Sobre Métodos de busca padrão para Minimização de funções com Restrições Lineares

Deise Gonçalves Ferreira
Orientadora: Profa. Dra. Maria Aparecida Diniz Ehrhardt

Mestrado em Matemática Aplicada
IMECC - Unicamp

1 de março de 2013

Introdução:

- É frequente nas aplicações problemas em que as derivadas não estão disponíveis;
- Há inúmeras aplicações cujas restrições são apenas lineares;
- Subproblema em métodos voltados para problemas gerais;
- Estudar e propor métodos sem derivadas para resolvê-los;
- Focamos nosso trabalho no Método de Busca Padrão aplicado ao problema com restrições lineares;
- Propomos novas estratégias de busca e um novo Padrão buscando melhorar o desempenho do método.

Definindo o problema a ser trabalhado

Trabalharemos com um Método de Busca padrão aplicado ao problema de Otimização com restrições lineares definido por:

$$(\mathrm{PRL}) \begin{cases} min & f(x) \\ s.a. & l \leq Ax \leq u \\ & bl \leq x \leq bu \end{cases}$$

$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad x \in \mathbb{R}^n, \quad A \in \mathbb{R}^{m \times n}, \quad l, u, bl, bu \in \mathbb{R}^m$

Métodos de Busca direta

Definição:

Dizemos que um método é de busca direta se, além de não computar derivadas, ele não utilizar os valores de função em nenhum cálculo.

Vantagens dessa classe de métodos

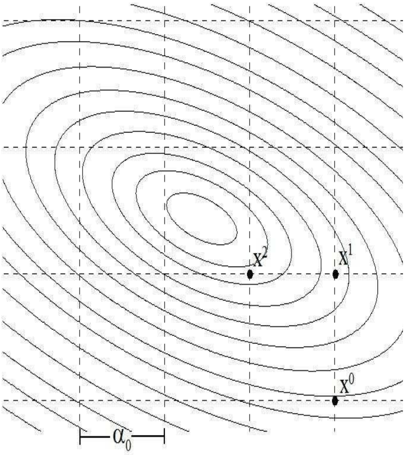
- Algoritmos simples e com forte apelo geométrico;
- Iteração barata cujo passo mais caro é avaliar o valor da função;
- Algoritmos fáceis de serem implementados;
- Aplicação quase que imediata;
- Sob certas hipóteses possui uma robusta teoria de convergência;
- Precisamos de muito pouco para poder aplicar o método;
- É uma alternativa quando métodos mais elaborados falham.

Desvantagens dessa classe de métodos

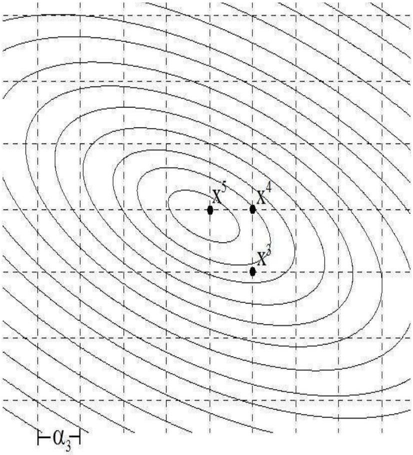
- Sem derivadas não podemos verificar se uma determinada direção é de descida;
- Quando a sequência converge não podemos testar se de fato é um ponto estacionário e muito menos se é um minimizador;
- Ao abandonar derivadas o desempenho do método é muito inferior;
- Embora a iteração seja barata em geral o número de iterações é muito alto;
- Não podemos impor condições de decréscimo suficiente;
- Decréscimo lento, ou passos muito curtos;
- Avaliar a função muitas vezes, encarecendo a iteração;

Métodos de Busca Padrão

- É considerado um método de Busca Direta;
- Utiliza busca monótona;
- A partir de um ponto x^k, o próximo iterando deve satisfazer $f(x^{k+1}) < f(x^k)$;
- A matriz P^k contém o conjunto de direções pelas quais será realizada a busca;
- A matriz P^k é o que chamamos de padrão e pode variar de uma iteração para outra.
- O novo iterando é então da forma: $x^{k+1} = x^k + \alpha_k d^k$; onde d^k é uma coluna de P^k e α_k é o tamanho do passo, que é atualizado de acordo com o status da iteração.



x^2
x^1
x^0
α_0



x^5
x^4
x^3
α_3

Generalizando o Padrão

- Em 1997, Virginia Torczon [1] generalizou o Método de Busca Padrão para minimização irrestrita;
- Torczon então apresenta uma robusta teoria de convergência.
- Em 2000, Lewis e Torczon [2] generalizam o método aplicando ao problema com restrições lineares;
- Os diferentes Métodos de Busca Padrão são definidos pelas diferentes direções de busca (Padrão) que utilizam e pelas estratégias de atualização do tamanho do passo.

---

[1] V. Torczon. On the convergence of pattern search algorithms. SIAM Journal on Optimization 7, pp. 1-25 (1997).

[2] R. M. Lewis, V. Torczon. Pattern search algorithms for linearly constrained minimization. SIAM Journal on Optimization 10, pp. 917-941 (2000).

Como é construído o Padrão?

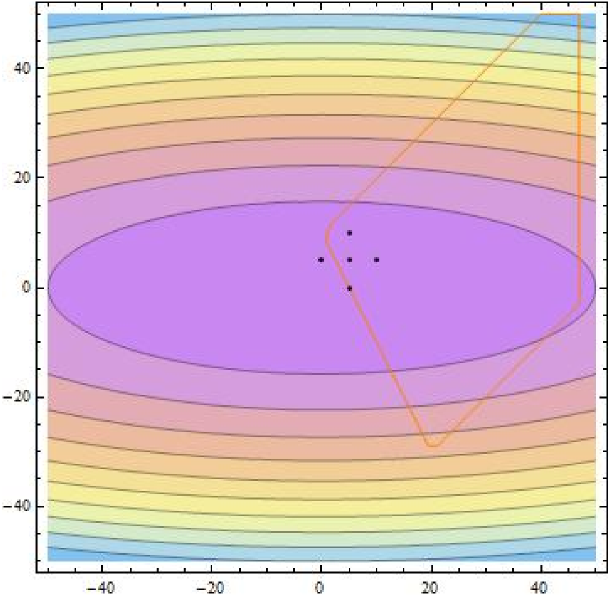
- Para o caso irrestrito as direções contidas em P^k devem gerar positivamente o $\mathbb{R}^n$;
- Com restrições o padrão deve conter um conjunto de geradores(positivos) para o cone de direções viáveis;
- É necessário preservar a viabillidade em todas as iterações;
- Devemos seguir a geometria da fronteira da região viável, quando estamos próximos desta;
- Garantir que o padrão contenham direções de descida viáveis;
- Considerar as restrições de forma a garantir passos suficientemente longos;

Problema irrestrito

- Para o caso irrestrito as direções contidas em P^k têm a característica de gerar positivamente o $\mathbb{R}^n$;
- Seja a matriz $B \in \mathbb{R}^{n \times n}$, não singular, (base para $\mathbb{R}^n$), e considere $M \subset \mathbb{Z}^{n \times n}$ um conjunto finito de matrizes não singulares e $p \in \mathbb{N}$ *tq* $p > 2n$, onde p é o número de direções em P^k.

$$P^k = \begin{bmatrix} BM^k & -BM^k & BL^k \end{bmatrix} = \begin{bmatrix} B\Gamma^k & BL^k \end{bmatrix}$$

Geometria da Fronteira

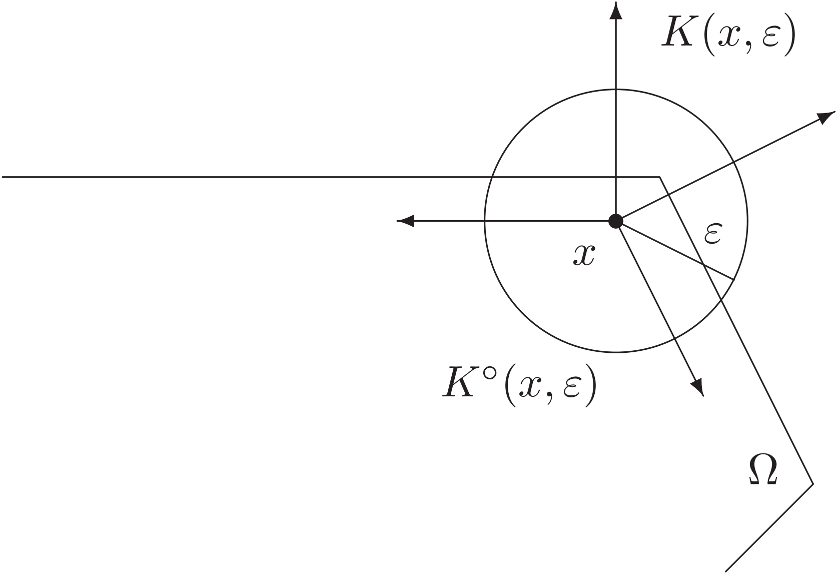


Padrão para o problema com restrições linerares

- $P^k \in \mathbb{Z}^{n \times p^k}$ (teoricamente) da forma:

$$P^k = \begin{bmatrix} \Gamma^k & L^k \end{bmatrix}$$

- $\Gamma^k \in \mathbb{Z}^{n \times r^k}$ $(r^k \geq n+1)$ - possui as restrições geométricas que deverão ser satisfeitas;
- Por simplicidade tomamos $B = I$;
- $L^k \in \mathbb{Z}^{n \times p^k - r^k}$ - deve conter pelo menos uma coluna, a coluna de zeros;
- Chamaremos de c_i^k a i-ésima coluna de P^k
- Dado o passo α_k então: $s_i^k = \alpha_k c_i^k$


