

UM ESTUDO NUMÉRICO DO MÉTODO DOS GRADIENTES CONJUGADOS PROJETADOS

Wéllington Moutinho Dias¹, Marcio Antônio de Andrade Bortoloti²

Universidade Estadual do Sudoeste da Bahia, BA, Brasil

wellingtonmoutinhodias@gmail.com¹, mbortoloti@uesb.edu.br²

Resumo

Neste trabalho apresentamos um estudo numérico do Método dos Gradientes Conjugados Projetados, para encontrar zeros de campos vetoriais continuamente diferenciáveis e monótonos, sujeito a um conjunto viável não-vazio, fechado e convexo. Consideramos dois campos vetoriais e dois conjuntos viáveis. Analisamos o tempo de CPU, o número de avaliação de função e o número de iterações com relação a dois parâmetros de controle que atuam nas direções conjugadas.

Palavras-Chave: Otimização contínua; Método dos Gradientes Conjugados; Zeros de campos vetoriais.

Introdução

Métodos dos gradientes conjugados são métodos iterativos que podem ser utilizados na minimização irrestrita de funções $f: \mathbb{R}^n \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 . Diversos autores têm estudado esses métodos, em especial Dai e Kou, [1], que propuseram um método para minimizar funções escalares sem restrições. Tal método foi estendido por Ding, Xiao e Li, [2], para determinar zeros de campos vetoriais com restrições. Formalmente, esse método, o qual chamaremos de Método dos Gradientes Conjugados Projetados (MGCP), resolve problemas do tipo

$$\mathcal{F}(x) = 0, x \in \Omega, \quad (1)$$

onde Ω (denominado conjunto viável) é um conjunto não-vazio, fechado e convexo e $\mathcal{F}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ é Lipschitz contínua, continuamente diferenciável e monótona, isto é, satisfaz a desigualdade $(\mathcal{F}(x) - \mathcal{F}(y))^\top (x - y) \geq 0$ para todo $x, y \in \mathbb{R}^n$. O MGCP proposto por [2] é uma variante do método proposto em [1] que gera uma sequência $\{x_k\}_{k \in \mathbb{N}}$ construída por meio do emprego de direções conjugadas d_k . Além disso, o método garante que $x_k \in \Omega$, para todo $k \in \mathbb{N}$, através da aplicação de um operador projeção $P_{\Omega}[\cdot]: \mathbb{R}^n \rightarrow \Omega$, definido por $P_{\Omega}[x] = \arg \min\{\|y - x\|; y \in \Omega\}$, para todo $x \in \mathbb{R}^n$.

O Método

Nesta seção apresentamos o Método dos Gradientes Conjugados Projetados, que estamos estudando, para determinar zeros de campos vetoriais. Esse método consiste na construção de direções conjugadas para a geração de uma sequência que espera-se convergir para a solução do problema. Para cada $k \in \mathbb{N}$, a direção conjugada d_k , que foi apresentada em [2], satisfaz a propriedade $\mathcal{F}(x_k)^\top d_k \leq -c\|\mathcal{F}(x_k)\|^2$ onde $c \in (0, 1)$ é uma constante. Essa direção é definida como

$$d_k = \begin{cases} -\mathcal{F}(x_0), & \text{se } k = 0 \\ -\mathcal{F}(x_k) + \beta_k^+(\tau_{k-1})d_{k-1} & \text{se } k \geq 1, \end{cases} \quad (2)$$

onde o parâmetro $\beta_k^+(\tau_{k-1})$ é dado por

$$\beta_k^+(\tau_{k-1}) = \max \left\{ \beta_k(\tau_{k-1}), \eta \frac{\mathcal{F}(x_k)^\top d_{k-1}}{\|d_{k-1}\|^2} \right\}, \quad (3)$$

com $\eta \in [0, 1]$, $\beta_k(\tau_{k-1})$ determinado como

$$\beta_k(\tau_{k-1}) = \frac{\mathcal{F}(x_k)^\top y_{k-1}}{d_{k-1}^\top y_{k-1}} - \left(\tau_{k-1} + \frac{\|y_{k-1}\|^2}{s_{k-1}^\top y_{k-1}} - \frac{s_{k-1}^\top y_{k-1}}{\|s_{k-1}\|^2} \right) \frac{\mathcal{F}(x_k)^\top s_{k-1}}{d_{k-1}^\top y_{k-1}},$$

e τ_{k-1} é definida por

$$\tau_{k-1} = \theta \frac{\|y_{k-1}\|^2}{s_{k-1}^\top y_{k-1}} + (1 - \theta) \frac{s_{k-1}^\top y_{k-1}}{\|s_{k-1}\|^2}, \theta \in [0, 1]. \quad (4)$$

