

Algoritmos para resolução de problemas de minimização irrestrita

Marina Andretta

ICMC-USP

4 de agosto de 2014

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

Algoritmos para minimização irrestrita

Veremos agora um resumo dos algoritmos existentes mais comuns para resolução de problemas de minimização irrestrita, ou seja,

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

onde

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

Algoritmos para minimização irrestrita

Todos os algoritmos para resolução de problemas de minimização irrestrita precisam de um **ponto inicial** x_0 .

O **usuário com conhecimento** da aplicação e do conjunto de dados referente ao problema a ser resolvido pode fornecer uma **boa aproximação** x_0 para a solução do problema. Quando uma **boa aproximação não é conhecida**, um **ponto arbitrário** pode ser escolhido como x_0 .

Iniciando em x_0 , algoritmos de otimização geram uma **sequência de iterandos** $\{x_k\}_{k=0}^{\infty}$ que termina quando **não é mais possível obter progresso** ou quando **uma solução é encontrada** com precisão suficiente.

Algoritmos para minimização irrestrita

Para decidir como mover de um ponto x_k para o próximo, **algoritmos usam a informação de f em x_k** e, possivelmente, informação sobre os iterandos anteriores x_0, x_1, \dots, x_{k-1} . Eles usam esta informação para **calcular o novo iterando x_{k+1} com valor de função menor do que x_k** .

Existem algoritmos não-monótonos, que não exigem que o valor da função decresça a cada iteração. Mas mesmo estes algoritmos exigem que o valor da função decresça depois de um número fixo m de iterações.

Há duas estratégias fundamentais para se mover do ponto x_k para o próximo iterando x_{k+1} : **busca linear** e **regiões de confiança**.

Estratégia de busca linear

Na estratégia de **busca linear**, o algoritmo calcula uma **direção p_k** e **busca ao longo desta direção**, a partir de x_k , um ponto x_{k+1} com valor de função menor do que x_k .

A distância a se mover ao longo de p_k pode ser encontrada resolvendo-se aproximadamente o problema unidimensional

$$\text{Minimizar}_{\alpha>0} f(x_k + \alpha p_k). \quad (2)$$

Resolvendo este problema exatamente, obteríamos o ponto com **melhor valor de função na direção p_k** (a partir de x_k).

Estratégia de busca linear

No entanto, resolver (2) de maneira exata é custoso e desnecessário.

Em vez disso, o algoritmo de busca linear gera um número limitado de tentativas de tamanho de passo até que encontre um α_k que aproxima a solução do problema (2). O novo iterando passa a ser

$$x_{k+1} = x_k + \alpha_k p_k.$$

A partir do novo ponto calculado x_{k+1} , uma nova direção p_{k+1} é calculada, uma nova busca pelo tamanho de passo é feita e um novo iterando é calculado. Este processo é repetido até que uma solução para o problema original (1) seja encontrada.

Estratégia de regiões de confiança

Na estratégia de **regiões de confiança**, a **informação obtida sobre a função f é usada para construir uma função modelo m_k** , cujo comportamento perto do ponto x_k é parecido com o comportamento de f .

Como o comportamento de m_k pode não ser similar ao comportamento de f em pontos distantes de x_k , **restringimos a busca por um minimizador de m_k a uma região em torno de x_k** . Em outras palavras, calculamos o candidato a passo p resolvendo aproximadamente o problema

$$\text{Minimizar}_p \quad m_k(x_k + p), \quad (3)$$

com $x_k + p$ pertencente à região de confiança.

Estratégia de regiões de confiança

Calculado o passo p , verifica-se se o ponto $x_k + p$ produz decréscimo suficiente da função f . Em caso negativo, entende-se que a região de confiança está muito grande, então ela é diminuída e calcula-se novamente uma solução aproximada para (3).

Se o decréscimo produzido na função objetivo f for suficiente, o próximo iterando passa a ser

$$x_{k+1} = x_k + p.$$

O processo é repetido até que uma solução para o problema original (1) seja encontrada.

Estratégia de regiões de confiança

Geralmente, a **região de confiança** é a bola dada por $\|p\| \leq \Delta$, com $\Delta > 0$. Δ é chamado de **raio da região de confiança**. Regiões elípticas ou com formato de caixa também podem ser usadas.

O modelo m_k geralmente é definido como uma função quadrática da forma

$$m_k(x_k + p) = f_k + p^T \nabla f_k + \frac{1}{2} p^T B_k p,$$

com $f_k = f(x_k)$, $\nabla f_k = \nabla f(x_k)$ e B_k uma aproximação (ou a própria) Hessiana de f em x_k .

Diferenças entre as duas estratégias

Note que, quando o tamanho da região de confiança é mudado, a direção a ser calculada muda, diferentemente do que acontece com a estratégia de busca linear - na qual, uma vez calculada a direção, apenas o tamanho de passo é modificado.

Basicamente, as estratégias de busca linear e regiões de confiança diferem na ordem em que calculam a direção e a distância de x_k para x_{k+1} .

A busca linear começa fixando uma direção e, então, calcula a distância apropriada que x_k deve ter de x_{k+1} (ou seja, o tamanho de passo α_k).

Em regiões de confiança, primeiro determina-se a maior distância permitida entre x_k e x_{k+1} (ou seja, o raio da região de confiança Δ_k). Então, calcula-se o passo que será dado, a partir de x_k , para calcular x_{k+1} .

Ordem de convergência

Uma das maneiras de medir a performance de um algoritmo é calcular sua **ordem de convergência**.

Seja $\{x_k\}$ a sequência em \mathbf{R}^n que converge a x^* . Dizemos que a convergência é **Q-linear** se existe uma constante $r \in (0, 1)$ tal que

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq r,$$

para todo k suficientemente grande.

Ordem de convergência

Isto significa que a distância à solução x^* diminui a cada iteração de, pelo menos, um fator constante.

Por exemplo, a sequência $1 + (0.5)^k$ converge Q-linearmente a 1.

O prefixo Q vem de “quociente”, já que o tipo de convergência é definido em termos dos quocientes dos erros sucessivos.

Ordem de convergência

A convergência é dita **Q-superlinear** se

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0.$$

Por exemplo, a sequência $1 + k^{-k}$ converge Q-superlinearmente para 1.

A convergência é dita **Q-quadrática** se

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} \leq M,$$

para todo k suficientemente grande. M é uma constante positiva, não necessariamente menor do que 1.

Por exemplo, a sequência $1 + (0.5)^{2^k}$ converge Q-quadraticamente para 1.

Ordem de convergência

A velocidade de convergência depende de r e, mais fracamente, de M . Estes valores dependem não só do algoritmo, mas das propriedades do problema particular. No entanto, independentemente destes valores, uma sequência Q-quadraticamente convergente sempre convergirá, a partir de um momento, mais rápido do que uma sequência Q-linearmente convergente.

Obviamente, se uma sequência converge Q-quadraticamente, ela também converge Q-superlinearmente. E se uma sequência converge Q-superlinearmente, ela também converge Q-linearmente.

Ordem de convergência

Estas são as ordens de convergência mais usadas na prática. Mais geralmente, dizemos que a **Q-ordem de convergência de uma sequência é p** (para $p > 1$) se existe uma constante positiva M tal que

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p} \leq M,$$

para todo k suficientemente grande.