral BB1 e a busca linear não monótona GLL1.

A Figura 5.15 apresenta a comparação entre os métodos utilizando o tempo de CPU para analisar o perfil de desempenho. Podemos observar que o método *ZH1-2* conseguiu resolver a maior parte dos problemas e também é mais eficiente comparado ao método SPG.

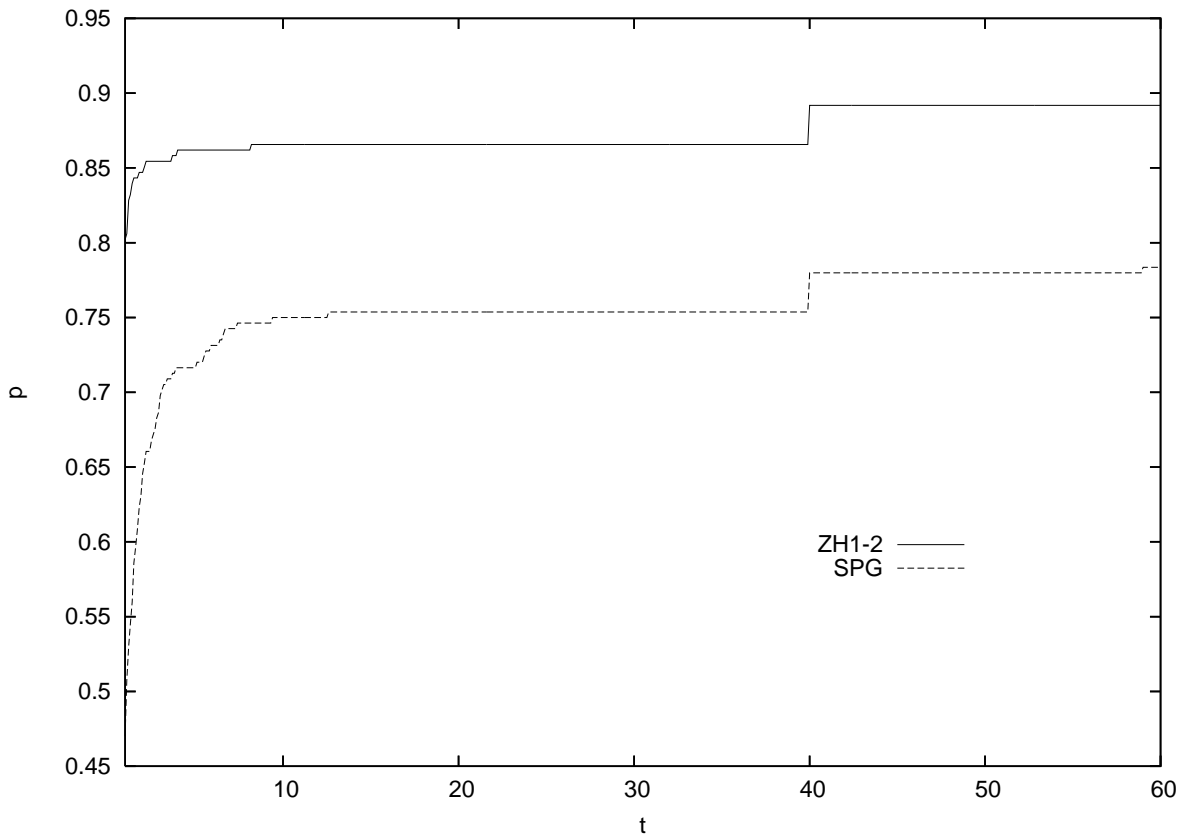


Figura 5.15: Curva do perfil de desempenho comparando os métodos métodos ZH1-2 e SPG.

A Tabela 5.18 mostra o resumo do desempenho comparativo entre os métodos *ZH1-2* e SPG aplicados aos 268 problemas do conjunto de testes.

