Universidade Federal de Minas Gerais

TRABALHO PRÁTICO 2
OTIMIZAÇÃO EM ENGENHARIA

OTIMIZAÇÃO COM ALGORITMOS QUASE-NEWTON E APROXIMAÇÕES QUADRÁTICAS

Sumário

Capítulo 1

Introdução

Em problemas de programação não linear, um dos primeiros algoritmos de otimização desenvolvido, a fim de obter o menor valor possível da função objetivo (no caso dos problemas de minimização), é o algoritmo Gradiente. Esse método é baseado no cálculo do gradiente da função objetivo que corresponde ao vetor que define, localmente, a direção e o sentido de maior crescimento da função. Dessa forma, no intuito de encontrar uma solução que minimize a função objetivo, basta caminhar no sentido contrário ao do vetor gradiente. Por meio desse processo iterativo uma nova solução é localizada e essa rotina se repete até que algum critério de convergência seja satisfeito (CHUNG, et al., 2001).

Após o estabelecimento do algoritmo Gradiente, muitas pesquisas foram desenvolvidas no intuito de obter mais informações da função objetivo. Até então, os algoritmos levavam em consideração apenas a informação do gradiente da função objetivo no ponto corrente. No entanto, essas pesquisas possibilitaram a introdução das informações sobre a curvatura da função, dando origem ao método de Newton e aos métodos Quase-Newton. No primeiro a Hessiana da função objetivo é determinada por meio de aproximações quadráticas baseadas nos termos da série de Taylor. Já no segundo são realizadas aproximações de segunda ordem por meio de metodologias que determinam uma estimativa da Hessiana. A aplicação desses métodos permitem uma significativa aceleração de convergência nos processos de otimização (HULL, 2002).

Muitos estudos computacionais recentes têm mostrado que os métodos de otimização de direção de busca, tais como os métodos Quase-Newton, são extremamente eficientes na resolução de problemas de otimização monobjetivo irrestritos. Uma classe desses problemas são os relacionados à programação quadrática irrestrita que podem ser descritos matematicamente como segue:

$$x^* = \arg\min_x f(x)\,, \qquad x \in \mathbb{R} \tag{1}$$

A programação quadrática é considerada uma classe de problemas de otimização que possui como função objetivo uma função quadrática. Muitos problemas de diversas áreas de conhecimento, tais como, indústria, medicina e economia, são modelados por meio dessas funções. Assim surgiu a necessidade de desenvolver métodos de otimização robustos para solucionar tais problemas de forma eficiente e eficaz (HULL, 2002).

O principal objetivo dos métodos de otimização é encontrar os valores das variáveis de decisão x que minimizem ou maximizem a função objetivo $f(x)$ do problema descrito em (1). Tais valores é a solução ótima do problema de otimização designada x^*.

Os algoritmos de direção de busca que solucionam o problema apresentado em (1) iniciam-se com uma solução inicial $x_0 \neq x^*$ e em seus processos iterativos determinam uma sequência de soluções x_k que tendem para a solução ótima do problema x^*. Abaixo é apresentado um pseudocódigo que implementa de maneira generalizada a ideia proposta pelos método de direção de busca (CHUNG, et al., 2001).

Pseudocódigo geral dos métodos de direção de busca

$k \leftarrow 0$
Enquanto (não critério de parada)
  | $d_k \leftarrow h\big(x_1, \dots, x_k, f(x_1), \dots, f(x_2)\big)$
  | $\alpha_k \leftarrow \arg\min_{\alpha} f(x_k + \alpha \cdot d_k)$
  | $x_{k+1} \leftarrow x_k + \alpha_k \cdot d_k$
  | $k = k + 1$
Fim

No pseudocódigo apresentado acima $h(\cdot)$ é uma função recursiva, d_k é a direção de busca e α_k é o passo de busca unidimensional para cada iteração k. Os métodos de otimização são classificados de acordo com a forma que é determinada a direção de busca. Sendo assim, as próximas seções são destinadas ao detalhamento de tais métodos.

