Universidade Federal do Espírito Santo
Programa Institucional de Iniciação Científica
Subprojeto de Iniciação Científica

O método do gradiente espectral projetado e variantes para minimização com restrições convexas

Resumo

Um dos procedimentos mais fundamentais para resolução de problemas de otimização com restrições convexas é o método do gradiente projetado. Este método baseia-se na ideia de que, dada uma função continuamente diferenciável, a direção de maior decréscimo é a contrária ao de seu gradiente. Apesar de seu uso em implementações práticas, sabe-se que a minimização iterativa através de passos de gradiente pode tornar-se demasiadamente lenta à medida que o método se aproxima da solução. Estratégias de aceleração incluem o controle do tamanho de passo. Em particular, o método do gradiente espectral projetado emprega um cálculo simples e barato do tamanho do passo que usa informações de segunda ordem advindas da equação secante, em que também se baseiam métodos quasi-Newton de grande sucesso. O método do gradiente espectral projetado, tal como formalizado por Raydan em 1997 e Birgin, Martínez e Raydan em 2000, usa ainda a técnica de interpolação quadrática para aceleração do cálculo do passo em conjunto com uma busca linear inexata não monótona. Esse método mostrou-se muito eficiente para minimização de funções gerais, inclusive de grande porte, com eficácia muito acima do método do gradiente tradicional, o que despertou a atenção da comunidade científica. Desde então, várias variantes foram propostas na literatura, sendo objeto de pesquisa até hoje. Nesta pesquisa, consideramos o método do gradiente espectral projetado e suas principais variantes. O objetivo é compará-las, do ponto de vista teórico e numérico, entre si.

Palavras-chave: Programação Não Linear. Gradiente projetado. Gradiente espectral e variantes.

1 Introdução

Nesta pesquisa, consideramos problemas de otimização com restrições convexas da forma

$$min_{x \in X} f(x) \ , \tag{1}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ é função continuamente diferenciável e $X \subset \mathbb{R}^n$ um conjunto convexo e fechado. Aqui estão contemplados problemas irrestritos, caso em que $X = \mathbb{R}^n$. Problemas da forma (1) são de grande importância, não só porque são utilizados diretamente em aplicações como *aprendizado de máquina supervisionado* (BOTTOU; CURTIS; NOCEDAL, 2018), mas também porque aparecem como subproblemas de métodos iterativos para minimização com restrições não lineares gerais, como Lagrangiano aumentado. De qualquer modo, problemas na forma (1) são largamente utilizados, e métodos para sua resolução estão presentes em vários pacotes computacionais modernos.

Dentre as ideias para resolução de (1), uma das mais fundamentais é caminhar iterativamente em direções que decrescem f localmente a partir de um ponto x (*direções de descida*). Sabemos dos cursos de cálculo que a direção de maior decréscimo para f a partir de x é $-\nabla f(x)$. Isso nos leva ao esquema iterativo do *método do gradiente projetado*, dado por

$$x^{k+1} = x^k + t_k d^k \ , \tag{2}$$

onde $t_k > 0$ é o *tamanho do passo* na direção de descida, e viável,

$$d^k = P_X(x^k - \nabla f(x)) - x^k, \tag{3}$$

e P_X é o operador projeção sobre X (no caso, a projeção existe e é única). Evidentemente, o tamanho do passo deve ser calculado a cada iteração, o que, na prática numérica, é guiado por uma condição que garanta "decréscimo suficiente" de f, usualmente a *condição de Armijo* (veja (RIBEIRO; KARAS, 2013) para detalhes). Tal condição é suficiente para convergência global teórica do método, isto é, é possível mostrar que a partir de qualquer ponto inicial x^0, os pontos de acumulação $\bar{x}$ gerados pelo método do gradiente satisfazem $P_X(\bar{x} - \nabla f(\bar{x})) = \bar{x}$ ($\bar{x}$ é ponto crítico/estacionário). Isso se reflete, por sua vez, no critério de parada prático $\|d^k\| \leq \varepsilon$.

A despeito de sua simplicidade, a ideia de descida pelo gradiente projetado é empregada em pacotes computacionais modernos. No entanto, sabe-se que a convergência pode ser lenta próximo a um ponto crítico (fenômeno "zig-zag", veja (RIBEIRO; KARAS, 2013)). Ou seja, são necessários muitos passos próximo à solução, o que leva ao aumento do número de avaliações da função f e seu de gradiente. Uma forma de contornar esse problema é tomar outra direção que não a do gradiente, pelos menos próximo à pontos críticos. Nesse sentido, métodos que fazem uso de derivadas segundas, por exemplo, tipo quasi-Newton, são por vezes empregados. Costumam ser eficazes, possuem boas taxas de convergência, porém são computacionalmente mais caros. Por outro lado, alternativas com baixo custo computacional funcionam muito bem em problemas específicos, como métodos de gradientes conjugados aplicados à problemas quadráticos estritamente convexos no caso irrestrito; porém as extensões para minimização de funções não lineares gerais são por vezes sofisticadas e não possuem tão bons resultados quanto no caso quadrático.

