

Estudo numérico entre métodos do Gradiente Projetado na minimização restrita de uma função quadrática

Emanuel Mendes Queiroz¹, Samara Viriato Vilar Dias², Marcio Antônio de Andrade Bortoloti³
emanuelmqueiroz.emq@gmail.com¹, samaravvilar@gmail.com², mbortoloti@uesb.edu.br³

Universidade Estadual do Sudoeste da Bahia - Vitória da Conquista, BA, Brasil

Resumo

Neste trabalho, realizamos um estudo comparativo entre métodos do Gradiente Projetado e do Gradiente Espectral Projetado. Para os métodos do Gradiente Projetado, utilizamos a busca de Armijo ao longo do arco de projeção e ao longo de direções viáveis. Já nos métodos do Gradiente Espectral Projetado, utilizamos o passo espectral e buscas lineares não monótonas também ao longo do arco de projeção e das direções viáveis. Em nossos experimentos, empregamos esses métodos para a resolução de problemas de programação quadrática. Os resultados evidenciaram que os métodos do Gradiente Espectral Projetado apresentam um desempenho superior em relação aos métodos do Gradiente Projetado, considerando o tempo de CPU, o número de iterações e a avaliação de função e de projeção.

Introdução

O desenvolvimento de métodos computacionais para a minimização de funções com restrições é fundamental para a resolução de muitos problemas práticos. Em geral, esses métodos buscam resolver o problema de minimizar f em Ω , onde $f: \Omega \rightarrow \mathbb{R}$ é uma função diferenciável e $\Omega \subset \mathbb{R}^n$ é convexo e fechado. Métodos importantes para a resolução desse problema são os do Gradiente Projetado, que podem ser vistos como uma combinação de métodos do Gradiente com a projeção em Ω dos termos obtidos no processo iterativo. A convexidade de Ω torna possível a utilização da projeção ortogonal $P_{\Omega}: \mathbb{R}^n \rightarrow \Omega$ a fim de se obter direções viáveis, as quais também são de descida. O esquema iterativo desses métodos é dado por

$$x_{k+1} = x_k + \gamma_k(x_k(\tau_k) - x_k), \quad (1)$$

onde $\tau_k, \gamma_k > 0$ e $x_k(\tau_k) = P_{\Omega}(x_k - \tau_k \nabla f(x_k))$. Uma possibilidade de escolha para τ_k foi proposta por [1] e é conhecida como passo espectral. Segundo [3], esse comprimento de passo requer pouco esforço computacional e acelera a convergência de métodos do Gradiente no contexto de problemas com restrição. A utilização do passo espectral e de uma busca linear não monótona caracteriza os métodos do Gradiente Espectral Projetado. Neste trabalho, realizamos um estudo numérico entre os métodos do Gradiente Projetado e do Gradiente Espectral Projetado apresentados por [3] e [4] na minimização de uma função quadrática.

Métodos do Gradiente Projetado

Uma estratégia de método do Gradiente Projetado consiste em utilizar a busca de Armijo ao longo do arco de projeção. Para o emprego dessa busca, dados $\bar{\alpha} > 0$ e $\theta, \eta \in (0, 1)$, escolhemos α_k como o maior entre todos os números da forma $\alpha = \bar{\alpha}\theta^i$, $i = 0, 1, 2, \dots$, que satisfaça

$$f(x_k(\alpha_k)) \leq f(x_k) + \eta \nabla f(x_k)^{\top} d_k, \quad (2)$$

onde $d_k = x_k(\alpha_k) - x_k$ é uma direção de descida. Outra estratégia, também baseada na busca de Armijo, mas ao longo das direções viáveis, considera a direção $d_k = x_k(\beta_k) - x_k$, onde, dados $\hat{\beta} \geq \beta > 0$, toma-se $\beta_k \in [\beta, \hat{\beta}]$, para $k \in \mathbb{N}$. Neste caso, dados $\bar{\alpha} > 0$, $\theta, \eta \in (0, 1)$ e considerando a desigualdade

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \eta \alpha_k \nabla f(x_k)^{\top} d_k, \quad (3)$$

escolhemos α_k sendo o maior entre todos os números da forma $\alpha = \bar{\alpha}\theta^i$, $i = 0, 1, 2, \dots$, que satisfaça essa desigualdade. Na primeira estratégia (método PG1), tomamos em (1) os parâmetros $\gamma_k = 1$ e $\tau_k = \alpha_k$ dado por (2), para $k \in \mathbb{N}$. Na segunda estratégia (método PG2), tomamos em (1) o parâmetro $\gamma_k = \alpha_k$ dado por (3) e $\tau_k = \beta_k \in [\beta, \hat{\beta}]$, para $k \in \mathbb{N}$. A seguir, apresentamos formalmente o algoritmo referente aos métodos PG1 e PG2.