1.1. Método do Gradiente

O algoritmo do Gradiente, como já abordado nesse trabalho, define a direção de busca por meio do cálculo do gradiente da função objetivo no ponto corrente. Como o gradiente define a direção do crescimento da função em um determinado ponto. No caso de problemas padrão de minimização, basta caminhar no sentido oposto determinado pelo gradiente. Dessa forma, tem-se que:

$$d_k = -\nabla f(x_k) \tag{2}$$

Vale ressaltar que o único requisito para que esse método seja aplicado é que a função objetivo seja diferenciável. Além disso, em muitas situações práticas, onde não se conhece a função objetivo (caixa preta) ou o cálculo analítico da derivada de primeira ordem da função objetivo é extremamente complexo, o gradiente é determinado numericamente.

1.2. Método de Newton

No algoritmo de Newton, além da obtenção do gradiente da função objetivo, calcula-se a hessiana da função de modo a obter uma convergência mais rápida. Assim, tem-se que:

$$d_k = -H_k^{-1} \ \nabla f(x_k) \tag{3}$$

Sendo H_k a Hessiana da função objetivo no ponto x_k. A definição apresentada na equação (3) baseia-se na aproximação quadrática da função objetivo, a partir dos termos da série de Taylor definida abaixo:

$$f(x) \;=\; f(x_0) \;+\; \nabla f(x_0)' \cdot (x - x_0) \;+\; \frac{1}{2} \cdot (x - x_0)' \cdot H(x_0) \cdot (x - x_0) \;+ \cdot \mathcal{O}(3) \tag{4}$$

Na equação acima, os termos de ordem igual ou superior a três $\mathcal{O}(3)$ são desprezados. Vale ressaltar que a Hessiana H_k também pode ser determinada numericamente.

Sabe-se que o método de Newton na sua forma iterativa aplicado a problemas de otimização irrestritos, inicia-se com uma solução inicial e determina uma próxima solução x_{i+1}, baseada na solução da iteração anterior x_i. Esse processo se repete até o algoritmo encontrar uma solução próxima da solução ótima x^*.

Uma desvantagem desse método é a necessidade de realizar um número considerável de avaliações da função objetivo que depende da dimensão do problema e é calculado por meio da seguinte expressão $(n + 1)^2$. Dessa forma, a medida que se aumenta a quantidade de variáveis de decisão do problema de otimização, aumenta-se de forma quadrática o número de avaliações da função objetivo. Isso implica em um esforço computacional significativo. Além disso, o cálculo da inversa da Hessiana pode acarretar em erros numéricos, o que traz dificuldades na aplicação desse método na resolução de diversos problemas. Assim como no método do Gradiente, o método de Newton não se aplica em problemas cuja função objetivo não é diferenciável e a matriz Hessiana não é semi-definida positiva.

1.3. Métodos de Quase-Newton

Diversos métodos da família Newton que solucionam problemas de otimização, requerem a determinação da matriz Hessiana que equivale a derivada de segunda ordem da função a ser minimizada. Os métodos Quase-Newton são utilizados em problemas de otimização caracterizados por possuírem uma função objetivo cuja Hessiana é custosa ou difícil de ser calculada. Estes métodos utilizam a informação somente da derivada de primeira ordem para obter uma aproximação da Hessiana sob um número finito de iterações. Esta aproximação é atualizada em cada iteração por uma matriz de rank baixo. Em problemas de otimização irrestritos, a direção de busca do método Quase-Newton é dada por:

$$B_{k+1} = B_k + C_k(\alpha) \rightarrow d_k = -B_{k+1} \quad \nabla f(x_{k+1}) \tag{5}$$

Ao observar a equação (5) verifica-se que a avaliação da inversa da Hessiana da função objetivo é substituída pela construção de uma estimativa recursiva dada por B_k, a qual é chamada durante todo o processo de otimização. Esta estimativa é essencial principalmente na otimização de funções não quadráticas cuja Hessiana não é constante, permitindo uma adaptação contínua da estimativa no local em que a mesma está sendo avaliada. Nesse caso, é crucial que a matriz H_k permaneça sempre no mínimo semi-definida positiva, e preferencialmente bem condicionada, ou seja, com autovalores não muito distanciados entre si (GROENWOLD, et al., 2008).