Ordenado por Robustez			Ordenado por Eficiência		
Método	Robustez	Eficiência	Método	Robustez	Eficiência
ZH1-2	89.18	80.22	ZH1-2	89.18	80.22
SPG	78.36	47.02	SPG	78.36	47.02

Tabela 5.18: Desempenho dos métodos ZH1-2 e SPG - Robustez x Eficiência.

Capítulo 6

Conclusões e pesquisas futuras

Apresentamos algumas versões de buscas lineares não monótonas e passos espectrais mais utilizados recentemente. Combinamos as diferentes buscas lineares com as versões de passos espectrais e aplicamos ao método do Gradiente Espectral Projetado.

Realizamos experimentos considerando problemas irrestritos e com restrições de caixa. De acordo com os resultados numéricos podemos concluir que métodos com uma variação não monótona relativamente simples do critério de Armijo (neste caso, as buscas *ZH1*, *ZH2* e *LMR*) associada à utilização de passos espectrais que combinam um certo número de passos consecutivos (passo *Multi-ponto*) ou que se alternam entre as duas versões de passos espectrais de Barzilai e Borwein (passos *ABB-DF* e *ABB-GS*) podem apresentar um desempenho melhor do que métodos que utilizam buscas lineares com técnicas mais sofisticadas e passos espectrais clássicos. Também constatamos, assim como os autores de [28], que a versão cíclica do passo BB1 (CBB) com tamanho fixo de ciclo não apresenta bons resultados na maioria dos casos pois nos testes que realizaram a escolha ótima do tamanho do ciclo mudava de um caso de teste para outro (ou seja, dependia muito das características do problema testado).

As sugestões de pesquisas futuras para este trabalho seguem na direção do desenvolvimento de um método para a determinação ótima de parâmetros dos algoritmos pois, como observamos ao longo deste trabalho, as diferentes buscas lineares não monótonas utilizam um número razoavelmente grande e variado de parâmetros. Os autores de cada uma das versões dos passos espectrais e buscas lineares mencionam que o desempenho de seus métodos depende (fortemente em alguns casos) da escolha desses parâmetros. É possível formular o problema

de encontrar os parâmetros ótimos desses métodos como um problema de otimização. As variáveis seriam os parâmetros que se deseja encontrar e o valor de função poderia ser o tempo gasto pelo método, utilizando-se desses parâmetros, para resolver um determinado problema. Vários inconvenientes deveriam ser superados para resolver um problema de otimização como esse.

Referências Bibliográficas

- [1] M. ARIOLI, T. F. CHAN, I. S. DUFF, N. I. M. GOULD E J. K. REID, *Computing a search direction for large-scale linearly constrained nonlinear optimization calculations*, Technical Report TR/PA/93/34, CERFACS, Toulouse, France (1993).
- [2] M. ANDRETTA, E. G. BIRGIN E J. M. MARTÍNEZ, *Practical active-set Euclidian trust-region method with spectral projected gradients for bound-constrained minimization*, Optimization 54, pp. 305–325 (2005).
- [3] B. M. AVERICK, R. G. CARTER, J. J. MORÉ E G.L. XUE, *The MINPACK-2 test problem collection*, Argonne National Laboratory, Preprint MCS-P153-0692, Junho (1992).
- [4] D. AZOFEIFA, N. CLARK E W. VARGAS, *Optical and electrical properties of terbium films as a function of hydrogen concentration*, Physica Status Solidi B – Basic Solid State Physics 242, pp. 2005-2009 (2005).
- [5] J. BARZILAI E J. M. BORWEIN, *Two point step size gradient methods*, IMA Journal of Numerical Analysis 8, pp. 141–148 (1988).
- [6] L. BELLO E M. RAYDAN, *Convex constrained optimization for the seismic reflection tomography problem*, Journal of Applied Geophysics 62, pp. 158–166 (2007).
- [7] D. P. BERTSEKAS, *On the Goldstein-Levitin-Polyak gradient projection method*, IEEE Transactions on Automatic Control 21, pp. 174-184 (1976).
- [8] E. G. BIRGIN, I. E. CHAMBOULEYRON E J. M. MARTÍNEZ, *Optimization problems in the estimation of parameters of thin films and the elimination of the influence of the substrate*, Journal of Computational and Applied Mathematics 152, pp. 35-50 (2003).
- [9] E. G. BIRGIN E J. M. MARTÍNEZ, *A box-constrained optimization algorithm with negative curvature directions and spectral projected gradients*, Computing 15, pp. 49–60 (2001).
- [10] E. G. BIRGIN E J. M. MARTÍNEZ, *Large-scale active-set box-constrained optimization method with spectral projected gradients*, Computational Optimization and Applications 23, pp. 101–125 (2002).
- [11] E. G. BIRGIN, J. M. MARTÍNEZ, W. F. MASCARENHAS E D. P. RONCONI, *Method of sentinels for packing items within arbitrary convex regions*, Journal of the Operational Research Society 57, pp. 735-746 (2006).
- [12] E. G. BIRGIN, J. M. MARTÍNEZ, F. H. NISHIHARA, D. P. RONCONI, *Orthogonal packing of rectangular items within arbitrary convex regions by nonlinear optimization*, Computers and Operations Research 33, pp. 3535-3548 (2006).

