XXX Congresso de Iniciação Científica UNICAMP

Estudo sobre uma família de métodos com as propriedades teóricas de gradientes conjugados para funções quadráticas

Palavras-chave: Programação matemática. Gradientes conjugados. Métodos de primeira ordem.

Orientador: Prof. Dr. Roberto Andreani – IMECC | UNICAMP

Aluno: Filipe Lacerda Benevides 173494 – UNICAMP

Campinas – 2022

Introdução

O CELEBRADO método do gradiente conjugado tem quase a mesma complexidade computacional que o método do gradiente simples, porém produz os melhores resultados quando se trata de minimizar funções quadráticas em que a Hessiana é definida positiva. Esse método é utilizado na resolução de subproblemas que surgem quando aplicados os métodos de regiões de confiança, fazendo parte das subrotinas dos mais diversos pacotes de otimização. Seu resultado depende fortemente da exatidão da busca linear em cada passo, se isso é feito com certo grau de erro, inevitável em problemas mal condicionados, o método perde boa parte de sua eficácia.

Andreani *et al* [1] apresentaram recentemente uma família de métodos que possuem as mesmas propriedades teóricas de gradientes conjugados e de implementação tão eficiente quanto. Bem como o gradiente conjugado minimiza a função objetivo sobre o espaços de Krilov ou variedades lineares de busca, os membros dessa família minimizam uma combinação convexa da função objetivo e sua norma ao quadrado.

Objetivos

O principal objetivo desse projeto é o estudo de uma família de métodos com as propriedades teóricas de gradientes conjugados para funções quadráticas. A ideia por trás desse estudo é analisar se essa família preserva melhor essas propriedades quando extendida às funções estritamente convexas, dado que o gradiente conjugado clássico tende a perder eficâcia com problemas que não possuem alta precisão na busca linear.

De maneira concisa, os objetivos do projeto estão delineados a seguir:

- Estudar algoritmos de otimização de primeira ordem, em especial o método do gradiente conjugado;
- Entender teoricamente a razão desses métodos funcionarem;
- Implementar computacionalmente e fazer estudos numéricos desses métodos;
- [OBJETIVO PRINCIPAL] Estudar do ponto de vista teórico e prático uma família de métodos com a melhor propriedade para funções quadráticas.

Resultados e Discussão

Durante o estudo dos métodos de descida, foi formulado um novo método para computar o passo de descida que se mostrou mais eficiente do que o método clássico, mesmo quando o problema é mal condicionado.

Seja a função $f(x) = \frac{1}{2}x^t Ax + b^t x + c$ com matriz hessiana da forma:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & q \end{bmatrix}$$

em que q é um número aleatório gerado uniformemente no intervalo $[10^{-8}, 10^2]$. Esse intervalo faz com que algumas matrizes hessianas sejam extremamente mal condicionadas, principalmente quando $q \ll 1$ ou $1 \ll q$. Foram realizados 5000 gerações randômicas e as funções foram minimizadas utilizando tanto o método clássico quanto o novo método proposto. As iterações foram computadas até que a norma euclidiana do gradiente fosse inferior a 10^{-8}.

Em quase 90% das gerações o método proposto precisou de menos da metade da quantidade de iterações que o método clássico precisou, ou seja, foi certamente mais eficiente do ponto de vista computacional. Em menos de 0.5% das gerações o método proposto precisou de mais iterações do que o método clássico. O método proposto precisou, em média, de 4.5 vezes menos iterações do que o clássico.

Repetindo a mesma análise para matrizes quadradas de dimensão randômica entre 2 e 10 (incluindo os extremos), temos que em quase 80% das gerações o método proposto precisou de menos da metade da quantidade de iterações que o método clássico precisou. Em cerca de 6% das gerações o método proposto precisou de mais iterações do que o método clássico. O método proposto precisou, em média, de 5 vezes menos iterações do que o método clássico.

A mesma comparação foi feita com o método proposto por Barzilai e Borwein [2], contudo o método proposto só performou melhor com matrizes de ordem pequena. Quando a dimensão da hessiana era próxima de 100, o método proposto convergia com até três vezes mais iterações do que o método de BB.