Diante disso, nas próximas subseções são pontuados e detalhados os métodos que definem a forma de determinar as estimativas recursivas da matriz de aproximação

B_k. Nesse caso são estudados os métodos BFGS e DFP, além das adaptações propostas por Huang e Biggs (GUO, et al., 2007).

1.3.1. Método BFGS

O método BFGS (Broyden-Fletcher-Goldfarb-Shanno) determina uma aproximação para a matriz Hessiana por meio de um processo iterativo. Os passos que definem a rotina do algoritmo desse método são apresentados abaixo:

Passo 1. Inicialização do índice das iterações em $k = 0$. Dessa forma a solução é inicializada por meio de um ponto inicial x_0 e uma matriz $n \times n$ definida positiva B_0 é estabelecida como estimativa inicial da inversa da Hessiana da função objetivo $f(x)$. Nos casos em que não se possui informações adicionais relacionadas ao problema a ser otimizado, a matriz B_0 é igual a matriz identidade. Além disso, o vetor gradiente da primeira iteração é dado por $\nabla f(x_0)$.

Passo 2. Determinação do gradiente da função $\nabla f(x_k)$ no ponto x_k e determinação da direção d_k como segue abaixo:

$$d_k = -[B_k] \cdot \nabla f(x_k) \tag{6}$$

Passo 3. Determinação do passo ótimo λ_k^* na direção d_k e, assim, calcula-se:

$$x_{k+1} = x_k + \lambda_k^* \cdot d_k \tag{7}$$

Passo 4. Verificação da otimalidade do ponto x_{k+1}, por meio de uma precisão ε preestabelecida, em que se a condição $\|\nabla f(x_{k+1})\| \leq \varepsilon$ for verdadeira, a solução x_{k+1} equivale a solução do problema x^* e o processo iterativo é finalizado. Caso a condição seja falsa, o processo avança para o **Passo 5**.

Passo 5. Atualização da matriz Hessiana por meio de:

$$[B_{k+1}] = [B_k] + \left(1 + \frac{g_k^T \cdot [B_k] \cdot g_k}{s_k^T \cdot g_k}\right) \cdot \frac{s_k \cdot s_k^T}{s_k^T \cdot g_k} - \frac{s_k^T \cdot g_k^T \cdot [B_k]}{s_k^T \cdot g_k} - \frac{[B_k] \cdot g_k \cdot s_k^T}{s_k^T \cdot g_k} \tag{8}$$

Sendo que:

$$s_k = x_{k+1} - x_k = \lambda_k^* \cdot d_k$$

$$g_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

Passo 6. Atualização da iteração em $k = k + 1$ e reinicialização a partir do **Passo 2**.

Vale ressaltar que se o passo λ_k^* a cada iteração do algoritmo BFGS equivale a um valor com boa precisão na direção da solução ótima do problema de programação quadrática, a matriz de aproximação da Hessiana B_k permanecerá definida positiva nas próximas iterações. Todavia, em muitas situações práticas, essa afirmação nem sempre é verdadeira, o que resulta em uma matriz B_k indefinida. Nesse caso, uma saída é atualizar essa matriz pela matriz identidade de forma periódica durante o processo de otimização a fim de evitar erros indesejados.

1.3.2. Método DFP

O método DFP (Davidon-Fletcher-Powell) busca encontrar uma aproximação da inversa da matriz Hessiana. E esse método pode ser definido por meio da seguinte rotina de execução:

Passo 1. Definição de um ponto inicial x_0 e de uma matriz $n \times n$ definida positiva B_0 como uma estimativa inicial da inversa da Hessiana da função objetivo $f(x)$. Nos casos em que não se possui informações adicionais relacionadas ao problema a ser otimizado, a matriz B_0 é igual a matriz identidade. Além disso, o vetor gradiente da primeira iteração é dado por $\nabla f(x_0)$.

