

# Gradiente Projetado em Conjuntos Proximalmente Suaves

Christian H. B. de Andrade\*      Douglas S. Gonçalves

Departamento de Matemática, CFM, UFSC,

88040-900, Florianópolis, SC

E-mail: christian.barlera@grad.ufsc.br,    douglas.goncalves@ufsc.br

## RESUMO

Problemas de otimização restrita da forma minimizar  $f(x)$  sujeito a  $x \in \Omega \subset \mathbb{R}^n$  com  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , aparecem em diversas aplicações reais [2, Seção 1.1]. Entre os métodos clássicos para tratar esse tipo de problema, temos o **método do gradiente projetado clássico**, que assume que o conjunto viável  $\Omega$  é convexo, fechado e não vazio. Em linhas gerais, a ideia é realizar a iteração padrão de descida de gradiente e então projetar o ponto (possivelmente fora de  $\Omega$ ) de volta ao conjunto viável, ou seja,  $x_{k+1} = P_\Omega(x_k - \eta \nabla f(x_k))$  para algum  $\eta > 0$ , em que  $P_\Omega(z)$  denota a projeção ortogonal de  $z \in \mathbb{R}^n$  sobre o conjunto  $\Omega$ .

A proposta neste trabalho é explorar a viabilidade do método de gradiente projetado para conjuntos **não convexos**, buscando conhecer as diferenças e limitações em relação ao cenário convexo e como trazer propriedades suficientes para obter resultados de convergência. Em particular, estudamos a estrutura de conjunto *R-proximalmente suave* sobre o conjunto viável  $\Omega$ , tratada em [1, 3].

Seja  $\emptyset \neq A \subset \mathbb{R}^n$ . Definimos o **tubo de “raio”**  $r > 0$  ao redor de  $A$  como o conjunto  $U_A(r) := \{y \in \mathbb{R}^n : \text{dist}(y, A) \leq r\}$ . Um subconjunto  $B \subset \mathbb{R}^n$  fechado é dito **R-proximalmente suave** quando existe um escalar  $R > 0$  tal que a projeção é única para todo  $x \in U_B(R)$ . Dado  $\emptyset \neq Q \subset \mathbb{R}^n$  conjunto *R-proximalmente suave* e fechado, para  $x \in Q$ , o **cone normal proximal de Q em x** é definido como  $N_Q^P(x) := \{v \in \mathbb{R}^n \mid \exists \delta > 0, x = P_Q(x + \delta v)\}$ .

Esses conceitos são particularmente interessantes pois permitem que a iteração de gradiente projetado esteja bem definida mesmo no caso de  $\Omega$  um conjunto não convexo (mas proximalmente-suave). Com isso, é possível obter o seguinte resultado de convergência.

**Teorema 1** ([1]). *Dado  $\emptyset \neq Q \subset \mathbb{R}^n$  compacto e R-proximalmente suave, se  $f : Q \rightarrow \mathbb{R}$  for função Lipschitz de constante  $L > 0$  (em  $U_Q(R)$ ),  $\nabla f$  Lipschitz de constante  $L_1 > 0$  e  $\eta \in \mathbb{R}$  tal que  $0 < \eta \leq \min\left\{\frac{1}{L_1}, \frac{R}{L}\right\}$ , então o método de gradiente projetado produz uma sequência  $\{x_k\} \subset Q$  tal que  $\lim_{k \rightarrow \infty} \text{dist}(x_k, \Phi) = 0$ , sendo  $\Phi = \{y \in \Omega : -\nabla f(y) \in N_\Omega^P(y)\}$  (o conjunto de pontos estacionários).*