-
- [13] E. G. BIRGIN, J. M. MARTÍNEZ E M. RAYDAN, *Nonmonotone spectral projected gradient methods on convex sets*, SIAM Journal on Optimization 10, pp. 1196–1211 (2000).
- [14] E. G. BIRGIN, J. M. MARTÍNEZ E M. RAYDAN, *Algorithm 813: SPG – software for convex-constrained optimization*, ACM Transactions on Mathematical Software 27, pp. 340–349 (2001).
- [15] E. G. BIRGIN, J. M. MARTÍNEZ E M. RAYDAN, *Inexact Spectral Projected Gradient methods on convex sets*, IMA Journal on Numerical Analysis 23, pp. 539–559 (2003).
- [16] E. G. BIRGIN, J. M. MARTÍNEZ E M. RAYDAN, *Spectral Projected Gradient Methods*, Encyclopedia of Optimization C. A. Floudas and P. M. Pardalos (Eds.), Second edition, a publicar.
- [17] E. G. BIRGIN, J. M. MARTÍNEZ E D. P. RONCONI, *Minimization subproblems and heuristics for an applied clustering problem*, European Journal of Operational Research 146, pp. 19–34 (2003).
- [18] E. G. BIRGIN, J. M. MARTÍNEZ E D. P. RONCONI, *Optimizing the packing of cylinders into a rectangular container: A nonlinear approach*, European Journal of Operational Research 160, pp. 19–33 (2005).
- [19] I. BONGARTZ, A. R. CONN, N. I. M. GOULD E P. L. TOINT, *CUTE: constrained and unconstrained testing environments*, ACM Transactions on Mathematical Software 21, pp. 123–160 (1995).
- [20] A. BUCKLEY E A. LENIR, *QN-like variable storage conjugate gradients*, Mathematical Programming 27, pp. 155–175 (1983).
- [21] A. CAUCHY, *Méthode générale pour la résolution des systèmes d'équations simultanées*, Comp. Rend. Sci. Paris, 25, pp. 536–538 (1847).
- [22] R. M. CHAMBERLAIN, M. J. D. POWELL, C. LEMARECHAL E H. C. PEDERSEN, *The watchdog technique for forcing convergence in algorithms for constrained optimization*, Math. Prog. Study 16, pp. 1–17 (1982).
- [23] D. CORES E M. LORETO, *A generalized two-point ellipsoidal anisotropic ray tracing for converted waves*, Optimization and Engineering 8, pp. 373–396 (2007).
- [24] F. CUIEL, W. E. VARGAS E R. G. BARRERA, *Visible spectral dependence of the scattering and absorption coefficients of pigmented coatings from inversion of diffuse reflectance spectra*, Applied Optics 41, pp. 5969–5978 (2002).
- [25] Y. H. DAI, *Alternate Step Gradient Method*, Optimization 52, pp. 395–415 (2003).
- [26] Y. H. DAI E R. FLETCHER, *Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming*, Numerische Mathematik 100, pp. 21–47 (2005).
- [27] Y. H. DAI E R. FLETCHER, *New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds*, Mathematical Programming 106, pp. 403–421 (2006).
- [28] Y. H. DAI, W. W. HAGER, K. SCHITTKOWSKI E H. C. ZHANG, *The cyclic Barzilai-Borwein method for unconstrained optimization*, IMA Journal of Numerical Analysis 26, pp. 604–627 (2006).
- [29] Y. H. DAI E L. Z. LIAO, *R-linear convergence of the Barzilai and Borwein gradient method*, IMA Journal on Numerical Analysis 22, pp. 1–10 (2002).
- [30] Y. H. DAI E H. C. ZHANG, *Adaptive two-point stepsize gradient algorithm*, Numerical Algorithms 27, pp. 377–385 (2001).
-