Uma alternativa computacionalmente barata, e logo adequada à problemas de grande porte, e que apresenta bom desempenho numérico para minimização de funções gerais é o *método do gradiente espectral projetado* (RAYDAN, 1997), (BIRGIN; MARTÍNEZ; RAYDAN, 2000). Nele, a direção (3) é usada, mas o passo $t_k = 1/\alpha_k > 0$ é computado de forma que a matriz múltiplo da identidade $\alpha_k I$ aproxime a equação secante $B(x^k - x^{k-1}) = \nabla f(x^k) - \nabla f(x^{k-1})$ (ou, resumidamente, $B s_k = y_k$), no sentido de α_k de ser solução do problema

$$min_\alpha \|\alpha s_k - y_k\|_2^2 .$$

A solução pode ser computada por fórmula fechada, a saber, $\alpha_k = (s_k^T y_k)/(s_k^T s_k)$, o que torna o método facilmente implementável e barato. A equação secante é inspirada na aproximação linear de $\nabla f(x)$ ao redor de x^k; nesse sentido, B faz o papel de $\nabla^2 f(x)^{-1}$. Há outra versão em que α_k provém da equação secante obtida da anterior multiplicada por B^{-1}, o que gera outra versão de gradiente espectral. Na literatura, a primeira versão é conhecida como BB1 e a segunda BB2 ("BB" é um acrônimo dos nomes Barzilai e Borwein, que foram os idealizadores da ideia para quadráticas com 2 variáveis (BARZILAI; BORWEIN, 1988)).

Dado seu bom desempenho numérico, variantes do método do gradiente espectral projetado surgiram na literatura. Um exemplo é o algoritmo de Frassoldati, Zanni e Zanghirati (2008), conhecido como ABBmin. Neste método, assim como em seu antecessor ABB (do inglês, *Adaptive Barzilai-Borwein*) (ZHOU; GAO; DAI, 2006), os passos BB1 e BB2 são combinados de maneira heurística, porém muito bem fundamentada, de modo que o método resultante tenha melhor convergência quando próximo à solução. Mais tarde, a filosofia do famoso método dos gradientes conjugados para quadráticas foi agregada (DAI; KOU, 2016), (LIU; LIU, 2018). Cabe destacar que variantes do gradiente espectral projetado continuam sendo tema de trabalhos publicados em periódicos relevantes da área; veja por exemplo (HUANG; DAI; LIU; ZHANG, 2022).

Nesta pesquisa, consideramos o método do gradiente espectral e suas variantes ABB (ZHOU; GAO; DAI, 2006), ABBmin (FRASSOLDATI; ZANNI; ZANGHIRATI, 2008), o método de Dai e Kou (2016), dentre outras. **O objetivo é compará-los entre si do ponto de vista teórico e numérico.** Ressaltamos que este **não** é um tópico abordado em disciplinas da graduação em matemática.

Referências

BARZILAI, J.; BORWEIN, J. M. Two-point step size gradient methods. **IMA Journal of Numerical Analysis**, v. 8, p. 141-148, 1988.

BIRGIN, E. G.; MARTÍNEZ, J. M.; RAYDAN, M. Nonmonotone Spectral Projected Gradient Methods on Convex Sets. **SIAM Journal on Optimization**, v. 10, p. 1196-1211, 2000.

BOTTOU, L.; CURTIS, F. E.; NOCEDAL, J. Optimization Methods for Large-Scale Machine Learning. **SIAM Review**, v. 60, n. 2, p. 223-311, 2018.

DAI, Y.; KOU, C. A Barzilai-Borwein conjugate gradient method. **Science China Mathematics**, v. 59, p. 1511-1524, 2016.

FRASSOLDATI, G.; ZANNI, L.; ZANGHIRATI, G. New adaptive stepsize selections in gradient methods. **J. Ind. Manag. Optim.**, v. 4, p. 299-312, 2008.

HUANG, Y.; DAI, Y.-H.; LIU, X.-W.; ZHANG, H. On the acceleration of the Barzilai–Borwein method. **Computational Optimization and Applications**, v. 81, p. 717-740, 2022.

LIU, H.; LIU, Z. An Efficient Barzilai–Borwein Conjugate Gradient Method for Unconstrained Optimization. **Journal of Optimization Theory and Applications**, v. 180, p. 879-906, 2018.

NOCEDAL, J.; WRIGHT, S. J. **Numerical optimization**. Springer, 2006.

RAYDAN, M. The Barzilai and Borwein Gradient Method for the Large Scale Unconstrained Minimization Problem. SIAM Journal on Optimization, v. 7, p. 26-33, 1997.

RIBEIRO, A. A.; KARAS, E. W. **Otimização Contínua**. São Paulo: Cengage, 2013.

ZHOU, B.; GAO, L.; DAI, Y.-H. Gradient Methods with Adaptive Step-Sizes. **Computational Optimization and Applications**, v. 35, p. 69-86, 2006.