$K(x,\varepsilon)$
ε
x
$K^\circ(x,\varepsilon)$
Ω

Algoritmo de busca padrão - Restrições Lineares

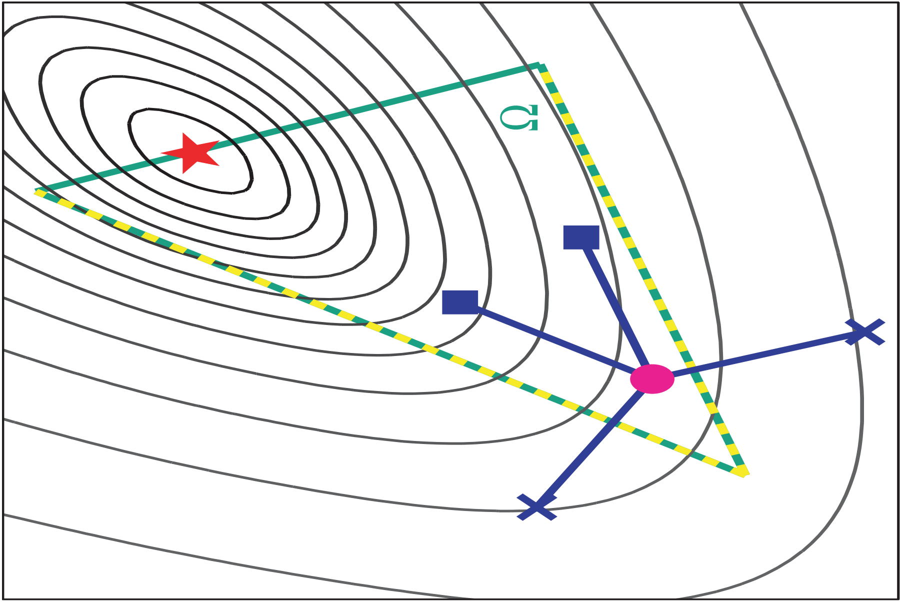
$$\Omega = \{x \in \mathbb{R}^n \mid l \leq Ax \leq u\}$$

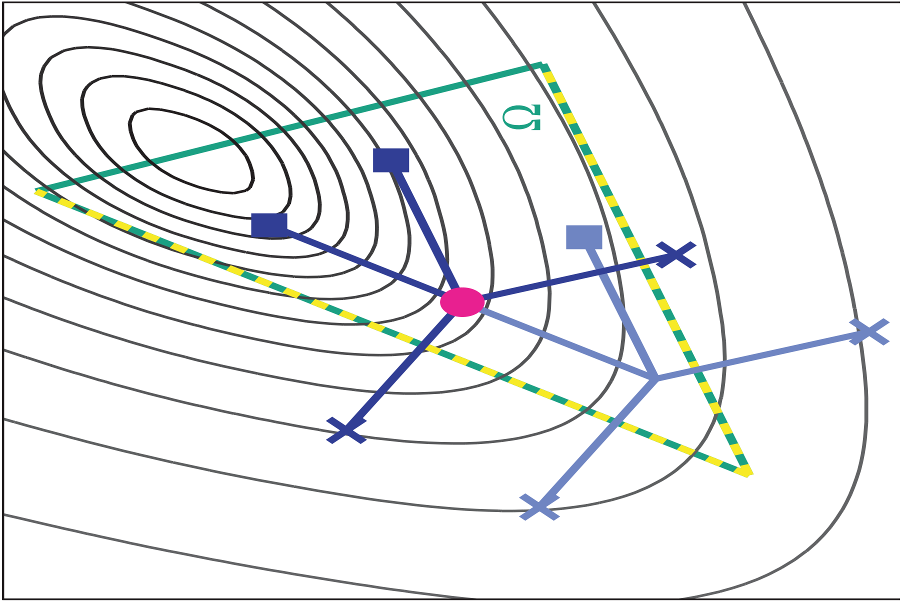
Algoritmo

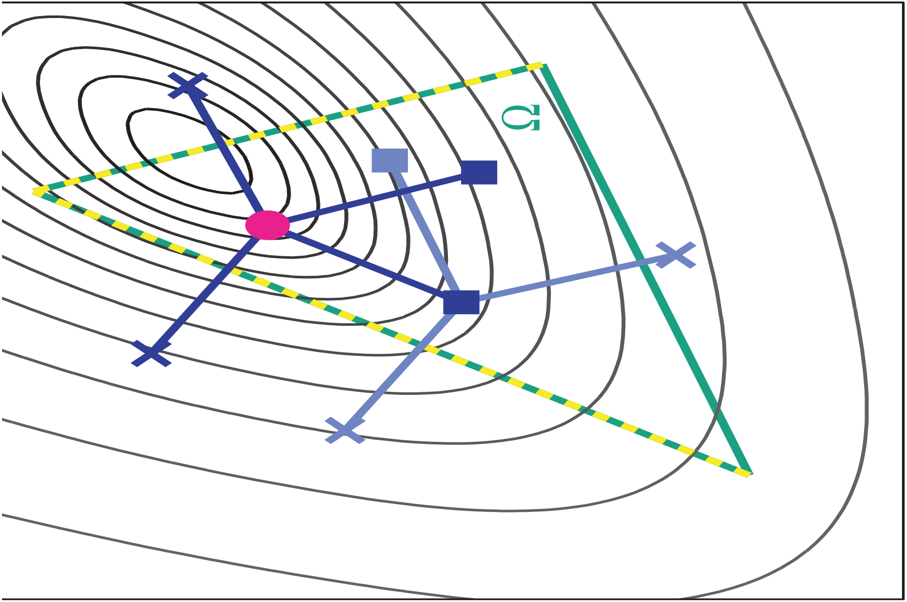
Dado $x^0 \in \Omega$ e $\alpha_0 > 0$

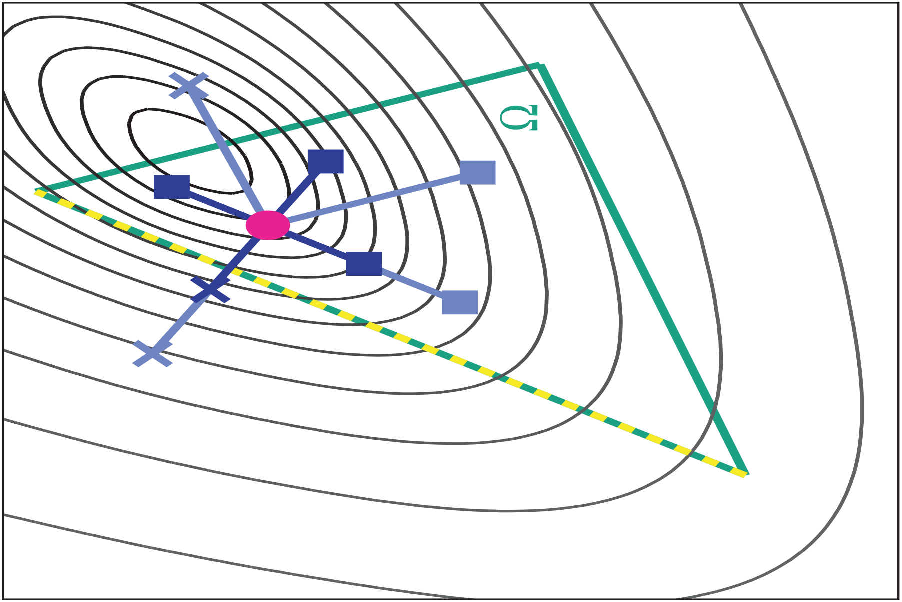
Para $k = 0, 1, \ldots,$

1. Calcule $f(x^k)$
2. Determine o passo $s_i^k = \alpha_k c_i^k$
3. Verifique a viabilidade: $(x^k + s_i^k) \in \Omega$
4. Se $f(x^k + s_i^k) < f(x^k)$, então $x^{k+1} = (x^k + s_i^k)$ e declare sucesso, caso contrário faça $x^{k+1} = x^k$;
5. Atualize P^k e α_k


Ω


Ω


Ω


Ω

Atualização do tamanho do Passo

Atualização do tamanho do Passo

- Em caso de sucesso fazemos $\alpha_{k+1} = \lambda_k \alpha_k$, onde $\lambda_k \in [1, +\infty)$
- Em caso de fracasso fazemos $\alpha_{k+1} = \theta_k \alpha_k$, onde $\theta_k \in (0, 1)$

Os parâmetros λ_k e θ_k devem ser da forma:

- Seja $\tau \in \mathbb{Q}$, $\tau > 1$, e $\{w_0, ..., w_L\} \subset \mathbb{Z}$, $w_0 < 0$, $w_L \geq 0$, ordenados, então:
- $\theta_k = \tau^{w_i}$ para algum $w_i \in \{w_0, ..., w_L\}$ tal que $w_i < 0$;
- $\lambda_k = \tau^{w_j}$ para algum $w_j \in \{w_0, ..., w_L\}$ tal que $w_j \geq 0$;

O que propomos...