-
- [31] G.P. DEIDDA, E. BONOMI E C. MANZI, *Inversion of electrical conductivity data with Tikhonov regularization approach: some considerations*, Annals of Geophysics 46, pp. 549-558 (2003).
- [32] E. D. DOLAN E J. J. MORÉ, *Benchmarking optimization software with performance profiles*, Mathematical Programming 91, pp. 201-213 (2002).
- [33] R. FLETCHER, *On the Barzilai-Borwein method*, Department of Mathematics, University of Dundee, NA/207, Dundee, Scotland (2001).
- [34] J. B. FRANCISCO, J. M. MARTÍNEZ E L. MARTÍNEZ, *Density-based globally convergent trust-region methods for self-consistent field electronic structure calculations*, Journal of Mathematical Chemistry 40, pp. 349-377 (2006).
- [35] G. FRASSOLDATI, L. ZANNI E G. ZANGHIRATI, *New Adaptive Stepsize Selections in Gradient Methods*, Technical Report n. 77, University of Modena and Reggio Emilia, Italy (2007).
- [36] A. FRIEDLANDER, J. M. MARTÍNEZ, B. MOLINA E M. RAYDAN, *Gradient method with retards and generalizations*, SIAM Journal on Numerical Analysis 36, pp. 275-289 (1998).
- [37] A. FRIEDLANDER, J. M. MARTÍNEZ E M. RAYDAN, *A new method for large-scale box constrained convex quadratic minimization problems*, Optimization Methods and Software 5, pp. 57-74 (1995).
- [38] K. GARG E R. A. TAPIA, *A Variable Storage Algorithm for Unconstrained Optimization*, Technical report, Department of Mathematical Sciences, Rice University, Houston, TX (1977).
- [39] N. I. M. GOULD, D. ORBAN E PH. L. TOINT, General CUTEr documentation, *CERFACS Technical Report TR/PA/02/13* (2002).
- [40] A. A. GOLDSTEIN, *Convex Programming in Hilbert Space*, Bulletin of the American Mathematical Society 70, pp. 709-710 (1964).
- [41] L. GRIPPO, F. LAMPARIELLO E S. LUCIDI, *A nonmonotone line search technique for Newton's method*, SIAM Journal on Numerical Analysis 23, pp. 707-716 (1986).
- [42] L. GRIPPO, F. LAMPARIELLO E S. LUCIDI, *A class of nonmonotone stabilization methods in unconstrained optimization*, Numerische Mathematik 59, pp. 779-805 (1991).
- [43] L. GRIPPO E M. SCIANDRONE, *Nonmonotone globalization techniques for the Barzilai-Borwein gradient method*, Computational Optimization and Applications 23, pp. 143-169 (2002).
- [44] Z. JIANG, *Applications of conditional nonlinear optimal perturbation to the study of the stability and sensitivity of the Jovian atmosphere*, Advances in Atmospheric Sciences 23, pp. 775-783 (2006).
- [45] W. LA CRUZ, J. M. MARTÍNEZ E M. RAYDAN, *Spectral residual method without gradient information for solving large-scale nonlinear systems of equations*, Mathematics of Computations, pp. 1429-1448 (2006).
- [46] E. S. LEVITIN E B. T. POLYAK, *Constrained Minimization Problems*, USSR Computational Mathematics and Mathematical Physics 6, pp. 1-50 (1966).
- [47] D. H. LI E M. FUKUSHIMA, *A derivative-free line search and global convergence of Broyden-like method for nonlinear equations*, Optimization Methods and Software 13, pp. 181-201 (2000).
-