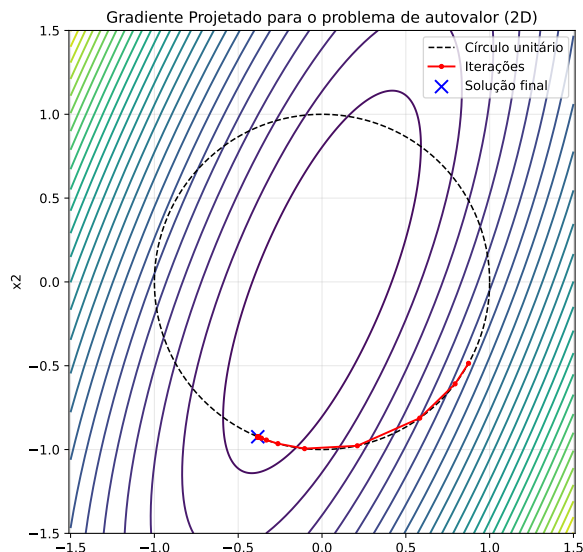
Como experimentos numéricos preliminares, executamos o método para o problema de encontrar o menor autovalor de uma matriz  $A \in \mathbb{R}^{n \times n}$  simétrica e positiva definida, isto é, o seguinte problema de otimização:  $\min f(x) = x^\top Ax$  sujeito a  $\|x\|_2 = 1$ . Observe que o conjunto  $Q = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$  é fechado, limitado, não convexo mas proximalmente suave de raio  $0 < R < 1$  pois o tubo  $U_Q(R)$  não contém a origem (que seria o único ponto em que a unicidade da projeção sobre  $Q$  falharia). Veja também que essa função objetivo satisfaz as hipóteses de diferenciabilidade e Lipschitz-continuidade. Portanto podemos aplicar o método de gradiente projetado neste problema, com a escolha de tamanho de passo  $\eta$  descrita no Teorema 1.

---

\*bolsista de Iniciação Científica PIBIC/CNPq

A Figura 1 ilustra as iterações do método de gradiente projetado neste tipo de problema em duas dimensões.

Figura 1: Teste para  $n = 2$  e número de condição  $\kappa = 10$ , com as curvas de nível mais interiores representando valores menores da função  $f$  e o trajeto em vermelho saindo do ponto inicial  $x_0$  e passando por cada iteração  $x_k$  até a solução final. O círculo unitário representa o conjunto viável  $Q$ .



Os resultados observados indicam que nesses experimentos o número de condição influencia no tamanho de  $\eta$  (por exemplo mudando da ordem de  $10^{-3}$  para  $10^{-4}$ , ao passar de  $\kappa = 100$  para  $\kappa = 1000$  em 5 dimensões). Entretanto existem casos em que  $\eta$  menor não implica em mais iterações e tampouco maior tempo de execução. Essas características parecem ser mais afetadas pela escolha do ponto inicial  $x_0$ .

Dimensão $n$	Erro relativo	Tempo (s)	Núm. de iterações	$\eta$
2	7.90e-15	6.69e-04	3.55e+01	1.10e-02
5	1.03e-14	5.57e-03	3.07e+02	6.56e-03
10	7.80e-15	1.93e-02	1.05e+03	4.30e-03
50	2.07e-15	3.36e-01	1.73e+04	1.40e-03
100	3.06e-15	9.94e-01	5.18e+04	9.22e-04

Tabela 1: Médias dos experimentos numéricos

**Palavras-chave:** *Otimização, gradiente, projeção, conjunto não convexo, proximalmente suave*

## Referências

- [1] M.V. Balashov, B.T. Polyak e A.A. Tremba, Gradient Projection and Conditional Gradient Methods for Constrained Nonconvex Minimization, *Numerical Functional Analysis and Optimization*, 41 (2020), 822–849.
- [2] S. Boyd e L. Vandenberghe, “Convex Optimization”, Cambridge University Press, Cambridge, 2004.
- [3] F.H. Clarke, R.J. Stern e P.R. Wolenski, Proximal Smoothness and the Lower- $C^2$  Property, *Journal of Convex Analysis*, 2 (1995), 117–144.