Fundamentos de otimização irrestrita

Marina Andretta

ICMC-USP

4 de agosto de 2014

Baseado no livro Numerical Optimization, de J. Nocedal e S. J. Wright.

Problema irrestrito

Em minimização irrestrita, queremos minimizar uma função de variáveis reais a um valor real sem que haja restrições aos valores das variáveis.

Ou seja, estamos interessados em resolver o seguinte problema:

$$\boxed{\text{Minimizar} \quad f(x)} \tag{1}$$

onde

- $x \in \boldsymbol{R}^n$;
- $f \in \boldsymbol{R}^n \to \boldsymbol{R}$ uma função suave.

Problema irrestrito

Geralmente não dispomos de uma visão geral da função f. Tudo o que sabemos são os valores de f e, talvez, os valores de suas derivadas em um conjunto de pontos x_0, x_1, x_2, ...

Felizmente, os algoritmos podem escolher estes pontos e eles tentam escolhê-los de modo a identificar a solução de maneira segura e sem usar muito tempo ou espaço computacional.

Muitas vezes, informações sobre f não são obtidas de maneira "barata". Por isso, preferem-se algoritmos que não pedem esta informação desnecessariamente.

Exemplo

Suponha que queremos encontrar a curva que se encaixa em alguns dados experimentais.

Pelos dados e pelo conhecimento do problema, deduzimos que o sinal tem comportamento exponencial e oscilatório de alguns tipos e decidimos modelá-lo usando a função

$$\boxed{\phi(t; x) = x_1 + x_2 e^{-(x_3 - t)^2/x_4} + x_5 \cos(x_6 t).}$$

Exemplo

Os números reais x_i, $i = 1, ..., 6$ são os parâmetros do modelo. Gostaríamos de escolhê-los para fazer com que os valores de $\phi(t_j; x)$ estejam o mais próximo possível dos valores y_j observados.

Para escrever isto como um problema de otimização, agrupamos os parâmetros x_i em um vetor de variáveis $x = (x_1, x_2, ..., x_6)$ e definimos os resíduos

$$\boxed{r_j(x) = y_j - \phi(t_j; x), \quad j = 1, ..., m,}$$

que medem a discrepância entre o modelo e o valor observado.

Exemplo

Nossas estimativas dos parâmetros $x \in \boldsymbol{R}^6$ serão dados pela solução do problema

$$\boxed{\text{Minimizar} \quad f(x) = r_1^2(x) + \ldots + r_m^2(x).} \qquad (2)$$

Este é um problema de quadrados mínimos não-linear, um caso particular de otimização irrestrita.

Exemplo

O problema (2) ilustra como o cálculo de algumas funções objetivos pode ser custoso, mesmo quando o número de variáveis é pequeno.

Neste caso, o número de variáveis é $n = 6$, mas, se o número de observações m for grande, o cálculo do valor de $f(x)$ pode ser computacionalmente alto.

Exemplo

Suponha que, para um certo conjunto de dados, a solução ótima de (2) seja aproximadamente $x^* = (1.1, 0.01, 1.2, 1.5, 2.0, 1.5)$ e a função objetivo valha $f(x^*) = 0.34$.

Como o valor da função objetivo não é nulo, deve haver discrepâncias entre os valores observados y_j e os valores estimados por $\phi(t_j; x^*)$ para algum j (geralmente, vários). Ou seja, o modelo não pode reproduzir exatamente todos os pontos dados.

Como podemos verificar que x^* é, de fato, um minimizador de f?

O que é uma solução

Idealmente, ficaríamos felizes se pudéssemos encontrar o minimizador global de f, ou seja, o ponto no qual o valor de f é o menor possível.

Formalmente, temos que

Um ponto x^* é um minimizador global se $f(x^*) \leq f(x)$ para todo x,

com x variando em todo $\boldsymbol{R}^n$ (ou pelo menos no domínio de interesse do modelador).

O que é uma solução

O minimizador global pode ser difícil de ser encontrado, pois temos apenas conhecimento local de f.

Como esperamos que o algoritmo para encontrar a solução de (1) não visite muitos pontos, não temos uma visão geral do formato de f e não podemos ter certeza de que a função não decresce muito em algum ponto não visitado pelo algoritmo.

O que é uma solução

A maioria dos algoritmos são capazes somente de encontrar minimizadores locais, que são pontos que atingem o menor valor de f em relação a uma vizinhança.

Formalmente, dizemos que

> Um ponto x^* é um minimizador local se há uma vizinhança $\mathcal{N}$ de x^* tal que $f(x^*) \leq f(x)$ para todo $x \in \mathcal{N}$,

lembrando que uma vizinhança de x^* é um conjunto aberto que contém x^*.