- Propomos uma estratégia para atualizar o tamanho do passo que leva em consideração as iterações bem-sucedidas consecutivas.
- Criamos a variável *suc* que é incrementada sempre que a iteração é bem-sucedida;
- Quando a iteração fracassa atribuímos zero à esta variável que volta a ser incrementada somente quando o método obter sucesso novamente.
- Então a atualização do tamanho do passo é feita da seguinte forma:

Atualização do tamanho do passo

- Em caso de sucesso fazemos $\alpha_{k+1} = \tau^{suc}\alpha_k$, onde $\tau \geq 1$ e $\|\tau^{suc}\| < +\infty$.
- Em caso de fracasso fazemos $\alpha_{k+1} = 0.5\alpha_k$.

Condições sobre os movimentos exploratórios

Hipóteses sobre os movimentos exploratórios

1. $s^k \in \alpha_k P^k$
2. $(x^k + s^k) \in \Omega$.
3. Se $min\left\{f(x^k + y)\,\middle|\,y \in \alpha_k P^k \text{ } e \text{ } (x^k + y) \in \Omega\right\} < f(x^k)$, então:
$$f(x^k + s^k) < f(x^k)$$

Hipóteses fortes sobre os movimentos exploratórios

1. $s^k \in \alpha_k P^k$
2. $(x^k + s^k) \in \Omega$.
3. Se $min\left\{f(x^k + y)\,\middle|\,y \in \alpha_k P^k \text{ } e \text{ } (x^k + y) \in \Omega\right\} < f(x^k)$, então:
$$f(x^k + s^k) \leq min\left\{f(x^k + y)\,\middle|\,y \in \alpha_k P^k \text{ } e \text{ } (x^k + y) \in \Omega\right\}$$

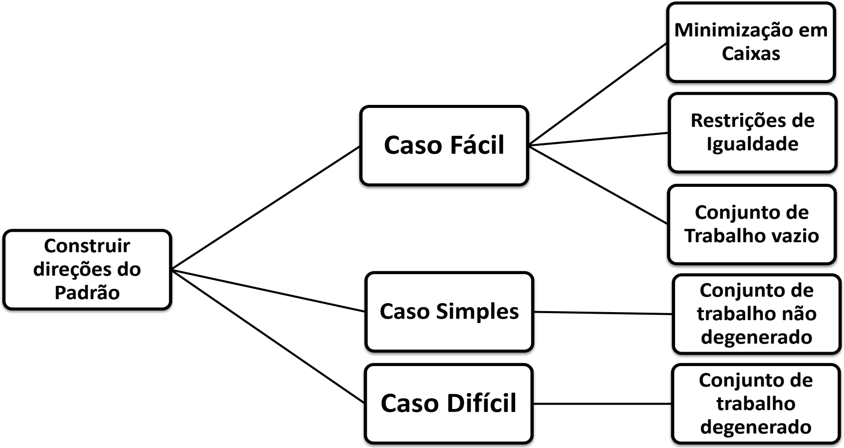
Resultados de convergência

- Assumindo que a função objetivo $f(x)$ é continuamente diferenciável;
- O conjunto de nível de $f(x)$ é compacto;
- Adicionando algumas hipóteses sobre as restrições, Padrão de busca e atualizações;
- Obtemos resultados de convergência global para o método;
- Com busca simples garantimos que existe uma subseqência gerada pelo Método que converge à um ponto KKT do problema;
- Com busca completa garantimos que toda subsequência convergente convege para a solução.
- Testes computacionais mostraram um bom desempenho do método mesmo sem satisfazer todas as condições que garantem a convergência.

Gerando o conjunto de Direções de busca

- As direções de busca são determinantes para o sucesso do método;
- Construir um padrão adequado para o tipo de problema que estamos trabalhando é essencial;
- Escolhas erradas podem fazer o método convergir para pontos que estão longe da solução;
- Dependendo da escolha o método pode não sair do lugar;
- Podemos dividir os problemas em alguns casos o que nos ajuda a determinar o padrão adequado;

Considerações ao gerar o Padrão



Minimização em caixas

O Problema de Minimização em caixas é dado por:

$$\begin{cases} min & f(x) \\ s.a. & l \leq x \leq u \end{cases}$$

- Minimizar em caixas seria o caso mais fácil de restrições lineares onde A=I;
- Neste caso sabemos de antemão os possíveis geradores para os cones $K(x,\epsilon)$ e $K^o(x,\epsilon)$ que serão sempre subconjuntos dos vetores coordenados $\pm e_i$, e este conjunto será sempre LI;
- Neste caso podemos tomar $\Gamma = [I \; -I]$ e usá-lo como padrão em todas as iterações;
- No pior caso teremos $2n$ avaliações de função por iteração;

Restrições de Igualdade

- Temos que manter a viabilidade a cada iteração;
- Para restrições de Igualdade pequenos passos podem fazer com que a viabilidade seja perdida;
- A cada iteração devemos ter:

$$A(x^k + s^k) = u$$
$$Ax^k + As^k = u, \text{ mas } Ax^k = u$$
$$As^k = 0$$

- Só conseguimos sair do ponto por direções no $\mathcal{N}(\mathcal{A})$;
- Logo meu padrão deve conter uma base positiva para $\mathcal{N}(\mathcal{A})$;
- Podemos e devemos utilizar o mesmo padrão em todas as iterações;

Restrições lineares gerais

- No caso geral o problema pode ter restrições de igualdade, desigualdade e de caixas;
- Sendo assim é importante saber as características das restrições que são ϵ-ativas no ponto corrente;
- Chamaremos de conjunto de trabalho as restrições que são ϵ- ativas na iteração corrente;

Conjunto de trabalho degenerado

Seja V a matriz cujas colunas geram o cone $K(x^k, \epsilon)$. Se as colunas de V formam um conjunto linearmente independente, diremos que o conjunto de trabalho em x^k, ou seja, na iteração corrente, é não-degenerado, caso contrário, diremos que o conjunto de trabalho na iteração corrente é degenerado.

Construindo o conjunto de trabalho

- Teremos uma restrição de igualdade quando $l_i = u_i$;

$$\mathbb{I} = \{i : \quad l_i = u_i\} \ \ e \ \ \mathbb{D} = \{i : \quad l_i < u_i\}.$$

- As restrições de desigualdade ativas serão identificadas pelos seguintes conjuntos de índices:

$$\mathbb{D}_l(x,\epsilon) = \{i : x \in \partial\Omega_{l_i}(\epsilon)\}, \ \ e \ \ \mathbb{D}_u(x,\epsilon) = \{i : x \in \partial\Omega_{u_i}(\epsilon)\},$$

- Podemos ter restrições momentaneamente de igualdade identificadas pelo conjunto:

$$\mathbb{D}_e(x,\epsilon) = \{i : i \in \mathbb{D}_l(x,\epsilon) \cap \mathbb{D}_u(x,\epsilon)\}.$$

- O conjunto de trabalho é dado por:

$$V_e = \{a_i : \ i \in \mathbb{I}\} \bigcup \{a_i : \ i \in \mathbb{D}_e\},$$

$$V_d = \left\{a_i^T : \ i \in \mathbb{D}_l \backslash \mathbb{D}_e\right\} \bigcup \left\{-a_i^T : \ i \in \mathbb{D}_u \backslash \mathbb{D}_e\right\}.$$

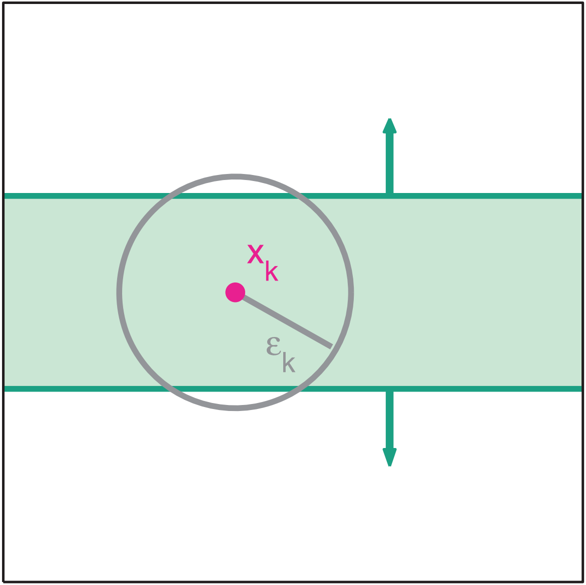


Figura: Restrições momentaneamente de igualdade

Padrão 1: Proposto por Lewis, Torczon e Shepherd

- Encontre uma base ortonormal Z para o espaço nulo de V_e;
- Calcule a matriz $B = Z^T V_d$;
- Verifique se as colunas de B formam um conjunto linearmente independe de vetores;
- Caso B não possua posto completo, o conjunto de trabalho é degenerado e outra estratégia deve ser utilizada;
- Se B tem posto completo então calcule a pseudoinversa R da matriz B^T ;
- Calcule uma base ortonormal positiva N para o núcleo de B^T;
- Então o Padrão P^k é dado por:

$$P^k = [-ZR \;\; ZN].$$

---

[1] R. M. Lewis, A. Shepherd, V. Torczon. Implementing generating set search methods for linearly constrained minimization. SIAM Journal on Scientific Computing 29, 6, pp. 2507-2530 (2007).