Algoritmo 1: MÉTODO DO GRADIENTE PROJETADO

- 1 Sejam $x_0 \in \Omega$, $\theta, \eta \in (0, 1)$, $\varepsilon > 0$, $\bar{\alpha} > 0$, $\hat{\beta} \geq \beta > 0$, $\beta_0 \in [\beta, \hat{\beta}]$ e $k = 0$.
- 2 Se $\|x_0 - x_0(1)\| < \varepsilon$, então pare e declare que x_0 é um ponto estacionário.
- 3 Determine α_k por (2) e calcule $x_{k+1} = x_k(\alpha_k)$ para o PG1; ou determine $\beta_k \in [\beta, \hat{\beta}]$, α_k por (3) e calcule $x_{k+1} = x_k + \alpha_k(x_k(\beta_k) - x_k)$ para o PG2.
- 4 Se $\|x_k - x_{k+1}\| < \varepsilon$, então pare e declare que x_k é um ponto estacionário.
- 5 Faça $k \leftarrow k + 1$ e retorne para o passo 3.

Ressaltamos que, neste trabalho, o cálculo do parâmetro β_k foi feito por interpolação quadrática. Os resultados de convergência para o Algoritmo 1 podem ser vistos em [2] e [4].

Métodos do Gradiente Espectral Projetado

O método do Gradiente Espectral Projetado, proposto por [3], representa uma variante do método do Gradiente Projetado. Ele incorpora duas adaptações em relação a esse método: a utilização do passo espectral, proposto por [1], e de uma busca linear não monótona. Para a escolha da busca linear não monótona, foram sugeridas em [3] duas estratégias. A primeira estratégia é empregar uma busca linear não monótona, veja [6], ao longo do arco de projeção. Neste caso, dados $\eta \in (0, 1)$ e $M \in \mathbb{N}$, busca-se determinar um comprimento de passo $\alpha_k > 0$ tal que

$$f(x_k(\alpha_k)) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \eta \nabla f(x_k)^{\top} d_k, \quad (4)$$

onde $d_k = x_k(\alpha_k) - x_k$. Nota-se que, em (2) e (4), o produto $\nabla f(x_k)^{\top} d_k$ é calculado para cada ponto teste $x_k(\alpha_k)$. Segundo [3], este emprego sucessivo de projeções causa um aumento no esforço computacional. Por isso, foi proposta uma nova estratégia para a busca linear não monótona. Essa estratégia consiste em, dados $\eta \in (0, 1)$ e $M \in \mathbb{N}$, determinar $\alpha_k > 0$ tal que

$$f(x_k + \alpha_k d_k) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x_{k-j}) + \eta \alpha_k \nabla f(x_k)^{\top} d_k, \quad (5)$$

com $d_k = x_k(\lambda_k) - x_k$, onde λ_k é o passo espectral. Nessa busca, ao rejeitar o primeiro ponto teste, os próximos são calculados ao longo da mesma direção viável, assim como em (3). Isso resulta em uma única direção d_k e uma única projeção por iteração. Na primeira estratégia (método SPG1), consideramos em (1) o parâmetro $\gamma_k = 1$ e $\tau_k = \alpha_k$ dado por (4), para $k \in \mathbb{N}$. Já na segunda estratégia (método SPG2), consideramos em (1) o parâmetro $\gamma_k = \alpha_k$ dado por (5) e $\tau_k = \lambda_k$ sendo o passo espectral. O Algoritmo 2 apresenta formalmente os métodos SPG1 e SPG2.