Passo 2. Determinação do gradiente da função $\nabla f(x_k)$ da solução corrente x_k e determinação da direção d_k como segue abaixo:

$$d_k = -[B_k] \cdot \nabla f(x_k) \tag{9}$$

Passo 3. Determinação do passo ótimo λ_k^* na direção d_k e, assim, calcula-se:

$$x_{k+1} = x_k + \lambda_k^* \cdot d_k \tag{10}$$

Passo 4. Verificação da otimalidade do ponto x_{k+1}, por meio de uma precisão ε preestabelecida, em que se a condição $\|\nabla f(x_{k+1})\| \leq \varepsilon$ for verdadeira, a solução x_{k+1} equivale a solução do problema x^* e o processo iterativo é finalizado. Caso a condição seja falsa, o processo avança para o **Passo 5**.

Passo 5. Atualização da inversa da matriz Hessiana por meio de:

$$[B_{k+1}] = [B_k] + [M_k] + [N_k] \tag{11}$$

Sendo que:

$$[M_k] = \lambda_k^* \cdot \frac{d_k \cdot d_k^T}{d_k^T \cdot g_k}$$

$$[N_k] = -\frac{([B_k] \cdot g_k) \cdot ([B_k] \cdot g_k)^T}{g_k^T \cdot [B_k] \cdot g_k}$$

$$g_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

Passo 6. Atualização da iteração em $k = k + 1$ e reinicialização a partir do **Passo 2.**

Assim como no algoritmo BFGS, no algoritmo DFP a matriz B_k será definida positiva somente se o passo λ_k^* for determinado com precisão. Caso contrário, torna-se necessário acrescentar procedimentos que recalculem o valor de λ_k^* de forma que a matriz B_k se torne definida positiva. Vale ressaltar que esses procedimentos pode aumentar o esforço computacional do método, e, além disso, convém destacar que o método DFP é mais influenciado por erros no parâmetro λ_k^* do que o método BFGS.

1.3.3. Adaptação de Huang

Em 1969, o professor Huang do departamento de Engenharia Mecânica e Aeroespacial da Universidade de Rice no Texas propôs uma metodologia para a atualização da matriz Hessiana por meio de funções quadráticas. Essas funções foram definidas da seguinte forma:

$$f(x) = a + b^T \cdot x + \frac{1}{2} \cdot x^T \cdot c \cdot x \tag{12}$$

O gradiente da função quadrática $f(x)$ apresentada em (12) é dado por:

$$g(x) = \nabla f(x) = b + c \cdot x \tag{13}$$

De acordo com o método adaptado de Huang, a atualização da Hessiana é dada por:

$$\overline{H} = H + \rho \cdot \left(\frac{\Delta x \cdot (C_1 \cdot \Delta x + C_2 \cdot H^T \cdot \Delta g)^T}{(C_1 \cdot \Delta x + C_2 \cdot H^T \cdot \Delta g)^T \cdot \Delta g} \right) - \frac{H \cdot \Delta g \cdot (K_1 \cdot \Delta x + K_2 \cdot H^T \cdot \Delta g)^T}{(K_1 \cdot \Delta x + K_2 \cdot H^T \cdot \Delta g)^T \cdot \Delta g} \tag{14}$$

Onde $\overline{H}$ corresponde à matriz atualizada e ρ, C_1, C_2, K_1 e K_2 são constantes, sendo que apenas três delas são independentes. Além disso tem-se as seguintes relações:

$$\Delta x = \bar{x} - x$$

$$\Delta g = \bar{g} - g$$

Caso a matriz de atualização da Hessiana $\overline{H}$ escrita em (14) for forçada para resultar em matrizes simétricas, ou seja, $(\overline{H} - H)^T = \overline{H} - H$, tem-se que:

$$\frac{(\Delta x \cdot \Delta g^T - H \cdot \Delta g \cdot \Delta x^T)}{(\beta \cdot \Delta x + H \cdot \Delta g)^T \cdot \Delta g} = \frac{(H \cdot \Delta g \cdot \Delta x^T - \Delta x \cdot \Delta g^T)}{(\Delta x + \gamma \cdot H \cdot \Delta g)^T \cdot \Delta g} \tag{15}$$

Onde $\beta = C_1/C_2$ e $\gamma = K_1/K_2$.