Ademais, $y_k = \gamma_k + \lambda_k t_k \|\mathcal{F}(x_k)\| d_k$, com $\gamma_k = \mathcal{F}(x_{k+1}) - \mathcal{F}(x_k)$, $s_k = t_k d_k$, com $t_k > 0$ e

$$\lambda_k = 1 + \|\mathcal{F}(x_k)\|^{-1} \max \left\{ 0, -\frac{\gamma_k^\top d_k}{t_k \|d_k\|^2} \right\}.$$

O Algoritmo abaixo apresenta formalmente o MGCP proposto por [2].

Algoritmo 1: MGCP

Passo 0 Escolha $x_0 \in \Omega$; $\epsilon > 0, \xi, \sigma, \rho \in (0, 1)$; $\eta \in [0, 1]$ e $\theta \in [0, 1]$. Faça $k := 0$;

Passo 1 Pare se $\|\mathcal{F}(x_k)\| < \epsilon$. Caso contrário, determine d_k por (2);

Passo 2 Encontre $t_k = \max\{\xi \rho^i : i = 0, 1, 2, \dots\}$, que satisfaça

$$-\mathcal{F}(x_k + t_k d_k)^\top d_k \geq \sigma t_k \|d_k\|^2. \quad (5)$$

Faça $z_k := x_k + t_k d_k$;

Passo 3 Se $\|\mathcal{F}(z_k)\| < \epsilon$, pare. Caso contrário, calcule $x_{k+1} = P_{\Omega}[x_k - \alpha_k \mathcal{F}(z_k)]$, em que

$$\alpha_k = \mathcal{F}(z_k)^\top (x_k - z_k) / \|\mathcal{F}(z_k)\|^2;$$

Passo 4 Faça $k := k + 1$. Retorne ao *Passo 1*.

Um estudo sobre as propriedades de convergência é apresentado em [2], onde é mostrado que o método converge globalmente com taxa linear. Além disso, também é provado que a busca (5) no **Passo 2** é bem definida.

Resultados Numéricos

Nesta seção vamos apresentar um estudo numérico do comportamento do MGCP em relação aos parâmetros η em (3) e θ em (4), que apresentam maior influência no desempenho do método. Para esse estudo consideramos os campos gradientes das funções *Sum Squares Function* e *Trid Function* disponibilizadas no endereço eletrônico <https://www.sfu.ca/~ssurjano/optimization.html>. Eles são definidos, respectivamente, por $\mathcal{F}, \mathcal{G}: \mathbb{R}^n \rightarrow \mathbb{R}^n$, com $\mathcal{F}(x) = 2x$ para todo $x \in \mathbb{R}^n$ e $\mathcal{G}(x) = (y_1, y_2, \dots, y_n)$ onde

$y_1 = 2(x_1 - 1) - x_2$, $y_i = 2(x_i - 1) - x_{i-1} - x_{i+1}$ para $1 < i < n$ e $y_n = 2(x_n - 1) - x_{n-1}$ para todo $x \in \mathbb{R}^n$. Em relação aos conjuntos viáveis, vamos considerar $\Omega_1 = \mathbb{R}^n$ e $\Omega_2 = \prod_{i=1}^n [0, +\infty)$. O problema (1) sujeito a Ω_1 é considerado irrestrito; portanto no **Passo 3** adotamos $P_{\Omega_1}[x] = x$ para todo $x \in \mathbb{R}^n$. De acordo com [3], definimos $P_{\Omega_2}[x] = (z_1, \dots, z_n)$, onde $z_i = \max\{0, x_i\}$, com $1 \leq i \leq n$ para todo $x \in \mathbb{R}^n$. Foram tomados os parâmetros $\epsilon = \sigma = 10^{-5}$, $\xi = 0.9$, $\rho = 0.74$, $\eta = 0.5$ e $\theta \in \{0.1, 0.3, 0.5\}$. Por fim, o algoritmo é interrompido quando o número de iterações atinge 55000 ou quando as condições de parada apresentadas nos passos 1 e 3 são satisfeitas. Os códigos utilizados neste trabalho foram implementados na Linguagem de Programação Julia e estão livremente disponibilizados em <https://github.com/petimatematica/CMECG>.