Algoritmo 2: MÉTODO DO GRADIENTE ESPECTRAL PROJETADO

- 1 Sejam $x_0 \in \Omega$, $\varepsilon > 0$, $\eta \in (0, 1)$, $1 > \sigma_2 > \sigma_1 > 0$, $M \in \mathbb{N}$, $\lambda^+ > \lambda^- > 0$, $\lambda_0 \in [\lambda^-, \lambda^+]$ e $k = 0$.
- 2 Se $\|x_k(1) - x_k\|_{\infty} < \varepsilon$, então pare e declare que x_k é um ponto estacionário.
- 3 Comece com $\alpha_k = \lambda_k$, calcule α_k por (4) e faça $x_{k+1} = x_k(\alpha_k)$ para o SPG1; ou comece com $\alpha_k = 1$, calcule α_k por (5) e faça $x_{k+1} = x_k + \alpha_k(x_k(\lambda_k) - x_k)$ para o SPG2.
- 4 Se (4) não for válida ou (5) não for válida, tome $\alpha_{\text{new}} \in [\sigma_1 \alpha_k, \sigma_2 \alpha_k]$, faça $\alpha_k = \alpha_{\text{new}}$ e retorne para o passo 3.
- 5 Calcule $s_k = x_{k+1} - x_k$ e $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.
- 6 Se $s_k^{\top} y_k \leq 0$, então faça $\lambda_{k+1} = \lambda^+$. Caso contrário, faça $\lambda_{k+1} = \max\{\lambda^-, \min\{s_k^{\top} s_k / s_k^{\top} y_k, \lambda^+\}\}$. Faça $k = k + 1$ e retorne para o passo 2.

Vale ressaltar que, quando o comprimento de passo α_k não satisfaz (4) ou (5), o ajuste para obtenção do novo α_k , no passo 4, é realizado por meio de interpolação quadrática.

Resultados

Nesta seção, vamos analisar o desempenho computacional dos métodos PG1, PG2, SPG1 e SPG2 relativo ao tempo de CPU, número de iterações e avaliação de função e de projeção. Os experimentos numéricos foram desenvolvidos a partir de 12 problemas de programação quadrática, da coleção CUTEst, definidos em conjuntos com restrições de limitação das variáveis. Estes correspondem à minimização de funções oriundas da modelagem de problemas de torção elasto-plástica, veja [5]. Consideramos $\Omega = \{x \in \mathbb{R}^n; x_j \in [-100, 50], j = 1, \dots, n\}$ com $n = 16, 100, 484, 1024$, totalizando 48 problemas. Foi utilizada a linguagem de programação Julia e os códigos estão livremente disponibilizados em <https://github.com/petimatematica/gradprojanalise>. Nos algoritmos 1 e 2, tomamos $\varepsilon = 10^{-5}$ e $\eta = 10^{-4}$. Além disso, no Algoritmo 1, assumimos $\bar{\alpha} = 1.0$, $\theta = 0.5$, $\hat{\beta} = 0.01$, $\beta = 0.9$ e $\beta_0 = 0.9$. No Algoritmo 2, assumimos $\lambda^- = 10^{-30}$, $\lambda^+ = 10^{30}$, $\sigma_1 = 0.1$, $\sigma_2 = 0.9$, $M = 10$ e o comprimento de passo inicial $1/\|x_0(1) - x_0\|_{\infty}$. Os algoritmos são interrompidos quando o número de iterações atinge 40000. No caso do Algoritmo 1, quando o comprimento do passo atinge um valor menor que 10^{-5} ele também é interrompido.

Conforme as figuras 1 e 2, os métodos SPG1 e SPG2 apresentaram melhor desempenho relativo ao tempo de CPU e ao número de iterações. Isso pode ser atribuído à melhoria que o passo espectral representa em relação a métodos tradicionais de cálculo do comprimento de passo.