A fim de simplificar as operações, elimina-se o parâmetro β da equação (15) por meio da seguinte manipulação algébrica:

$$\beta = -1 - (1 - \gamma) \cdot \left(\frac{\Delta g^T \cdot H \cdot \Delta g}{\Delta x^T \cdot \Delta g} \right) \tag{16}$$

A equação acima é substituída em (15), resultando em uma forma geral de atualização da Hessiana com base em apenas um único parâmetro de seleção, o γ. Assim, finalmente, tem-se que:

$$\overline{H} = \frac{\left[\overline{H}_{BFGS} + \gamma \cdot \left(\frac{\Delta g^T \cdot H \cdot \Delta g}{\Delta x^T \cdot \Delta g}\right) \cdot \overline{H}_{DFP}\right]}{\left[1 + \gamma \cdot \left(\frac{\Delta g^T \cdot H \cdot \Delta g}{\Delta x^T \cdot \Delta g}\right)\right]} \tag{17}$$

Ao analisar a equação acima, percebe-se que:

→ Se $\gamma = \infty \Rightarrow \overline{H} = \overline{H}_{DFP}$
→ Se $\gamma = 0 \Rightarrow \overline{H} = \overline{H}_{BFGS}$
→ Se γ assumir valores arbitrários, $\overline{H}$ será uma combinação linear de $\overline{H}_{DFP}$ e $\overline{H}_{BFGS}$.

1.3.4. Adaptação de Biggs

Conforme apresentado nesse trabalho, os método de Quase-Newton possuem como diferencial a determinação de uma aproximação da matriz Hessiana H, ou de sua inversa, de modo que tal matriz deve ser obrigatoriamente definida positiva. Dessa forma, deve garantir que:

$$H_{k+1} \cdot y_k = \alpha_k \cdot s_k, \qquad \alpha_k > 0 \tag{18}$$

Onde $s_k = x_{k+1} - x_k$ e $y_k = g_{k+1} - g_k$, sabendo que g é o gradiente da função objetivo $f(x)$.

A fim de melhorar o desempenho dos algoritmos Quase-Newton, Biggs propôs um novo método para determinar as atualizações da matriz hessiana por meio de um ajuste adequado do parâmetro α_k da equação acima. Baseado em modelos não quadráticos Biggs estabeleceu que o parâmetro α_k deve ser dado por:

$$\alpha_k = \frac{1}{t_k} \tag{19}$$

Sendo que:

$$t_k = \frac{6}{d_k^T \cdot y_k} \cdot \left(f(x_k) - f(x_{k+1}) + d_k^T \cdot g_{k+1}\right) - 2$$

Assim, a atualização da matriz Hessiana, utilizando essa adaptação para o método BFGS, por exemplo, ficaria na seguinte forma:

$$H_{k+1} = H_k + \frac{1}{s_k^T \cdot y_k} \cdot \left(\left(\alpha_k + \frac{y_k^T \cdot H_k \cdot y_k}{s_k^T \cdot y_k}\right) \cdot s_k \cdot s_k^T - s_k \cdot y_k^T \cdot H_k - H_k \cdot y_k \cdot s_k^T\right) \tag{20}$$

A adaptação proposta por Biggs faz com que o número de iterações e de avaliações da função objetivo/gradiente sejam menores do que os métodos BFGS e DFP tradicionais durante o processo de otimização.

Nesse trabalho são realizados diversos testes experimentais a fim de comparar o desempenho dos métodos Quase-Newton: o BFGS e o DFP originais e com as adaptações de Huang e Biggs. Em suma é avaliado o desempenho de cada método por meio do seu número de iterações, número de avaliações da função objetivo e erro percentual em relação ao ótimo de uma série de problemas de programação quadrática. O próximo capítulo é destinado a apresentação das funções quadráticas e funções não quadráticas perfeitas utilizadas nesse estudo.