Proposição:

Suponha que para algum δ, $K(x,\delta)$ tem um conjunto linearmente independente de geradores racionais V. Seja N uma base racional positiva para o espaço nulo de V^T. Então para qualquer ϵ, $0 \leq \epsilon \leq \delta$, um conjunto racional de geradores para $K^{\circ}(x,\epsilon)$ pode ser obtido entre as colunas de $N, V(V^TV)^{-1}$, e $-V(V^TV)^{-1}$.

Considere também o seguinte conjunto de índices que indicam se o ponto atingiu alguma das faces:

$$\mathbb{I}_2(x^k,\epsilon_2) = \left\{ i : \ |a_i^T x^k - l_i| \leq \epsilon_2 \ ou \ |a_i^T x^k - u_i| \leq \epsilon_2 \right\},$$

$$V_{e2} = \left\{ a_i : \ i \in \mathbb{I}_2(x^k,\epsilon_2) \right\}.$$

Padrão 2: Proposto neste trabalho

- Determine uma base racional T para o núcleo da matriz V_e;
- Determine uma base racional N para o núcleo da matriz V_d;
- Determine uma base racional T_2 para o núcleo da matriz V_{e2};
- Determine a matriz $F = V_d(V_d^T V_d)^{-1}$, a qual é obtida pela resolução de vários sistemas lineares utilizando a mesma matriz de coeficientes, sendo assim, utilizamos a fatoração de Cholesky da matriz $(V_d^T V_d)$;
- O Padrão P^k fica da seguinte forma:

$$[P^k] = [-T \quad T \quad -N \quad N \quad F \quad -F \quad T_2 \quad -T_2]$$

Padrão 1: Problemas degenerados

- Os autores não utilizam a Pseudo-inversa neste caso;
- Citam que trata-se se de um cone com vértice degenerado na origem;
- Encontrar o Padrão é um caso especial de encontrar os vértices e raios extremos de um poliedro;
- Esse é um problema que já foi bastante estudado;
- Vários algoritmos vem sendo desenvolvidos para resolver esse problema;
- Os autores utilizam o pacote cddlib desenvolvido por K. Fukuda;
- Segundo os autores mostrou-se eficiente na prática em boa parte dos problemas;

Padrão 2: Problemas degenerados

- Quando o conjunto de trabalho é degenerado não é possível obter a fatoração de Cholesky da matriz $(V_d^T V_d)$;
- Propomos substituir a solução dos sistemas lineares pela solução de Quadrados Mínimos;
- Utilizamos a fatoração QR da matriz $(V_d^T V_d)$;
- É mais cara que a fatoração de Cholesky mas é mais barata que calcular a SVD;
- Se as direções funcionarem não precisamos acionar uma rotina externa para calcular o Padrão.

Viabilidade dos pontos visitados pelo algoritmo

- Fornecemos um ponto inicial viável;
- A cada iteração devemos manter a viabilidade;
- Lewis, Shepherd e Torczon, utilizam direções de busca normalizadas e constroem o conjunto de restrições ϵ- ativas levando em consideração o tamanho do passo na iteração corrente;
- Usando esta estratégia **teoricamente** a viabilidade seria preservada;
- Observamos convergência para pontos não viáveis;
- Devemos tomar cuidado ao adicionar novas direções de busca que possam retornar pontos não viáveis.

O que propomos...

- Testar a viabilidade de todos os passos antes de aceitar de calcular o valor da função objetivo;
- Garantimos que todos os passos estarão no conjunto viável com a precisão que estipularmos;
- Só calculamos o valor da função objetivo se o passo puder ser aproveitado;
- Podemos adicionar direções de busca que não temos garantias de pertencer ao cone K°;
- Consideramos as restrições de caixa neste passo, não adicionando às restrições de desigualdade do problema;
- Diminuímos as chances de trabalhar com conjuntos de trabalho degenerados;
- Aumentamos o custo da iteração, um produto matriz vetor por ponto testado;
- Não precisamos normalizar as direções de busca;
- Podemos tentar passos mais ambiciosos.

Estratégias de busca

1. Busca simples
   - Finalizamos a busca ao encontrar um ponto com valor da função objetivo melhor;
2. Busca Completa
   - Testamos todas as direções do Padrão antes de finalizar a busca;
3. Busca Simples Ordenando o Padrão
   - Com busca simples a ordem com que testamos as direções pode interferir no número de avaliações da função objetivo;
   - ordenamos as direções considerando resultado de iterações anteriores na qual obtemos sucesso;
   - O custo de ordenar o Padrão deve ser baixo;
   - Testamos primeiramente a direção que ofereceu decréscimo na iteração anterior.

Testes Computacionais

- Todas as implementações e os testes foram realizados em Matlab versão 7.10;
- Realizamos testes com 62 problemas;
- Os problemas foram divididos em 2 conjuntos:
  - Problemas suaves: 42 problemas extraídos de [1, 3]
  - Problemas minimax (não suaves) : 15 problemas extraídos de [2]

[1] D. Orban, N.I.M. Gould, P.L. Toint. CUTEr: Constrained and Unconstrained Testing Environment, revisited,http://www.cuter.rl.ac.uk (2011).

[2] L. Luksan, J. Vlcek. Test Problems for Nonsmooth Unconstrained and Linearly Constrained Optimization. Technical report No. 798, (2000).

[3] W. Hock, K. Schittkowski. "Test examples for nonlinear programming code". Springer, Berlin, (1981).

- Testamos todas as estratégias propostas procurando comparar o desempenho de cada Padrão variando a estratégia de busca e atualização do tamanho do passo;
- Testamos a viabilidade para ambos os Padrões;
- Em todos os testes o critério de parada foi $\alpha_k < 10^{-8}$ ou ao atingir o número máximo de avaliações da função;
- Variamos de 10 à 10^5 o número máximo de avaliações da função objetivo permitidos;
- Consideramos que o problema foi resolvido se:

$$f(\bar{x}) - f(x^\star) \leq 10^{-7} \quad ou \quad \|\bar{x} - x^\star\| \leq 10^{-7}$$

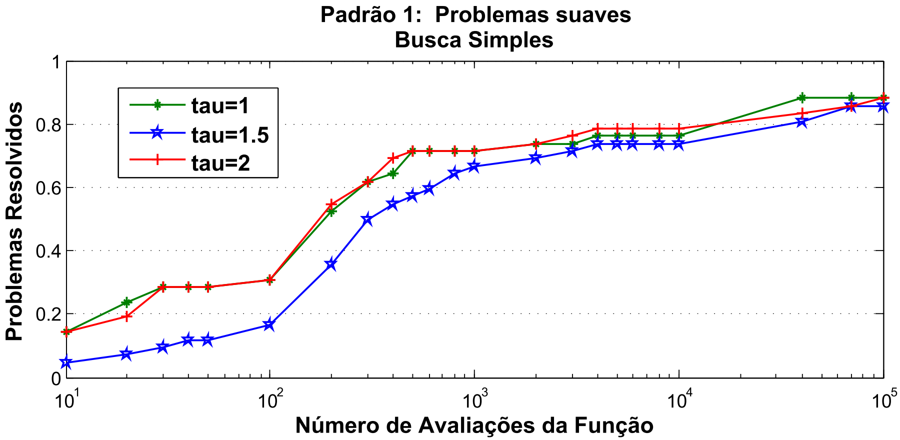


Figura: Problemas suaves: Padrão 1 - Busca Simples

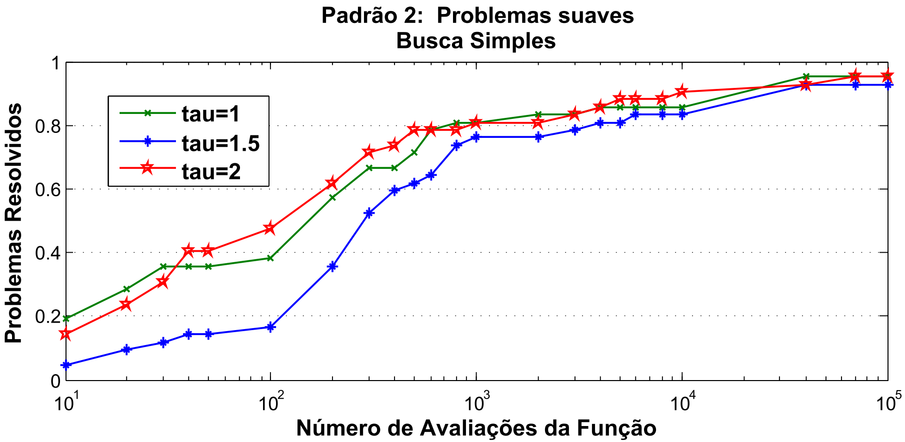


Figura: Atualização do tamanho do passo: Padrão 2 - Busca Simples

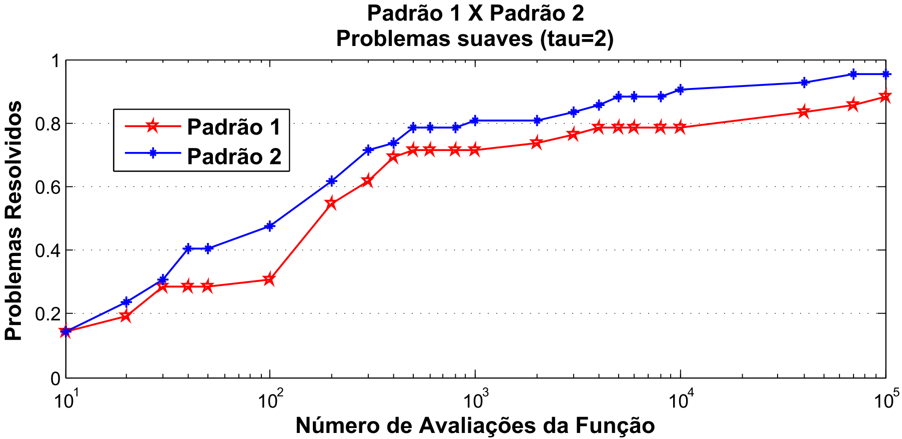


Figura: Padrão 1 $\times$ Padrão 2 - Busca Simples ($\tau = 2$)

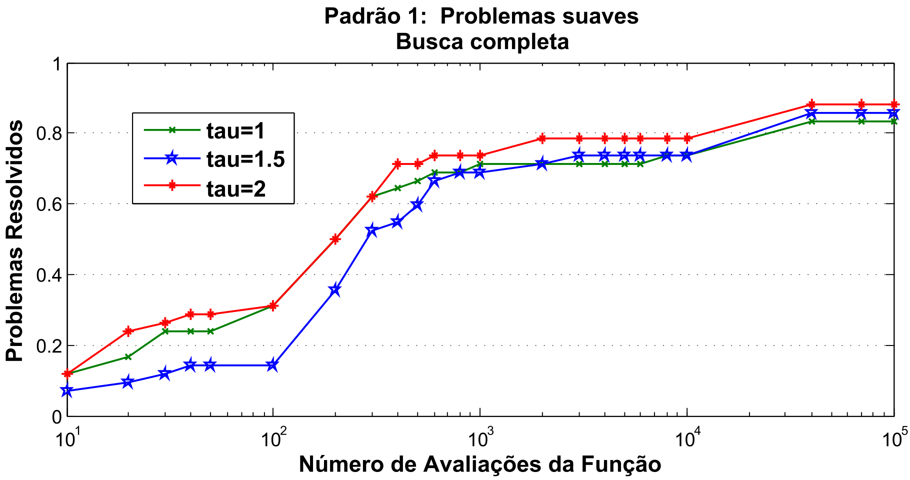


Figura: Atualização do tamanho do passo: Padrão 1 - Busca Completa

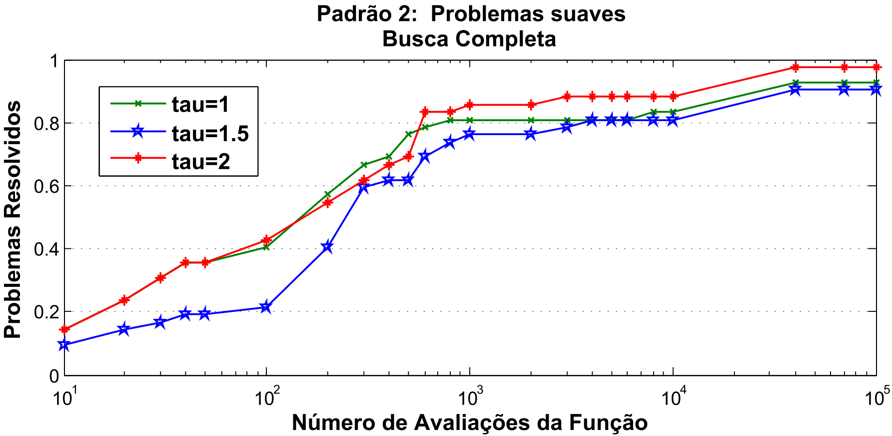


Figura: Atualização do tamanho do passo: Padrão 2 - Busca Completa

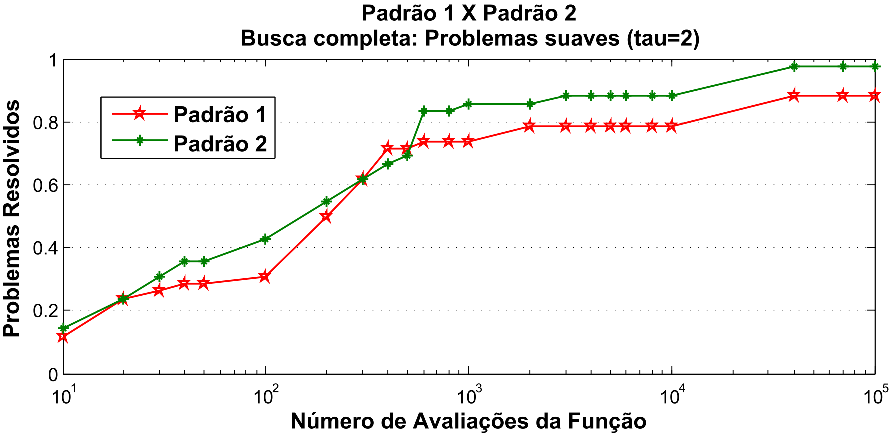


Figura: Padrão 1 $\times$ Padrão 2 - Busca Completa ($\tau = 2$)

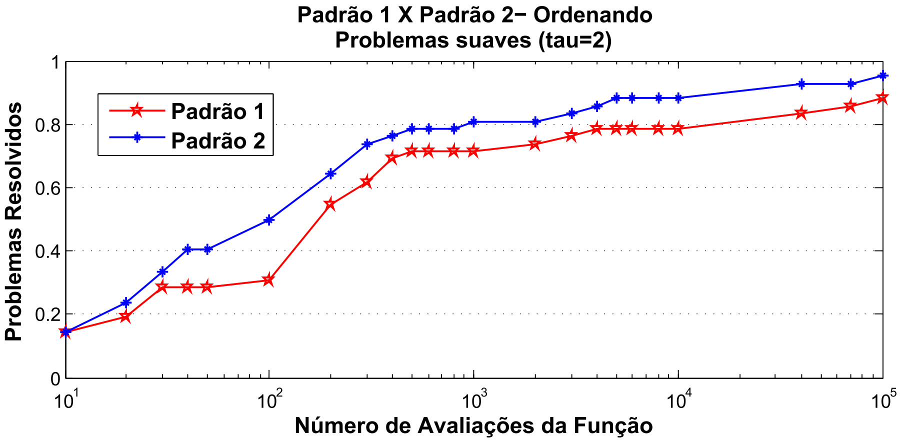


Figura: Padrão 1 $\times$ Padrão 2 - Busca Simples Ordenando o Padrão ($\tau = 2$)

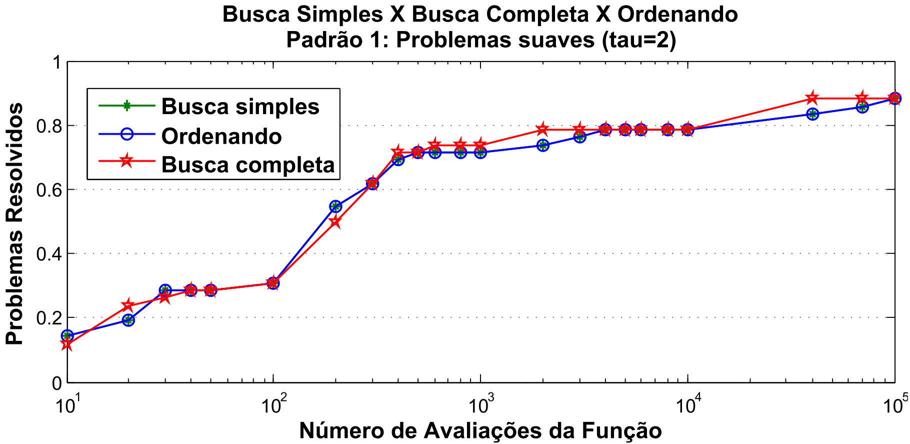


Figura: Padrão 1 - Estratégias de Busca ($\tau = 2$)