Inicialmente, analisamos o problema de determinar os zeros dos campos vetoriais \mathcal{F} e \mathcal{G} no conjunto viável Ω_1 . Esses resultados podem ser vistos nas figuras 1 a 3. Na Figura 2, observamos que o MGCP, com o parâmetro $\theta = 0.5$, resolveu cerca de 80% dos problemas com uma quantidade menor de iterações. Contudo a Figura 3 mostra que esse caso utilizou uma maior quantidade de avaliações de função. Isto contribuiu para piorar seu desempenho com relação ao tempo, como pode ser visto na Figura 1.

Agora, faremos o estudo do método para o conjunto viável Ω_2 . Ao analisarmos a Figura 5, observamos que o MGCP, com o parâmetro $\theta = 0.5$, resolveu 100% dos problemas com uma quantidade menor de iterações. Além disso, observando a Figura 4, notamos que ele apresentou bom desempenho com relação ao tempo de CPU. Por outro lado, na Figura 6, observamos que o caso com $\theta = 0.5$ apresentou pior desempenho com relação à quantidade de avaliações de função. Isso evidencia um emprego maior da busca. Contudo, isto não foi relevante o suficiente para prejudicar seu desempenho.

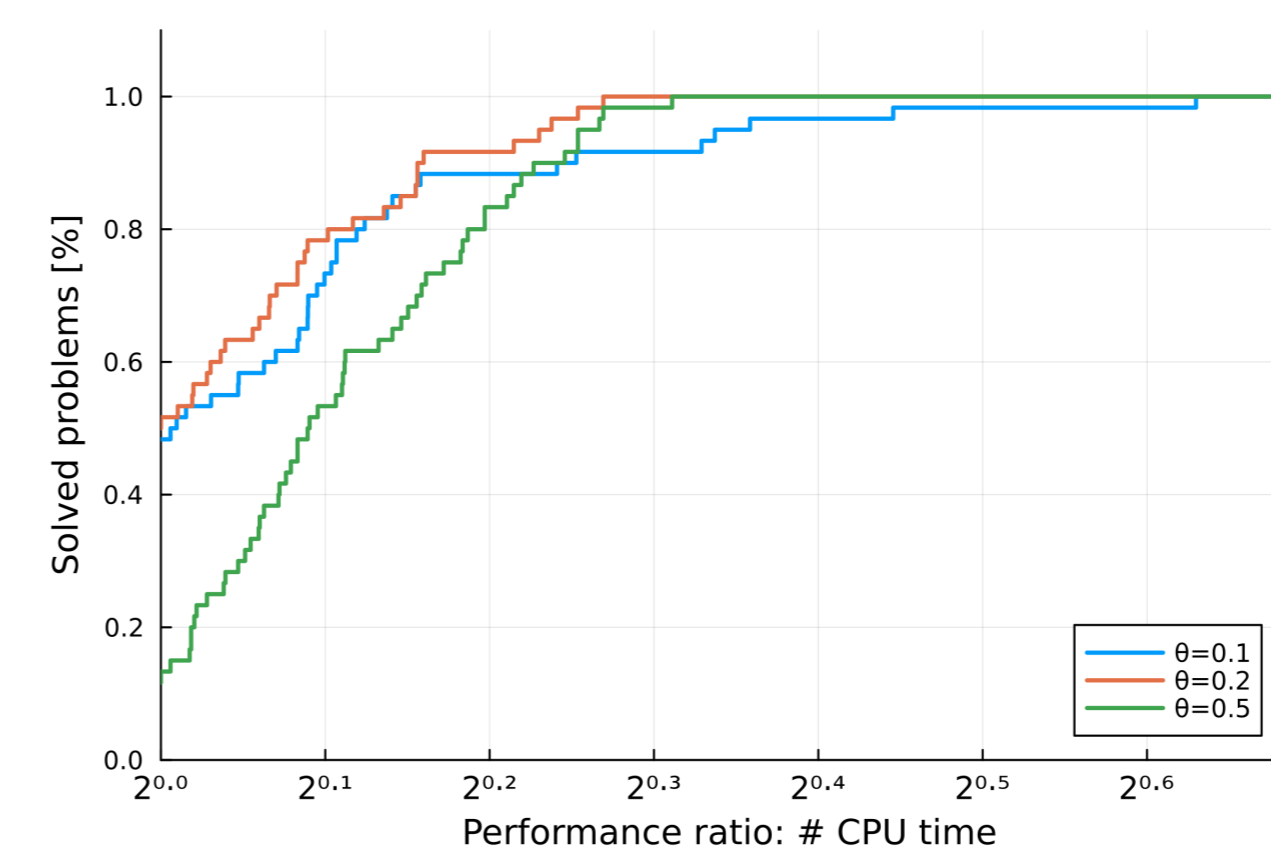


Figura 1: Performance profile relativo ao tempo de CPU considerando $\Omega = \Omega_1$.