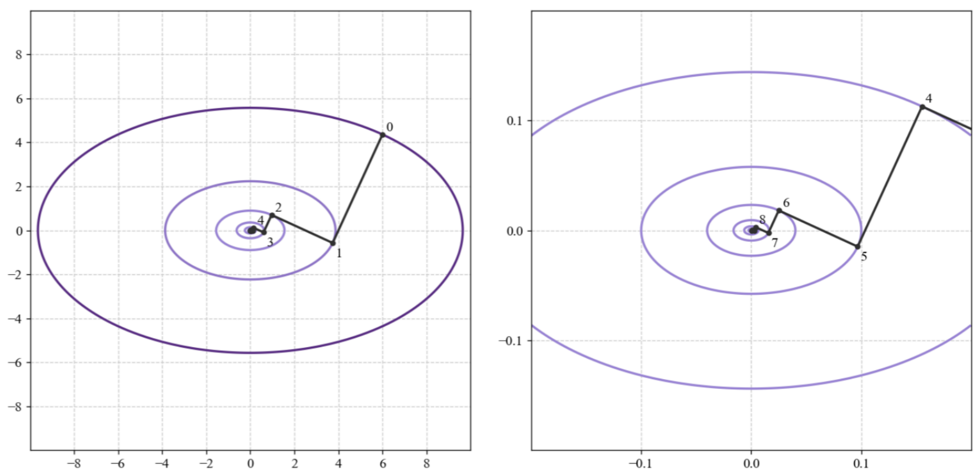

Figura 01. (a) quatro primeiras iterações do método clássico (b) *zoom* na região central para observar outras iterações. Esse método precisou de 11 iterações para que a norma euclidiana do gradiente fosse inferior a 10^{-3}.

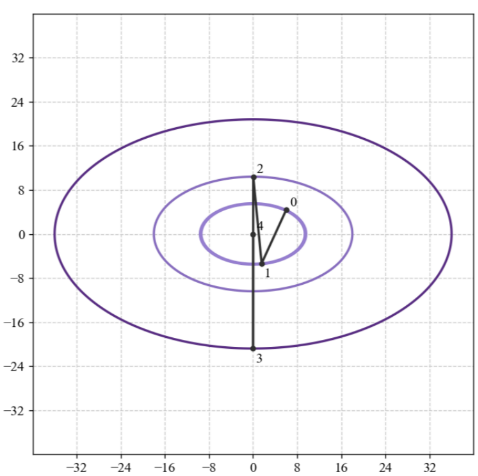

Figura 02. mesmas condições inicias da análise anterior. O método proposto precisou de 4 iterações para que a norma do gradiente fosse inferior a 10^{-3} (critério de parada).

Conclusões

Por mais que o novo método proposto tenha duas vezes o custo computacional do método clássico em cada iteração, ainda assim ele se mostrou mais eficiente do que aquele já consagrado com o passo de máxima descida para quadráticas. Vale ressaltar que o método proposto destaca-se do clássico principalmente quando a matriz hessiana é mal condicionada e x_0 favorece sua convergência lenta. Entretanto, esse novo método ainda é mais **ineficiente** do que alguns extremamente baratos do ponto de vista computacional, como o método de Barzilai e Borwein.

Referências

[1] R. Andreani, H. Oviedo and M. Raydan. A family of optimal weighted conjugate-gradient-type methods for strictly convex quadratic minimization, to appear in Numerical Algorithm. Optimization Online, 2021. ↩

[2] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. IMA Journal of Numerical Analysis, 8(1):141–148, 1988. ↩

[3] M. Bazaraa, H.D. Sherali and C.M. Shetty. Nonlinear Programming − Theory and Algorithms. John Wiley and Sons, New Jersey, 2006.

[4] Friedlander, Ana. Elementos de Programação Não Linear, 1994.

[5] David G. Luenberger and Yinyu Ye. Linear and Nonlinear Programming, Springer, 2008.

[6] J. Nocedal, S.J. Wright, Numerical Optimization, 2$^{\text{nd}}$ edition. Springer, New York, 2006.

[7] J.R. Shewchuk. An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, 1994.
https://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf