Programação Quadrática Sequencial

M. Fernanda P. Costa

Departamento de Matemática
Universidade do Minho

Outline

1. Programação Quadrática Sequencial
   - Método SQP Local
     - Método SQP
     - Método SQP (restrições de desigualdade)
   - Pré-visualização de Métodos SQP práticos
   - Hessiana da Lagrangeana: aproximações Quasi-Newton
   - Funções Mérito
     - Função de penalidade ℓ_1
   - Um método SQP de procura unidirecional prático

Introdução:

- O Método da Programação Quadrática Sequencial (do inglês *Sequential Quadratic Programming* (SQP)) é um dos métodos mais eficazes para otimização não linear com restrições, de grande ou pequena dimensão.
- O método SQP é uma técnica sequencial, na qual a direção de procura é obtida resolvendo subproblemas de Programação Quadrática (PQ).
- A função objetivo do problema original é substituída por uma aproximação quadrática e as funções de restrição por funções lineares.

Método SQP Local

Considere, inicialmente, o problema do otimização não linear com restrições de igualdade:

$$\begin{array}{ll} \underset{w \in \mathbb{R}^{\mathrm{I}}}{\text{minimizar}} & F(w) \\ \text{sujeito a} & c_n(w) = 0, \quad n \in \mathcal{E} \end{array} \tag{P1}$$

- A ideia subjacente ao método SQP é modelar o problema (P1) no iterando atual $w^{(k)}$ por um subproblema de Programação Quadrática e, em seguida, usar o minimizante deste subproblema para definir um novo iterando $w^{(k+1)}$.
- O desafio é projetar o subproblema de PQ de modo a que este produza boas direções de procura para o problema de otimização não linear.
- Talvez a dedução mais simples dos métodos SQP, que vamos já de seguida apresentar, os veja como uma aplicação do método de Newton às condições de otimalidade KKT para (P1).

A função Lagrangeana para o problema (P1) é:

$$L(w,\lambda) = F(w) - \sum_{n\in\mathcal{E}} \lambda_n c_n(w) \;\Leftrightarrow\; L(w,\lambda) = F(w) - c(w)^T\lambda \qquad (1)$$

As condições KKT de 1ª ordem para problema (P1) são:

$$\begin{cases} \nabla_w L(w,\lambda) &= 0 \\ \nabla_\lambda L(w,\lambda) &= 0 \end{cases} \Leftrightarrow \begin{cases} \nabla F(w) - A(w)^T\lambda &= 0 \\ c(w) &= 0 \end{cases} \qquad (2)$$

- $c(w) = (c_1(w), \ldots, c_m(w))^T$ é o vetor das funções de restrição;
  $\lambda = (\lambda_1, \ldots, \lambda_m)^T$ é o vetor dos multiplicadores de Lagrange;
  $m = |\mathcal{E}|$ denota o número de restrições;
- $A(w)$ é a matriz do Jacobiano das restrições:

$$A(w)^T = [\nabla c_1(w), \ldots, \nabla c_m]$$

- o sistema (2) é não linear de $\mathrm{I} + m$ equações em $\mathrm{I} + m$ incógnitas w e λ.

- Qualquer solução (w^*, λ^*) do problema com restrições de igualdade (P1) para o qual $A(w^*)$ tem característica completa por linhas, satisfaz (2).
- Um método que se sugere é resolver o sistema não linear (2) pelo **método de Newton** (descrito no Capítulo 11 em [1]).

Para usar o método de Newton, precisamos de formar a matriz do Jacobiano de (2) que é dada por

$$\left[\begin{array}{cc} \frac{\partial(\nabla F(w) - A(w)^T\lambda)}{\partial w} & \frac{\partial(\nabla F(w) - A(w)^T\lambda)}{\partial \lambda} \\ \frac{\partial c(w)}{\partial w} & \frac{\partial c(w)}{\partial \lambda} \end{array}\right] = \left[\begin{array}{cc} \nabla^2_{ww} L(w,\lambda) & -A(w)^T \\ A(w) & 0 \end{array}\right] \tag{3}$$

onde $\nabla^2_{ww} L(w,\lambda) = \nabla^2 F(w) - \sum_{n\in\mathcal{E}} \lambda_n \nabla^2 c_n(w)$

Assim, o passo de Newton a partir do iterando $(w^{(k)}, \lambda^{(k)})$ é dado por:

$$\begin{aligned} w^{(k+1)} &= w^{(k)} + s^k \\ \lambda^{(k+1)} &= \lambda^{(k)} + s_\lambda^k \end{aligned} \tag{4}$$

onde as direções de procura s^k e s_λ^k são a solução do sistema Newton-KKT:

$$\begin{bmatrix} \nabla^2_{ww} L\left(w^{(k)}, \lambda^{(k)}\right) & -A(w^{(k)})^T \\ A(w^{(k)}) & 0 \end{bmatrix} \begin{bmatrix} s^k \\ s_\lambda^k \end{bmatrix} = - \begin{bmatrix} \nabla F\left(w^{(k)}\right) - A\left(w^{(k)}\right)^T \lambda^{(k)} \\ c\left(w^{(k)}\right) \end{bmatrix} \tag{5}$$

- A iteração de Newton (4)-(5) está bem definida se a matriz KKT em (5) é não-singular.
- A matriz KKT é não singular se as hipóteses seguintes, se verificam para $(w, \lambda) = \left(w^{(k)}, \lambda^{(k)}\right)$:

Hipóteses:

(H1) O Jacobiano das restrições $A(w)$ tem característica completa por linhas.

(H2) A matriz $\nabla^2_{ww} L(w, \lambda)$ é definida positiva no espaço tangente das restrições, isto é, $s^T \nabla^2_{ww} L(w, \lambda) s \geq 0$ para todo $s \neq 0$ tal que $A(w)s = 0$.

- Se estas hipóteses se verificam, é possível mostrar que iteração de Newton (4)-(5) converge quadraticamente e constitui um excelente algoritmo para resolver problemas com restrições de igualdade, desde que o ponto inicial $w^{(0)}$ esteja suficientemente próximo de w^*.

nota:

- A hipoótese H1 é a qualificação da restrição de independência linear (do inglês *linear independence constraint qualification* - LICQ), que assumimos.
- A hipótese H2 verifica-se sempre que (w, λ) está próximo do ótimo (w^*, λ^*), e a condição suficiente de 2ª ordem é satisfeita no ótimo.

Método SQP

Há uma forma alternativa de ver a iteração (4)-(5). Suponhamos que no iterando $(w^{(k)}, \lambda^{(k)})$ modelamos o problema (P1) usando o seguinte programa quadrático:

$$\begin{array}{rl} \underset{s \in \mathbb{R}^{\mathrm{I}}}{\text{minimizar}} & F_k + \nabla F_k^T s + \frac{1}{2} s^T \nabla^2_{ww} L_k s \\ \text{sujeito a} & A_k s + c_k = 0 \end{array} \tag{Q1}$$

onde

- F_k, ∇F_k, $\nabla^2_{ww} L_k$, A_k e c_k denotam respetivamente $F(w^{(k)})$, $\nabla F(w^{(k)})$, $\nabla^2_{ww} L\left(w^{(k)}, \lambda^{(k)}\right)$, $A\left(w^{(k)}\right)$ e $c\left(w^{(k)}\right)$.

A função Lagrangeana para o problema (Q1) é:

$$L(s, \pi) = F_k + \nabla F_k^T s + \frac{1}{2} s^T \nabla^2_{ww} L_k s - (A_k s + c_k)^T \pi$$

- $\pi = (\pi_1, \ldots, \pi_m)$ é o vetor dos multiplicadores de Lagrange associados às restrições em (Q1).

As condições KKT de 1ª ordem para o problema (Q1) são:

$$\left\{\begin{array}{ll}\nabla_s L(s,\pi) & = 0\\ \nabla_\pi L(s,\pi) & = 0\end{array}\right. \Leftrightarrow \left\{\begin{array}{rl}\nabla F_k + \nabla^2_{ww} L_k s - A_k^T \pi & = 0\\ A_k s + c_k & = 0\end{array}\right. \tag{6}$$

Se as hipóteses H1 e H2 se verificam, o problema (Q1) tem uma única solução (s^k, π^k) que satisfaz (6), ou seja,

$$\begin{bmatrix}\nabla^2_{ww} L_k & -A_k^T\\ A_k & 0\end{bmatrix}\begin{bmatrix}s^k\\ \pi^k\end{bmatrix} = -\begin{bmatrix}\nabla F_k\\ c_k\end{bmatrix} \tag{7}$$

Uma observação importante é que os vetores s^k e π^k podem ser identificados com a solução das equações de Newton (5). Vejamos: se subtrairmos $A_k^T \lambda^{(k)}$ a ambos os lados da primeira equação de (5), tem-se

$$\nabla^2_{ww} L_k s^k - A_k^T s_\lambda^k - A_k^T \lambda^{(k)} = -\nabla F_k + A_k^T \lambda^{(k)} - A_k^T \lambda^{(k)} \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \nabla^2_{ww} L_k s^k - A_k^T (\lambda^{(k)} + s_\lambda^k) = -\nabla F_k$$

e por (4) obtemos

$$\nabla^2_{ww} L_k s^k - A_k^T \lambda^{(k+1)} = -\nabla F_k$$

▷ Assim, pela não singulariedade da matriz dos coeficientes, temos que:

- $\lambda^{(k+1)} = \pi^k$ e
- s^k é solução de (Q1) e (5).

▷ **Portanto**, o novo iterando $\left(w^{(k+1)}, \lambda^{(k+1)}\right)$ pode ser definido como a solução do subproblema quadrático (Q1) ou como o iterando gerado pelo método de Newton (4)-(5) aplicado às condições de otimalidade do problema.

▷ Ambos os pontos de vista são úteis. Do ponto de vista de Newton facilita a análise de convergência, enquanto o método SQP permite deduzir algoritmos práticos e estender a técnica para problemas com restrições de desigualdade.

Apresentamos o método SQP na sua forma mais simples.

Algoritmo1: Algoritmo SQP Local - Restrições de Igualdade

- Dar: o par inicial $\left(w^{(0)}, \lambda^{(0)}\right)$
- **Para** $k = 0, 1, \ldots$

1. Calcular F_k, ∇F_k, $\nabla^2_{ww} L_k$, c_k, A_k
2. Resolver o sistema (7) para obter (s^k, π^k) (solução, multiplicadores de Lagrange)
3. Fazer $w^{(k+1)} = w^{(k)} + s^k$ e $\lambda^{(k+1)} = \pi^k$
4. **Se** $\left(w^{(k+1)}, \lambda^{(k+1)}\right)$ satisfaz o critério de paragem para o problema (P1)
   Parar com a solução $\left(w^{(k+1)}, \lambda^{(k+1)}\right)$
   fim se

(nota: resolver o subproblema quadrático (Q1) é equivalente a resolver o sistema (7))

Alguns comentários:

- No subproblema quadrático (Q1) poderíamos substituir o termo linear $\nabla F_k^T s$ por $\nabla_w L(w^{(k)}, \lambda^{(k)})^T s$, já que a restrição de (Q1) torna as duas escolhas equivalente. (ver nota)

  Neste caso, a função objetivo de (Q1) é uma aproximação quadrática da função Lagrangeana.

- Este facto motiva a escolha do modelo quadrático (Q1): primeiro substitui-se o problema (P1) pelo problema de minimizar a função Lagrangeana sujeita às restrições de igualdade de (P1), depois faz-se uma aproximação quadrática da Lagrangeana e uma aproximação linear das restrições para obter (Q1).

nota: considerando a expansão em série de Taylor até à 2ª ordem da função $L(w, \lambda^{(k)})$ para $w = w^{(k)} + s$ numa vizinhança de $w^{(k)}$ tem-se:

$$
\begin{aligned}
L\left(w^{(k)} + s, \lambda^{(k)}\right) &\approx L\left(w^{(k)}, \lambda^{(k)}\right) + \nabla_w L\left(w^{(k)}, \lambda^{(k)}\right)^T s + \frac{1}{2} s^T \nabla^2_{ww} L\left(w^{(k)}, \lambda^{(k)}\right) s \\
&= F_k - \lambda^{(k)T} c_k + \nabla F_k^T s - \lambda^{(k)T} A_k s + \frac{1}{2} s^T \nabla^2_{ww} L_k s \\
&= F_k - \lambda^{(k)T} (A_k s + c_k) + \nabla F_k^T s + \frac{1}{2} s^T \nabla^2_{ww} L_k s \\
&= F_k + \nabla F_k^T s + \frac{1}{2} s^T \nabla^2_{ww} L_k s
\end{aligned}
$$

Exercício1: Considere o problema

$$\underset{w\in\mathbb{R}^2}{\text{minimizar}}\ F(w) = (w_1-2)^2 + (w_2-1)^2 \ \text{ sujeito a } \quad w_1^2 + (w_2-1)^2 - 1 = 0$$

Aplique o método SQP para resolver o problema, com $w^{(0)} = (-0.8, -0.8)^T$ e $\lambda^{(0)} = -0.9$.

Método SQP (restrições de desigualdade)

O método SQP pode ser facilmente estendido a um problema de otimização não linear com restrições de igualdade e desigualdade:

$$\begin{array}{rll} \underset{w \in \mathbb{R}^{\mathrm{I}}}{\text{minimizar}} & F(w) & \\ \text{sujeito a} & c_n(w) = 0, & n \in \mathcal{E} \\ & c_n(w) \geq 0, & n \in \mathcal{I} \end{array} \tag{P2}$$

Para modelar este problema, no iterando $(w^{(k)}, \lambda^{(k)})$, linearizamos ambas as restrições de igualdade e desigualdade sendo o programa quadrático dado por:

$$\begin{array}{rll} \underset{s \in \mathbb{R}^{\mathrm{I}}}{\text{minimizar}} & F_k + \nabla F_k^T s + \frac{1}{2} s^T \nabla^2_{ww} L(w^{(k)}, \lambda^{(k)}) s & \\ \text{sujeito a} & \nabla c_n(w^{(k)})^T s + c_n(w^{(k)}) = 0 & n \in \mathcal{E} \\ & \nabla c_n(w^{(k)})^T s + c_n(w^{(k)}) \geq 0 & n \in \mathcal{I} \end{array} \tag{Q2}$$

- O problema de PQ (Q2) pode ser resolvido por um algoritmo de Programação Quadrática (ver Capítulo 16 em [1]).
- Assim obtida a solução (s^k, π^k) de (Q2), o novo iterando é dado por: $w^{(k+1)} = w^{(k)} + s^k, \;\; \lambda^{(k+1)} = \pi^k$.

Algoritmo2: Algoritmo SQP Local - Restrições de Desigualdade

- Dar: o par inicial $\left(w^{(0)}, \lambda^{(0)}\right)$
- **Para** $k = 0, 1, \ldots$

1. Calcular F_k, ∇F_k, $\nabla^2_{ww} L(w^{(k)}, \lambda^{(k)})$, $c_n(w^{(k)})$, $\nabla c_n(w^{(k)})$
2. Resolver o subproblema quadrático (Q2) para obter (s^k, π^k)
   (solução, multiplicadores de Lagrange)
3. Fazer $w^{(k+1)} = w^{(k)} + s^k$ e $\lambda^{(k+1)} = \pi^k$
4. **Se** $\left(w^{(k+1)}, \lambda^{(k+1)}\right)$ satisfaz o critério de paragem para o problema (P2)
   Parar com a solução $\left(w^{(k+1)}, \lambda^{(k+1)}\right)$
   fim se

- Neste método SQP-restrições de desigualdade, o conjunto de restrições ativas $\mathcal{A}_k$ na solução do problema (Q2), constitui uma estimativa do conjunto ativo na solução do problema (P2).
- Se o método SQP for capaz de identificar corretamente este conjunto ativo ótimo (e não o altera nas iterações seguintes), então comporta-se como o método de Newton(4)-(5) para otimização com restrições de igualdade e convergirá rapidamente.
- O resultado seguinte dá as condições em que este comportamento desejável se verifica. Recordar que, dizemos que a complementaridade estrita verifica-se na solução (w^*, λ^*) se não existir nenhum índice $n \in \mathcal{I}$ tal que $\lambda_n^* = c_n(w^*) = 0$.

Teorema 1

Suponha que w^* *é uma solução local de* (P2) *na qual as condições KKT são satisfeitas para algum* λ^*. *Suponha-se, também, que a qualificação da restrição de independência linear (LICQ), a condição de complementaridade estrita, e as condições suficientes de segunda ordem se verificam em* (w^*, λ^*). *Então se* $\left(w^{(k)}, \lambda^{(k)}\right)$ *está suficientemente próximo de* (w^*, λ^*), *existe uma solução local do subproblema* (Q2) *cujo conjunto ativo* $\mathcal{A}_k$ *é o mesmo que o conjunto ativo* $\mathcal{A}(w^*)$ *do problema* (P2) *em* w^*.

Pré-visualização de Métodos SQP práticos

Para ser prático, um método SQP deve ser capaz de convergir a partir de pontos de iniciais distantes e para qualquer problema não convexo.

Vamos descrever como a estratégia SQP local pode ser adaptada para atingir estes objetivos:

- Começamos por fazer uma analogia com a otimização sem restrições. Na sua forma mais simples, a iteração de Newton para minimizar uma função F dá um passo para o minimizante do modelo quadrático

$$m_k(s) = F_k + \nabla F_k^T s + \frac{1}{2} s^T \nabla^2 F(w^{(k)}) s$$

▷ Este método é útil perto da solução, onde a Hessiana $\nabla^2 F(w^{(k)})$ é normalmente definida positiva e o modelo quadrático tem um minimizante bem definido.