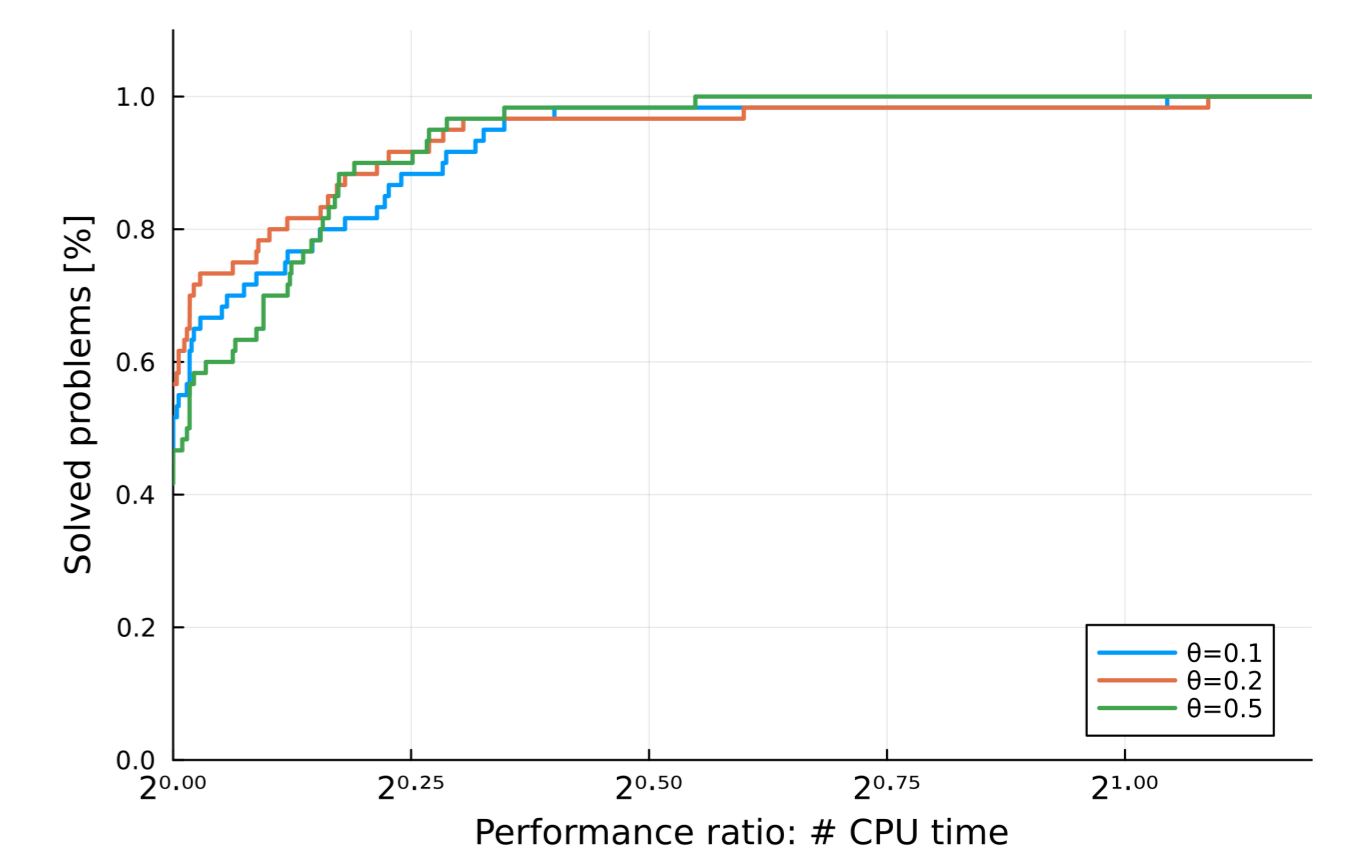


Figura 4: Performance profile relativo ao tempo de CPU considerando $\Omega = \Omega_2$.