- [48] D. C. LIU E J. NOCEDAL, *On the limited memory BFGS method for large scale optimization*, Mathematical Programming 45, pp. 503–528 (1989).
- [49] J. J. MORÉ, B. S. GARBOW E K. E. HILLSTROM, *Testing unconstrained optimization software*, ACM Transactions on Mathematical Software 7, pp. 17–41 (1981).
- [50] J. NOCEDAL, *Updating quasi-Newton matrices with limited storage*, Mathematics of Computation 35, pp. 773–782 (1980).
- [51] A. RAMIREZ-PORRAS E W. E. VARGAS-CASTRO, *Transmission of visible light through oxidized copper films: feasibility of using a spectral projected gradient method*, Applied Optics 43, pp. 1508-1514 (2004).
- [52] M. RAYDAN, *On the Barzilai and Borwein choice of steplength for the gradient method*, IMA Journal of Numerical Analysis 13, pp. 321–326 (1993).
- [53] M. RAYDAN, *The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem*, SIAM Journal on Optimization 7, pp. 26–33 (1997).
- [54] M. RAYDAN E B. F. SVAITER, *Relaxed Steepest Descent and Cauchy-Barzilai-Borwein Method*, Computational Optimization and Applications 21, pp. 155–167 (2002).
- [55] T. SERAFINI, G. ZANGHIRATI E L. ZANNI, *Gradient projection methods for quadratic programs and applications in training support vector machines*, Optimization Methods and Software 20, pp. 347-372 (2005).
- [56] P. L. TOINT, *Test Problems for Partially Separable Optimization and Results for the Routine PSPMIN*, Report Nr 83/4, Department of Mathematics, Facultés Universitaires de Namur, Namur, Belgium (1983).
- [57] P. L. TOINT, *An assessment of non-monotone line search techniques for unconstrained optimization*, SIAM J. Sci. Comput. 17, pp. 725–739 (1996).
- [58] P. L. TOINT, *A nonmonotone trust region algorithm for nonlinear optimization subject to convex constraints*, Mathematical Programming 77, pp. 69–94 (1997).
- [59] W. E. VARGAS, *Inversion methods from Kibelka-Munk analysis*, Journal of Optics A - Pure and Applied Optics 4, pp. 452-456 (2002).
- [60] W. E. VARGAS, D. E. AZOFEIFA E N. CLARK, *Retrieved optical properties of thin films on absorbing substrates from transmittance measurements by application of a spectral projected gradient method*, Thin Solid Films 425, pp. 1-8 (2003).
- [61] N. ZEEV, O. SAVASTA E D. CORES, *Non-monotone spectral projected gradient method applied to full waveform inversion*, Geophysical Prospecting 54, pp. 525-534 (2006).
- [62] H. ZHANG E W. W. HAGER, *A nonmonotone line search technique and its application to unconstrained optimization*, SIAM Journal on Optimization 14, pp. 1043–1056 (2004).