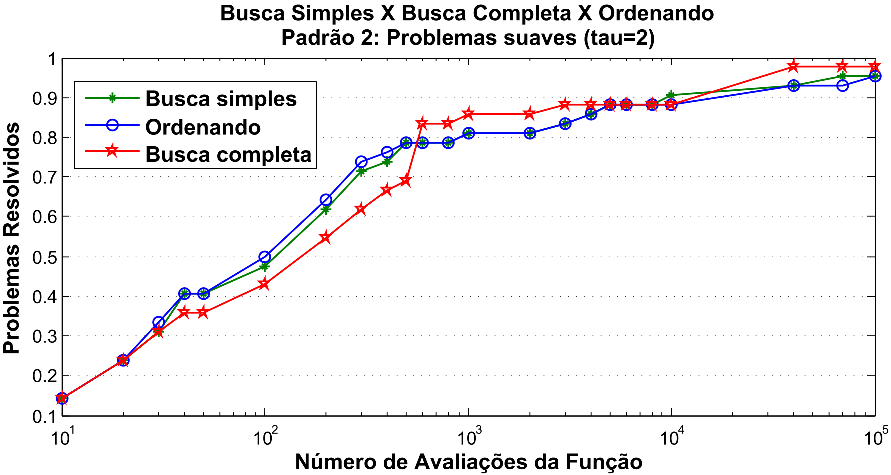


Figura: Padrão 2 - Estratégias de Busca ($\tau = 2$)

Problemas não suaves

- Trabalhamos com problemas minimax definidos por:

$$(P)\begin{cases} min & f(x) \\ s.a. & l \leq Ax \leq u \\ & bl \leq x \leq bu \end{cases}$$

onde $f(x) = max\{f_1(x), f_2(x), \ldots, f_p(x)\}$ tal que $f_i : \mathbb{R}^n \to \mathbb{R}, \quad x \in \mathbb{R}^n$, e $A \in \mathbb{R}^{m \times n}, \quad l, u \in \mathbb{R}^m$.

- O custo de avaliar a função objetivo equivale à avaliar p funções;
- Podemos abordar o problema diretamente ou usando um modelo suave para a função objetivo;
- Não temos teoria de convergência para esse tipo de problema;
- Esperamos um desempenho inferior ao observado para os problemas suaves.

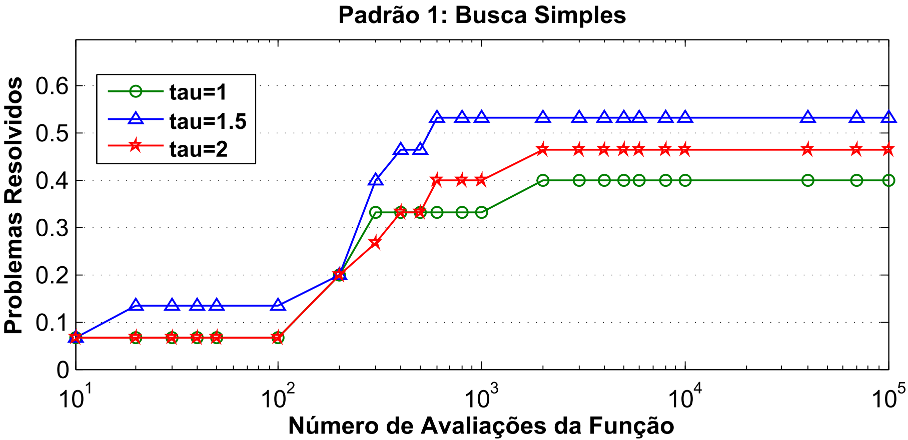


Figura: Problemas não suaves: Padrão 1 - Busca Simples

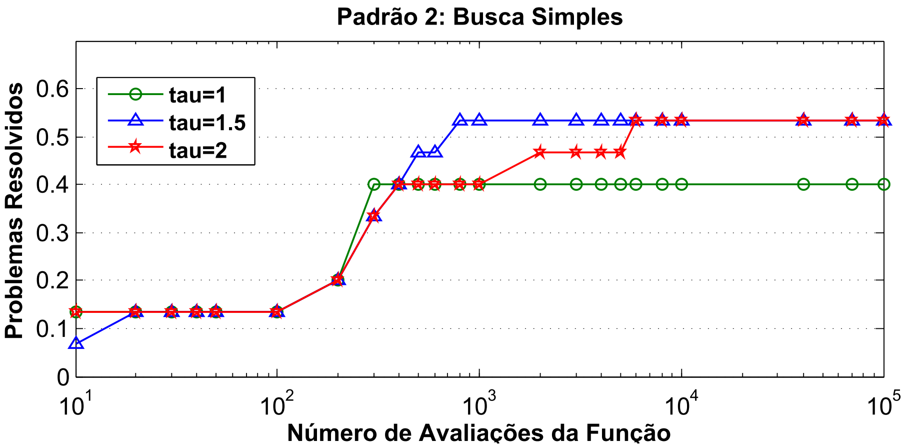


Figura: Problemas não suaves: Padrão 2 - Busca Simples

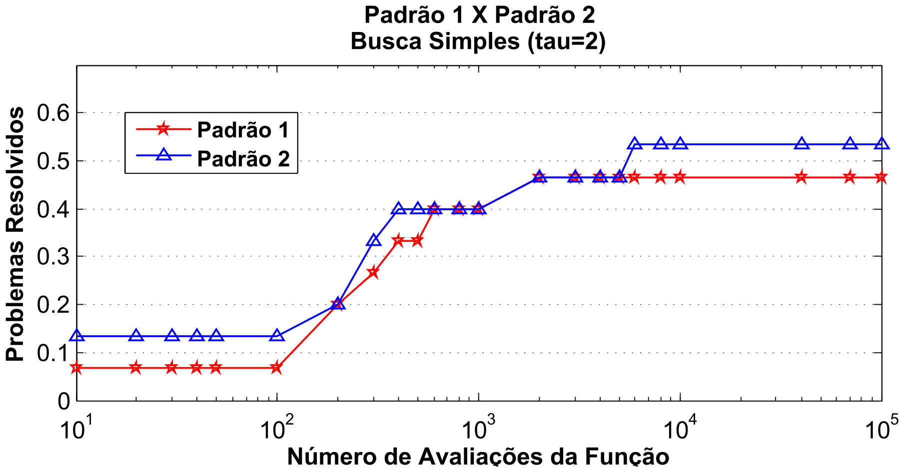


Figura: Padrão 1 $\times$ Padrão 2: Problemas não suaves - Busca Simples ($\tau = 2$)

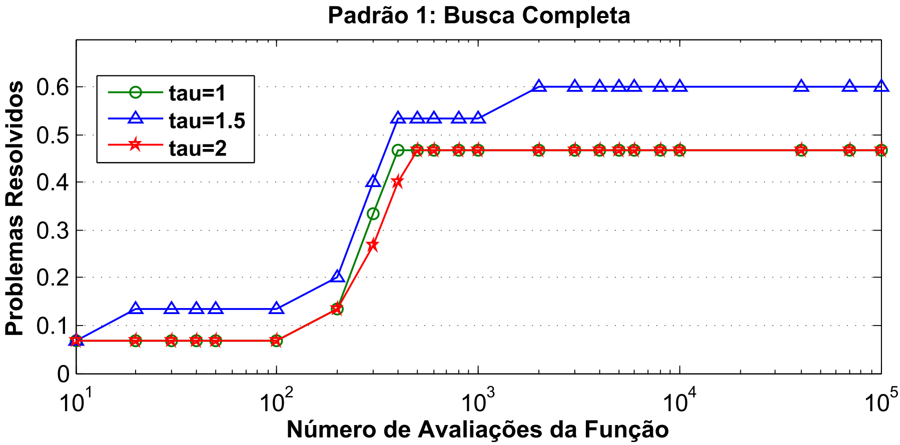


Figura: Problemas não suaves: Padrão 1 - Busca Completa

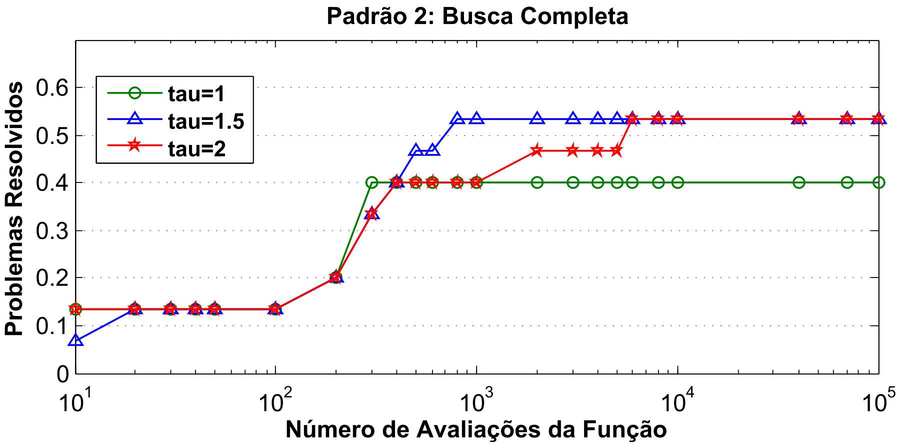


Figura: Problemas não suaves: Padrão 2 - Busca Completa

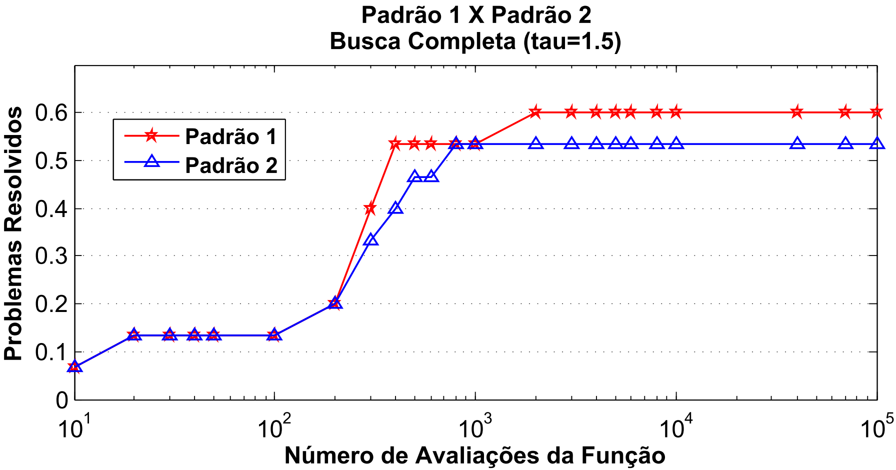


Figura: Padrão 1 $\times$ Padrão 2 - Busca Completa ($\tau = 1.5$)