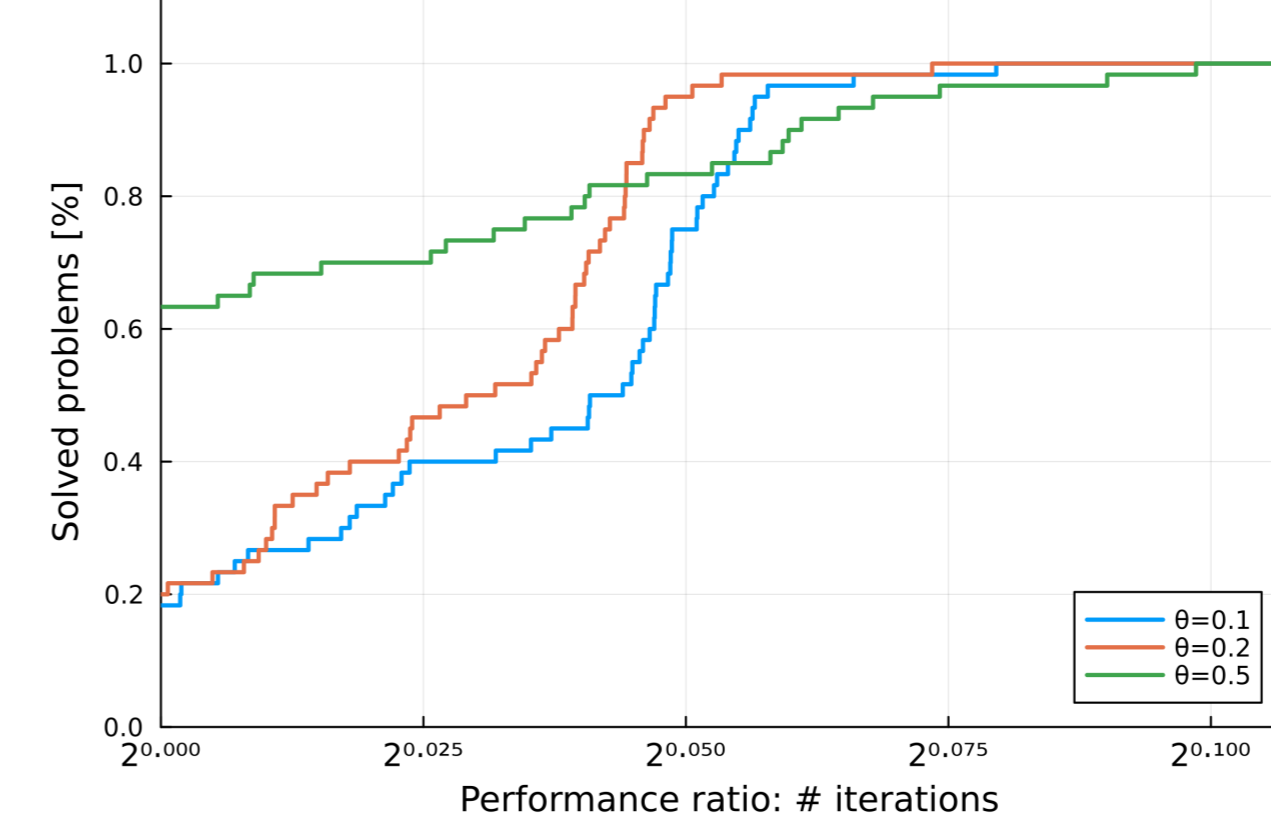


Figura 2: Performance profile relativo às iterações considerando $\Omega = \Omega_1$.