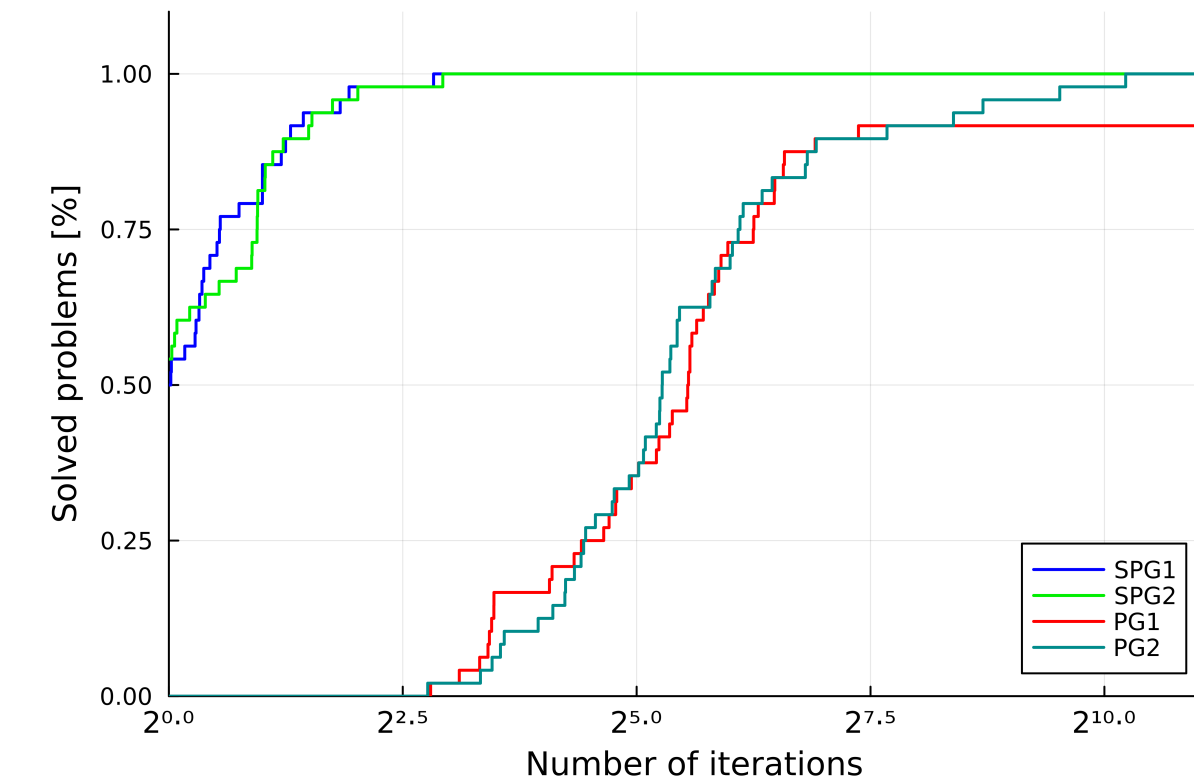


Figura 1: Performance profile relativo ao número de iterações.