Um ponto que satisfaz esta definição é comumente chamado de minimizador local fraco. Esta terminologia serve para distinguir este ponto de um minimizador local estrito (também chamado de minimizador local forte), que satisfaz a seguinte propriedade:

> Um ponto x^* é um minimizador local estrito se há uma vizinhança $\mathcal{N}$ de x^* tal que $f(x^*) < f(x)$ para todo $x \in \mathcal{N}$, com $x \neq x^*$.

Por exemplo, para a função constante $f(x) = 2$, todo ponto é um minimizador local fraco. Já a função $f(x) = (x-2)^4$ possui um minimizador local estrito em $x = 2$.

Uma outra definição de minimizador local é a seguinte:

> Um ponto x^* é um minimizador local isolado se há uma vizinhança $\mathcal{N}$ de x^* tal que x^* seja o único minimizador local em $\mathcal{N}$.

Note que alguns minimizadores locais estritos não são isolados. Tome a função

$$\boxed{f(x) = x^4 \cos(1/x) + 2x^4, \quad f(0) = 0.}$$

Ela tem segunda derivada contínua e possui um minimizador local estrito em $x^* = 0$. No entanto, há minimizadores locais estritos em muitos pontos x_n próximos e podemos rotular estes pontos de forma que $x_n \to 0$ quando $n \to \infty$.

Apesar de não ser verdade que todo minimizador local estrito é isolado, vale que todo minimizador local isolado é estrito.

Reconhecimento de mínimo local

Pelas definições dadas anteriormente, tem-se a impressão de que a única maneira de determinar se um ponto x^* é um minimizador local é examinar todos os pontos em uma vizinhança próxima para verificar se nenhum deles possui valor de função menor.

No entanto, quando uma função é suave, há maneiras mais práticas e eficientes para fazer identificar um minimizador local.

Em particular, se f possui segunda derivada contínua, é possível dizer se um ponto x^* é um minimizador local (possivelmente um minimizador local estrito) examinando apenas o gradiente $\nabla f(x^*)$ e a Hessiana $\nabla^2 f(x^*)$.

Reconhecimento de mínimo local

A ferramenta matemática usada para estudar minimizadores de funções suaves é o Teorema de Taylor.

Teorema de Taylor: *Suponha que* $f : \boldsymbol{R}^n \to \boldsymbol{R}$ *possua derivada contínua e que* $p \in \boldsymbol{R}^n$*. Então temos que*

$$f(x+p) = f(x) + \nabla f(x+tp)^T p, \quad t \in (0,1).$$

Mais ainda, se f *possui segunda derivada contínua, temos que*

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)p \, dt,$$

e

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp)^T p, \quad t \in (0,1).$$

Reconhecimento de mínimo local

Condições necessárias para otimalidade são feitas supondo que x^* seja um minimizador local e, então, concluindo fatos sobre $\nabla f(x^*)$ e $\nabla^2 f(x^*)$.

Teorema 1 (Condições necessárias de primeira ordem):
Se x^* é um minimizador local e f possui primeira derivada contínua em uma vizinhança aberta de x^*, então $\nabla f(x^*) = 0$.

Chamamos x^* de ponto estacionário se $\nabla f(x^*) = 0$. Ou seja, pelo Teorema 1, todo minimizador local deve ser um ponto estacionário.

Reconhecimento de mínimo local

Teorema 2 (Condições necessárias de segunda ordem):
Se x^* é um minimizador local de f e $\nabla^2 f$ é contínua em uma vizinhança aberta de x^*, então $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ é semi-definida positiva.

Reconhecimento de mínimo local

Agora descrevemos condições suficientes para que um ponto seja minimizador local de f. Ou seja, condições que, se satisfeitas por um ponto x^*, garantem que este ponto é um minimizador local de f.

Teorema 3 (Condições suficientes de segunda ordem): *Suponha que $\nabla^2 f$ seja contínua em uma vizinhança aberta de x^*, que $\nabla f(x^*) = 0$ e que $\nabla^2 f(x^*)$ seja definida positiva. Então x^* é um minimizador estrito de f.*

Reconhecimento de mínimo local

Note que as condições suficientes de segunda ordem do Teorema 3 garantem algo mais forte do que as condições necessárias apresentadas anteriormente, já que garantem que o ponto será um minimizador local estrito.

Note ainda que as condições suficientes do Teorema 3 não são necessárias: um ponto x^* pode ser um minimizador local estrito e não satisfazer as condições suficientes.

Um exemplo é a função $f(x) = x^4$, para a qual o ponto $x^* = 0$ é um minimizador local estrito no qual a Hessiana se anula.

Reconhecimento de mínimo local

Quando a função objetivo é convexa, minimizadores locais e globais são simples de caracterizar.

> **Teorema 4:** *Quando f é convexa, todo minimizador x^* é um minimizador global de f. Se f é diferenciável, então todo ponto estacionário x^* é um minimizador global de f.*

Estes resultados, baseados em cálculo elementar, são as bases para os algoritmos de otimização irrestrita. De alguma maneira, todo algoritmo busca um ponto no qual o gradiente ∇f se anula.