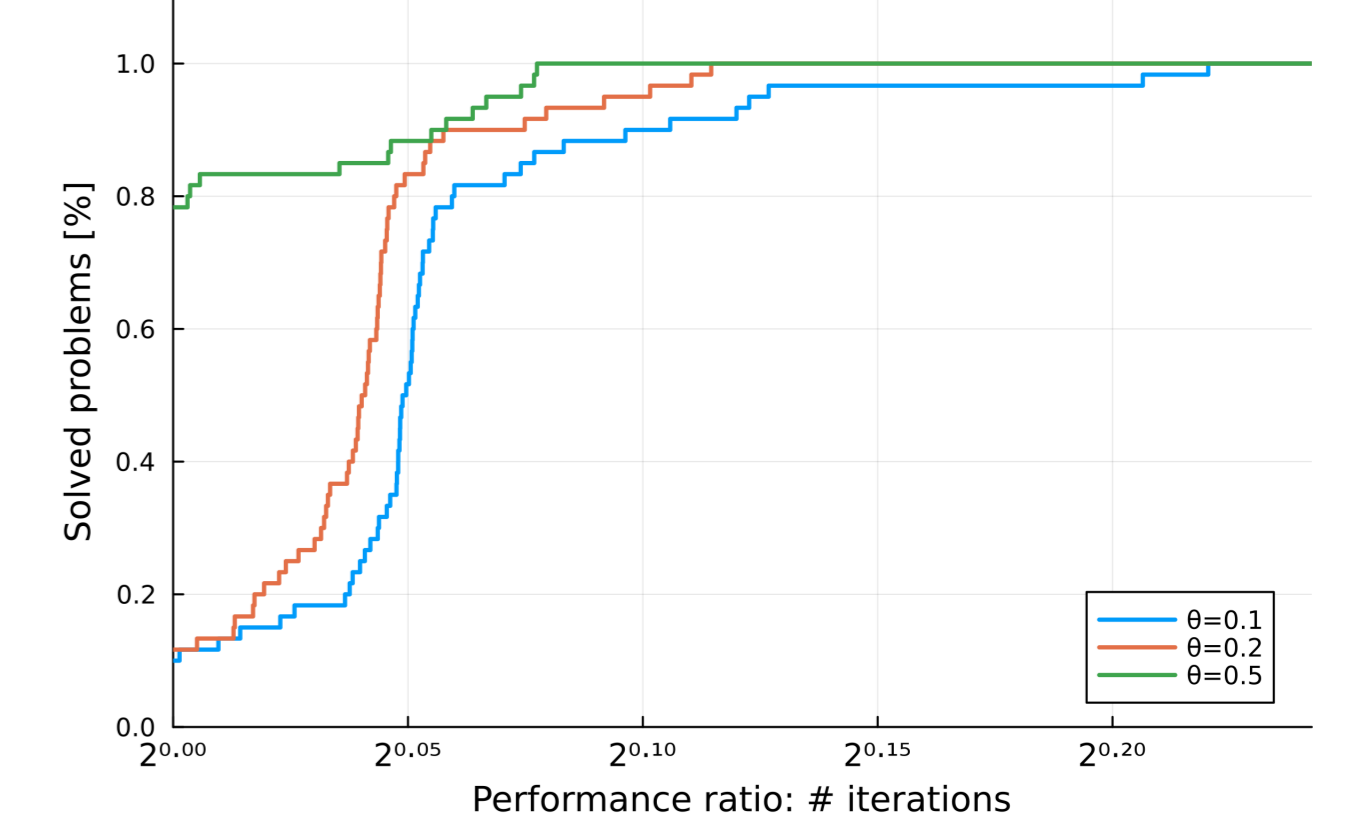


Figura 5: Performance profile relativo às iterações considerando $\Omega = \Omega_2$.

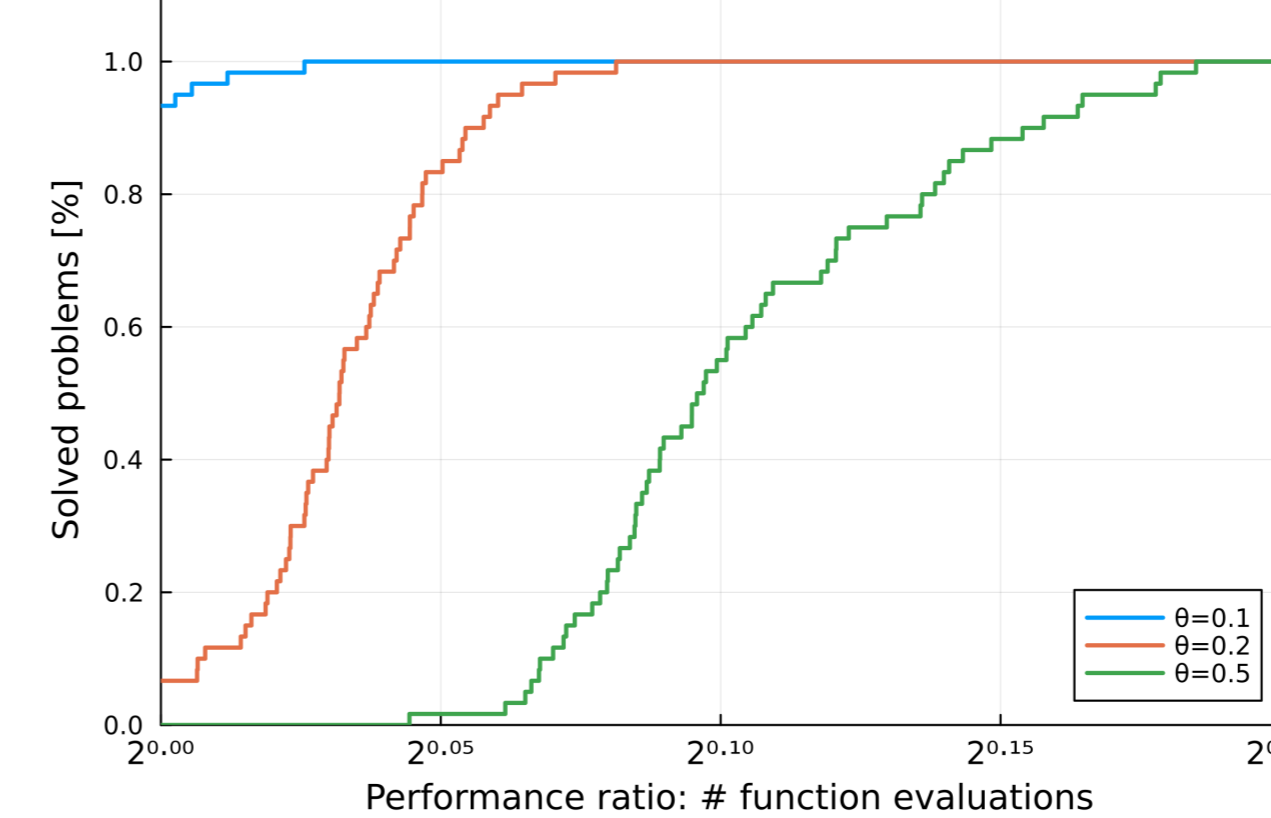


Figura 3: Performance profile relativo à avaliação de função considerando $\Omega = \Omega_1$.

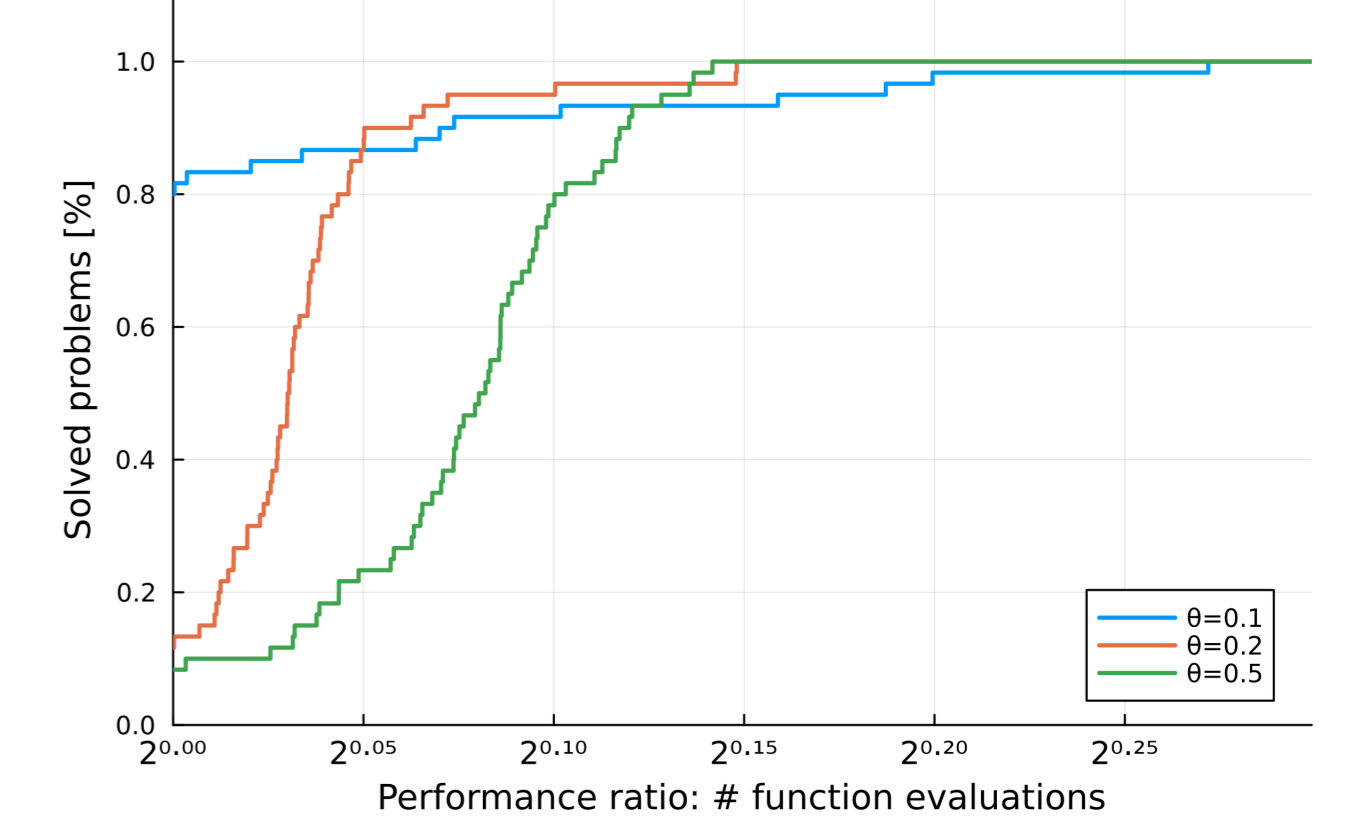


Figura 6: Performance profile relativo à avaliação de função considerando $\Omega = \Omega_2$.

Referências

- [1] Dai YH, Kou CX. A nonlinear conjugate gradient algorithm with an optimal property and an improved wolfe line search. *SIAM J Optim.*, 2013;23:296-320.
- [2] Yanyun Ding, Yunhai Xiao e Jianwei Li. A class of conjugate gradient methods for convex constrained monotone equations. Em: **Optimization** 66.12 (2017), pp. 2309-2328.
- [3] Alexey Izmailov e Mikhkhail Solodov. **Otimização - volume 2: métodos computacionais**. Terceira Edição. IMPA, 2018. ISBN: 978-85-244-0454-2.

Agradecimentos

O autor Wéllington Moutinho Dias agradece ao **Programa de Educação Tutorial Institucional da Universidade Estadual do Sudoeste da Bahia (PETI/UESB)** pela bolsa de estudo e à **UESB** pelo apoio financeiro na hospedagem e na alimentação.