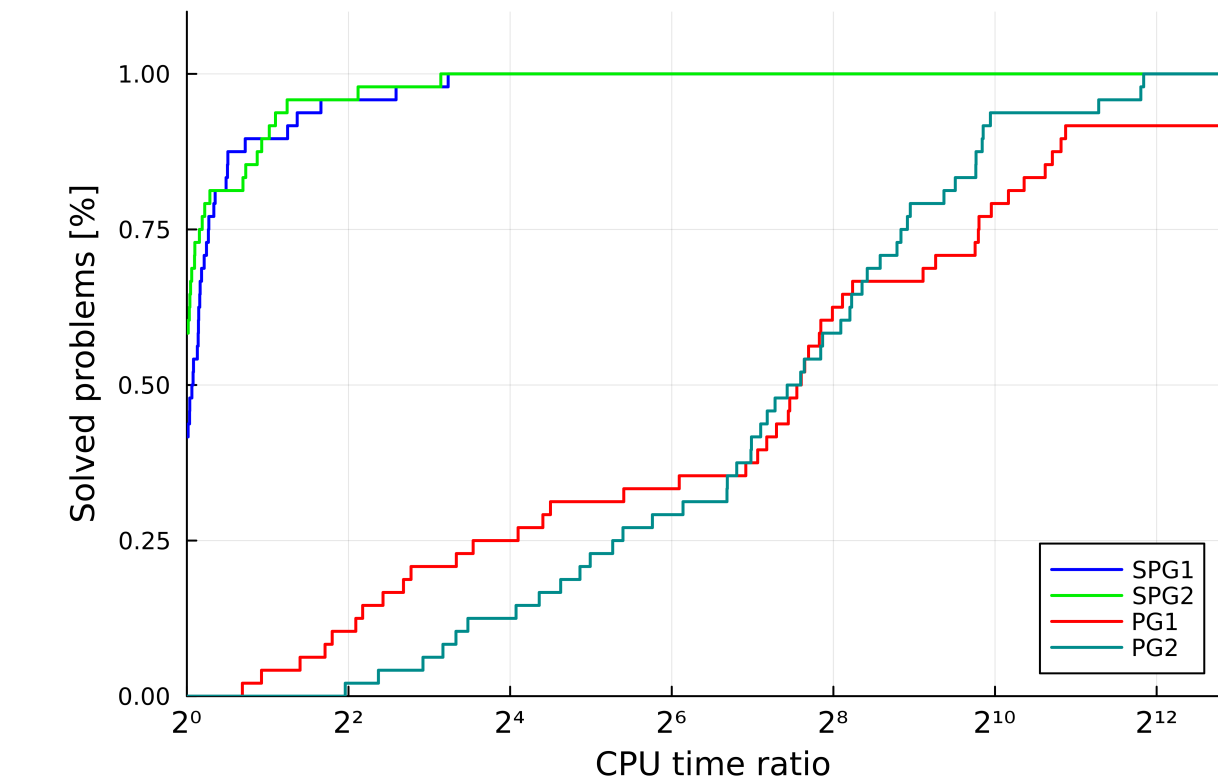


Figura 2: Performance profile relativo ao tempo de CPU.

Nas figuras 3 e 4, nota-se que o método SPG2 apresenta um melhor desempenho com relação ao número de avaliações de projeção. Além disso, o método PG2 apresenta melhor desempenho em relação ao PG1 quanto à avaliação de projeções, o que evidencia a eficiência das buscas ao longo das direções viáveis para os problemas estudados.

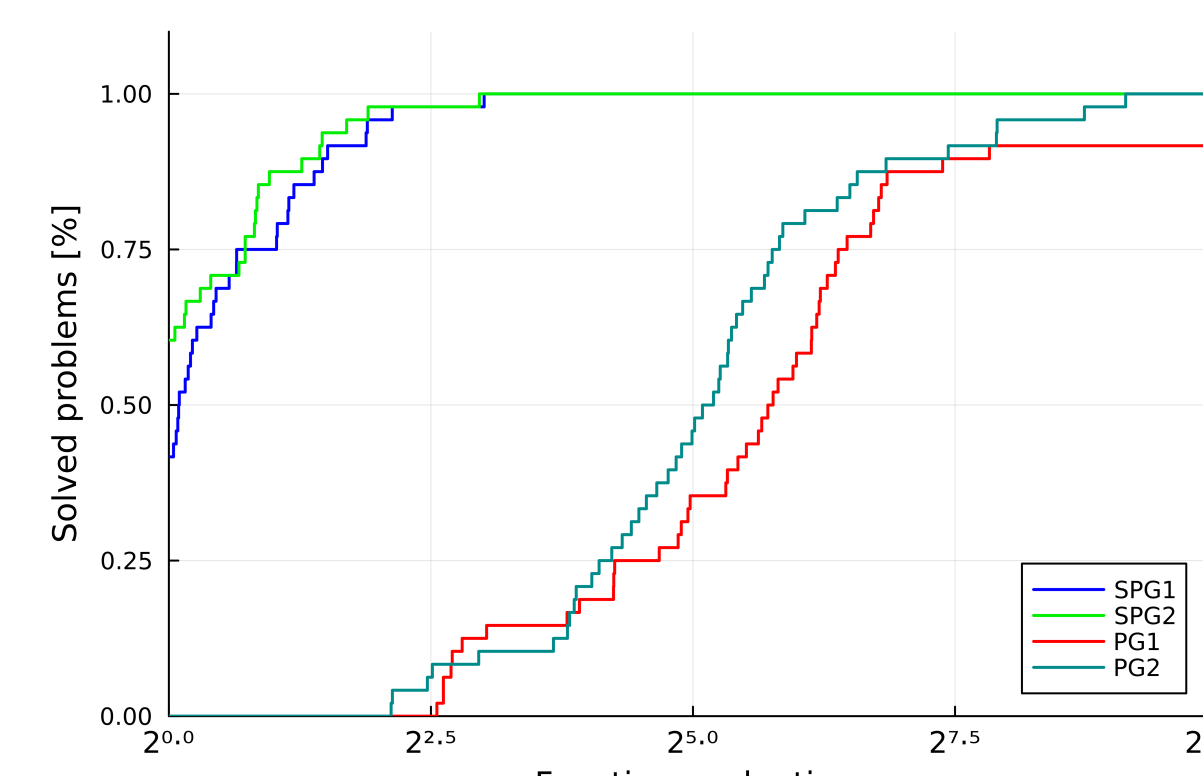


Figura 3: Performance profile relativo à avaliação de função.