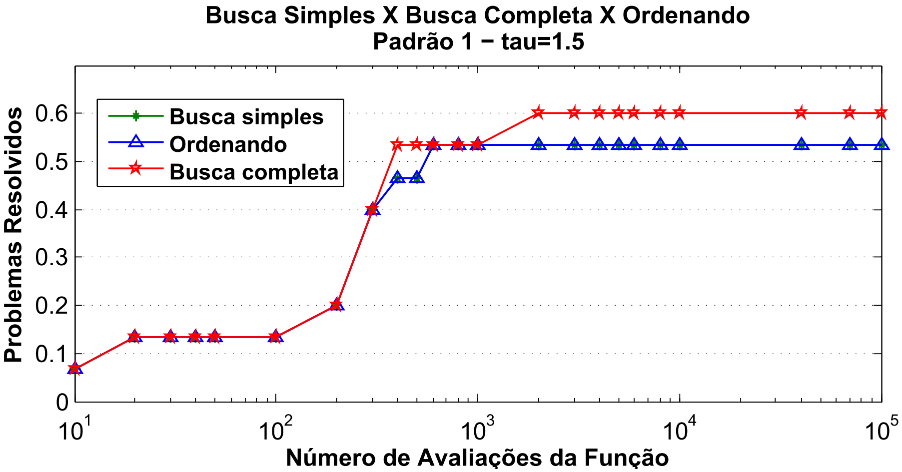


Figura: Problemas não suaves: Padrão 1 - Estratégias de Busca ($\tau = 1.5$)

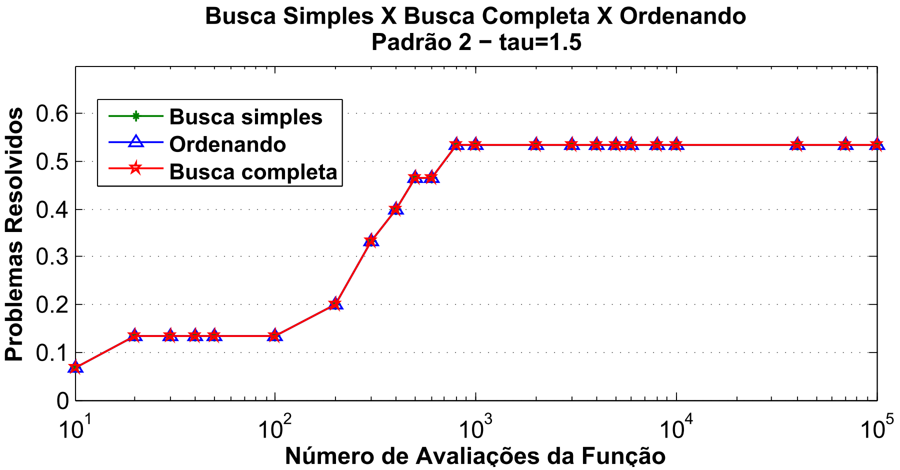


Figura: Problemas não suaves: Padrão 2 - Estratégias de Busca ($\tau = 1.5$)

Modelo suave para problemas Minimax

- O modelo que utilizamos foi proposto em [1] e é dado por:

$$f(x,\mu) = f(x) + \mu ln \sum_{i=1}^{q} exp\left(\frac{f_i(x) - f(x)}{\mu}\right).$$

- Com o uso do modelo esperávamos melhorar a robustez do método;
- No entanto a representação da função através de um modelo já possui um erro;
- Com a precisão considerada nos demais testes os resultados obtidos foram muito ruins.

---

[1] G. Liuzzi, S. Lucide, M. Sciandrone. A derivative-free algorithm for linearly constrained finite minimax problems. SIAM Journal on Optimization 16, pp. 1054-1075, (2006).

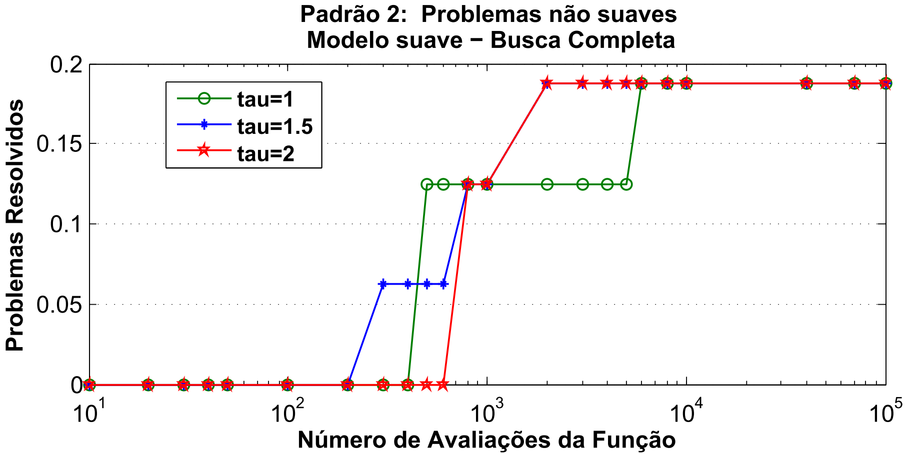


Figura: Busca completa utilizando o modelo exponencial - precisão 10^{-7}

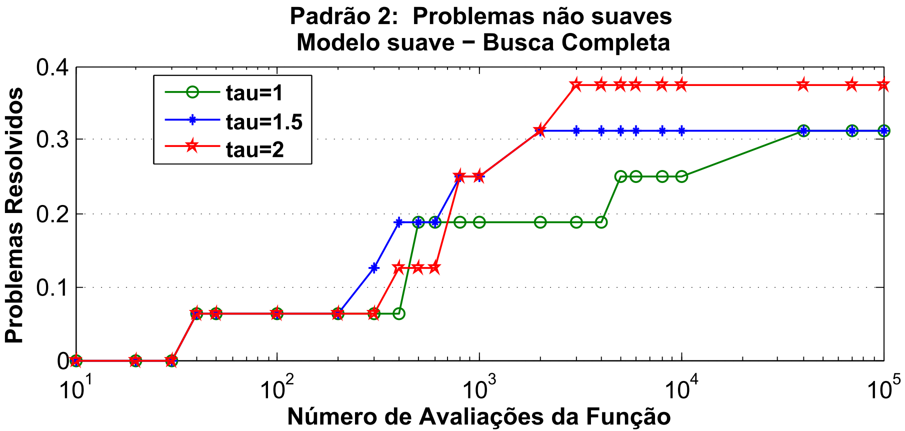


Figura: Busca completa utilizando o modelo exponencial - precisão 10^{-2}

Desempenho Geral do Método

Número total de avalições de Função						
Estratégia de busca	$\tau = 1$		$\tau = 1.5$		$\tau = 2$	
	Padrão 1	Padrão 2	Padrão 1	Padrão 2	Padrão 1	Padrão 2
Busca Simples	1157569	922526	954809	716911	1079561	681250
Ordenando o Padrão	1159984	919083	889906	652220	1104810	705836
Busca Completa	976152	798572	599952	524204	778933	545604

Tabela: Número total de Avaliações de Função

Problemas com restrições degeneradas

- Em [1] os autores reportam como problemas degenerados problemas contendo apenas restrições de igualdade e de caixas;
- Em nossa abordagem apenas problemas contendo restrições de desigualdade podem apresentar conjunto de trabalho degenerado;
- Os autores acrescentam as restrições de caixas à matriz de restrições;
- Em nossa abordagem não consideramos as restrições de caixa juntamente com as restrições do problema;
- Apenas fiscalizamos para que estas sejam sempre satisfeitas;
- Considerando restrições de caixa como restrições de desigualdade do problema muitos problemas não degenerados passam a ser degenerados;
- O desempenho do método cai drásticamente como podemos observar pela figura.

---

[1] R. M. Lewis, A. Shepherd, V. Torczon. Implementing generating set search methods for linearly constrained minimization. SIAM Journal on Scientific Computing 29, 6, pp. 2507-2530 (2007).