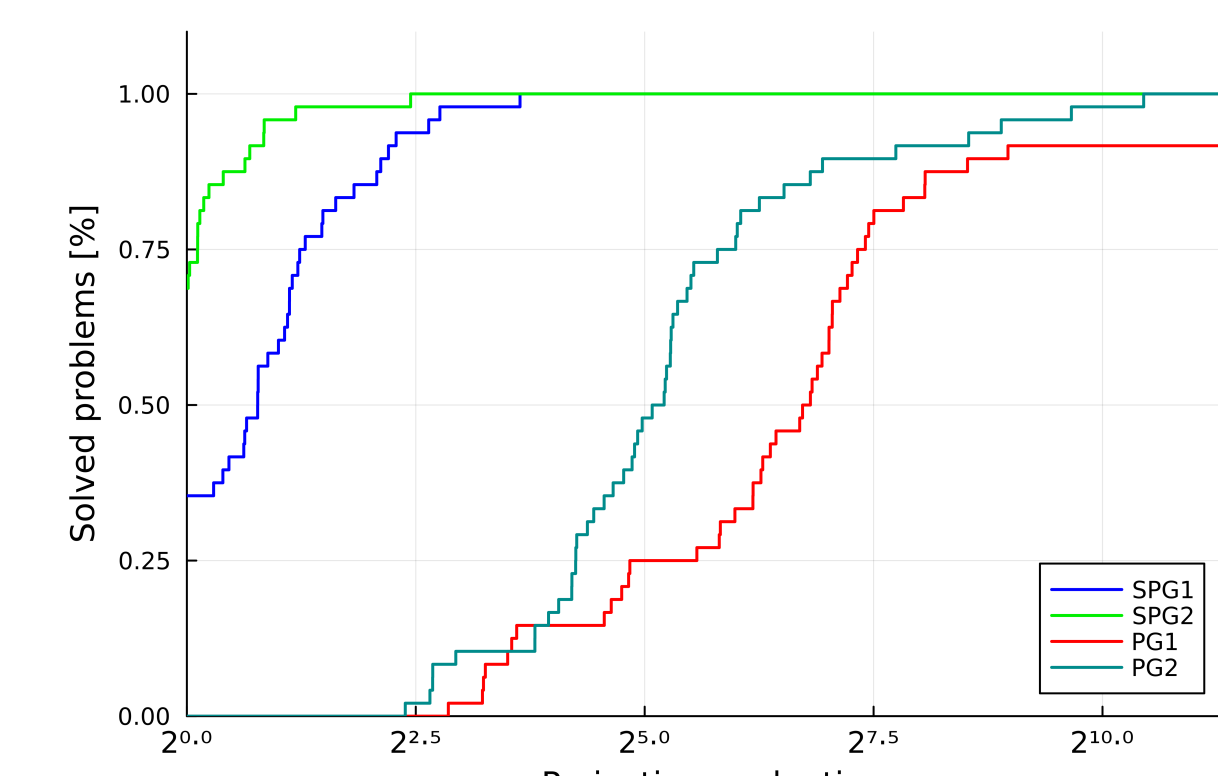


Figura 4: Performance profile relativo à avaliação de projeção.

Os perfis de desempenho evidenciaram também que quatro problemas com dimensão 1024, não foram resolvidos pelo PG1. Nestes casos, embora o algoritmo tenha atingido o número máximo de iterações, pôde-se observar que houve um decréscimo significativo na função objetivo.

Referências

- [1] Barzilai, J. & Borwein, J. M. (1988), "Two point step size gradient methods", *IMA Journal of Numerical Analysis*, 8, 141-148.
- [2] Bertsekas, D. P. (1997), "Nonlinear programming", *Journal of the Operational Research Society*, 48.3, 334-334.
- [3] Birgin, E. G., Martínez, J. M. & Raydan, M. (2000), "Nonmonotone spectral projected gradient methods on convex sets", *SIAM Journal on Optimization*, 10.4, 1196-1211.
- [4] Iusem, A. N. (2003), "On the convergence properties of the projected gradient method for convex optimization", *Computational & Applied Mathematics*, 22, 37-52.
- [5] Glowinski, R. (2013), "Numerical methods for nonlinear variational problems", Springer Science & Business Media.
- [6] Grippo, L., Lampariello, F. & Lucidi, S. (1986) "A nonmonotone line search technique for Newton's method", *SIAM journal on Numerical Analysis*, 23.4, 707-716.

Agradecimentos

Gostaríamos de agradecer ao Programa de Educação Tutorial Institucional da Universidade Estadual do Sudoeste da Bahia (PETI/UESB) pelas bolsas de estudo.