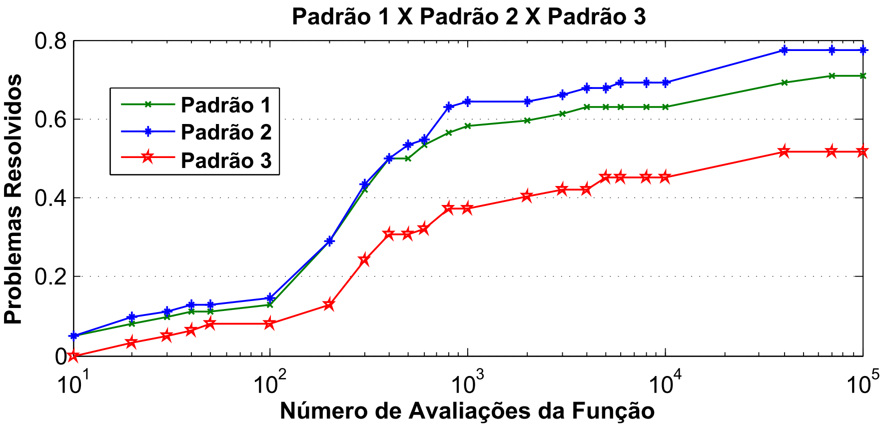


Figura: Teste de viabilidade

Problemas Degenerados				
Problema	# variáveis	# Desigualdades	Lower	Uper
#118 [3]	15	29	15	15
Pentagon [2]	6	15	0	0
Hatfldh [1]	4	7	4	4

Tabela: Problemas Degenerados

[1] D. Orban, N.I.M. Gould, P.L. Toint. CUTEr: Constrained and Unconstrained Testing Environment, revisited,http://www.cuter.rl.ac.uk (2011).

[2] L. Luksan, J. Vlcek. Test Problems for Nonsmooth Unconstrained and Linearly Constrained Optimization. Technical report No. 798, (2000).

[3] W. Hock, K. Schittkowski. "Test examples for nonlinear programming code". Springer, Berlin, (1981).

Busca Simples - $\tau = 2$				
Problema	$f(\bar{x}) - f(x^{\star})$		# avaliações da Função	
	Padrão 1	Padrão 2	Padrão 1	Padrão 2
#118 [7]	2.1224e+0	-2.0540e-8	100003	4596
Pentagon [3]	4.0689e-4	2.0224e-9	2996	2096
Hatfldh [1]	0.0000e+0	0.0000e+0	128	131

Tabela: Problemas Degenerados - Busca Simples

Busca Completa - $\tau = 2$				
Problema	$f(\bar{x}) - f(x^{\star})$		# avaliações da Função	
	Padrão 1	Padrão 2	Padrão 1	Padrão 2
#118 [7]	1.4473e+0	-4.8278e-8	42155	15149
Pentagon [3]	3.6431e-3	3.4565e-2	1539	1717
Hatfldh [1]	0.0000e+0	0.0000e+0	132	137

Tabela: Problemas Degenerados - Busca Completa

Considerações Finais

- O método estudado é simples e exigindo muito pouco do problema para ser aplicado possui teoria de convergência global;
- Obtemos uma excelente precisão na solução;
- As estratégias que propomos obtiveram bons resultados;
- O Padrão 2 que propomos se mostrou muito superior em praticamente todos os testes;
- Com busca simples vale a pena ordenar o Padrão de direções;
- Ainda que a convergência seja lenta o método se mostrou bastante robusto
- Não foi possível observar uma estratégia ótima para todos os problemas;
- Nada podemos concluir a respeito do Padrão proposto para problemas degenerados;

Referências Bibliográficas

- D. Orban, N.I.M. Gould, P.L. Toint. CUTEr: Constrained and Unconstrained Testing Environment, revisited,*http://www.cuter.rl.ac.uk* (2011).

- G. Liuzzi, S. Lucide, M. Sciandrone. A derivative-free algorithm for linearly constrained finite minimax problems. SIAM Journal on Optimization 16, pp. 1054-1075, (2006).

- L. Luksan, J. Vlcek. Test Problems for Nonsmooth Unconstrained and Linearly Constrained Optimization. Technical report No. 798, (2000).

- V. Torczon. On the convergence of pattern search algorithms. SIAM Journal on Optimization 7, pp. 1-25 (1997).

- R. M. Lewis, V. Torczon. Pattern search algorithms for linearly constrained minimization. SIAM Journal on Optimization 10, pp. 917-941 (2000).

- R. M. Lewis, A. Shepherd, V. Torczon. Implementing generating set search methods for linearly constrained minimization. SIAM Journal on Scientific Computing 29, 6, pp. 2507-2530 (2007).

- W. Hock, K. Schittkowski. “Test examples for nonlinear programming code”. Springer, Berlin, (1981).

OBRIGADA!

Critérios de Parada

- É intuitivo tomar como critério de parada $|\alpha_k| < tol$;
- Deixar o tamanho do passo suficientemente pequeno garante que estamos próximos de um ponto estacionário?

Proposição 1:

Seja $x \in \Omega$. Defina $q(x) = P_\Omega(x - \nabla f(x)) - x$. Então

$$\|q(x)\| \leq \|\nabla f(x)\|$$

e x é um ponto estacionário para o problema com restrições lineares (PRL) se e somente se $q(x) = 0$.

Critérios de Parada

Proposição 2:

Suponha que $\nabla f(x)$ é Lipschitz contínua em $L_\Omega(x^0)$, com constante de Lipschitz C. Se x^k é o ponto obtido numa iteraçãoo onde houve fracasso, então existe uma constante $\hat{c} > 0$ que satisfaz:

$$\left\| q(x^k) \right\|^2 \leq \hat{c}\alpha_k$$

- Logo o critério de parada mais intuitivo que pode ser tomado é um bom critério de parada, pois temos garantias que de fato estamos próximos da solução.
- Adicionalmente paramos o algoritmo ao atingir um número máximo de avaliações da função objetivo.

Hipóteses - "Fracas"

1. O Padrão $P^k = \left[\Gamma^k \; L^k\right] \in \mathbb{Z}^{n \times p^k}$, $p^k > n+1$, Então todas as direções de busca são inteiras, multiplicadas por $\alpha_k \in \mathbb{R}^n$;
2. $\Gamma^k \in \mathbb{Z}^{n \times r^k}$, $r^k \geq n+1$, pertence à Γ, que é um conjunto finito de matrizes inteiras cujas colunas incluem geradores para todos os cones $K^o(x^k, \epsilon)$;
3. A matriz $L^k \in \mathbb{Z}^{n \times (p^k - r^k)}$ contém pelo menos uma coluna (nula);
4. A atualização de α_k segue as regras especificadas;
5. São satisfeitas as hipóteses sobre os movimentos exploratórios;
6. A matriz de restrições A é racional;
7. O conjunto de nível $L_\Omega(x^0) = \left\{x \in \Omega \,\middle|\, f(x) \leq f(x^0)\right\}$ é compacto;
8. A função objetivo $f(x)$ é continuamente diferenciável em uma vizinhança aberta D de $L_\Omega(x^0)$.

Resultado de Convergência

Teorema

Assuma que as hipóteses 1-8 são satisfeitas. Seja $\{x^k\}$ uma sequência gerada pelo Método de busca Padrão Genaralizado para minimização com restrições lineares, então:

$$\liminf_{k\to\infty} \left\| q(x^k) \right\| = 0$$

Corolário

Existe um ponto limite de $\{x^k\}$ que é um ponto KKT para o problema (PRL).

Hipóteses - "Fortes"

1. O Padrão $P^k = \left[\Gamma^k \; L^k\right] \in \mathbb{Z}^{n \times p^k}$, $p^k > n+1$, Então todas as direções de busca são inteiras, multiplicadas por $\alpha_k \in \mathbb{R}^n$;
2. $\Gamma^k \in \mathbb{Z}^{n \times r^k}$, $r^k \geq n+1$, pertence à Γ, que é um conjunto finito de matrizes inteiras cujas colunas incluem geradores para todos os cones $K^o(x^k, \epsilon)$;
3. A matriz $L^k \in \mathbb{Z}^{n \times (p^k - r^k)}$ contém pelo menos uma coluna (nula);
4. A atualização de α_k segue as regras especificadas;
5. São satisfeitas as hipóteses fortes sobre os movimentos exploratórios;
6. A matriz de restrições A é racional;
7. O conjunto de nível $L_\Omega(x^0) = \left\{x \in \Omega \,\middle|\, f(x) \leq f(x^0)\right\}$ é compacto;
8. A função objetivo $f(x)$ é continuamente diferenciável em uma vizinhança aberta D de $L_\Omega(x^0)$.
9. As colunas do padrão P^k permanecem limitadas em norma;
10. temos que: $\lim_{k \to \infty} \alpha_k = 0$

Resultado de Convergência - “ Forte”

Teorema

Assuma que as hipóteses fortes 1-10 são satisfeitas. Seja $\{x^k\}$ uma sequência gerada pelo Método de busca Padrão Genaralizado para minimização com restrições lineares, então:

$$\lim_{k \to \infty} \left\| q(x^k) \right\| = 0$$

Corolário

Todo ponto limite de $\{x^k\}$ é um ponto KKT para o problema (PRL).