▷ No entanto, quando $w^{(k)}$ não está próximo da solução, a função m_k pode não ser convexa. Os métodos de procura unidirecional modificam a Hessiana em $m_k(s)$ para a tornar definida positiva (possivelmente substituindo-o por uma aproximação quasi-Newton B_k), para garantir que s^k é uma direção de descida para F.

▶ Nos métodos SQP usam-se estratégias idênticas para a sua globalização.

▷ Se $\nabla^2_{ww} L(w^{(k)}, \lambda^{(k)})$ é definida positiva no espaço tangente das restrições ativas, o subproblema quadrático (Q1) tem uma solução única.

▷ No entanto, quando $\nabla^2_{ww} L(w^{(k)}, \lambda^{(k)})$ não é definida positiva, os métodos SQP de procura unidirecional ou

- substituem $\nabla^2_{ww} L(w^{(k)}, \lambda^{(k)})$ por uma aproximação definida positiva B_k ou
- modificam $\nabla^2_{ww} L(w^{(k)}, \lambda^{(k)})$ diretamente durante o processo de fatorização da matriz.

Em qualquer um destes casos, o subproblema (Q1) fica bem definido.

▶ A técnica utilizada para aceitar ou rejeitar iterandos/pontos também tem impacto na eficiência dos métodos SQP.

▷ Na otimização sem restrições, a função mérito é a própria função objetivo F, e permanece fixa durante todo o processo de minimização.

▷ Para problemas com restrições, usamos uma função mérito (ou um filtro). Os parâmetros devem ser atualizados de forma a garantir que a direção procura, é uma direção de descida para esta função .

Aproximações Quasi-Newton

- A Hessiana da Lagrangeana $\nabla^2_{ww} L\left(w^{(k)}, \lambda^{(k)}\right)$ é formada pelas segundas derivadas da função objetivo e das funções de restrição:

$$\nabla^2_{ww} L\left(w^{(k)}, \lambda^{(k)}\right) = \nabla^2 F(w^{(k)}) - \sum_{n \in \mathcal{E}} \lambda_n \nabla^2 c_n(w^{(k)}) - \sum_{n \in \mathcal{I}} \lambda_n \nabla^2 c_n(w^{(k)})$$

  ▷ Em algumas aplicações, esta informação não é fácil de calcular.

  ▷ Além disso, a matriz $\nabla^2_{ww} L\left(w^{(k)}, \lambda^{(k)}\right)$ pode não ser definida positiva no espaço tangente das restrições ativas.

- Esta dificuldade pode ser ultrapassada substituindo a $\nabla^2_{ww} L\left(w^{(k)}, \lambda^{(k)}\right)$ em (Q2) por uma aproximação quasi-Newton.
- Como as fórmula BFGS provou ser bem sucedida no contexto da otimização sem restrições, podemos usá-la aqui também.

- A fórmula de atualização para B_k, que resulta do passo do iterando k para o iterando $k+1$, usa os vetores s_k e y_k definidos da seguinte forma:

$$s_k = w^{(k+1)} - w^{(k)}, \quad y_k = \nabla_w L(w^{(k+1)}, \lambda^{(k+1)}) - \nabla_w L(w^{(k)}, \lambda^{(k+1)}) \quad (8)$$

e calculamos a nova aproximação B_{k+1} usando a fórmula BFGS.

Fórmula BFGS:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}$$

Nota:

- Podemos ver este processo como a aplicação da atualização quasi-Newton ao caso em que a função objetivo é dada pela função Lagrangeana $L(w, \lambda)$ (com λ fixo). Este ponto de vista de imediato revela a força e a fragilidade destas aproximações.

▷ Se $\nabla^2_{ww} L$ é definida positiva na região onde se encontra o minimizante, então a aproximação B_k quasi-Newton BFGS refletirá alguma da informação sobre a curvatura do problema, e o processo iterativo irá convergir de forma robusta e rápida, tal como no método BFGS sem restrições.

▷ Se, no entanto, $\nabla^2_{ww} L$ tem valores próprios negativos, então o método BFGS de aproximá-la por uma matriz definida positiva pode ser problemática.

De facto, a atualização BFGS exige que s_k e y_k satisfaçam a condição de curvatura $s_k^T y_k > 0$, o que pode não se verificar, quando s_k e y_k são dados por (8), mesmo quando os iterandos estão próximos da solução.

Apresenta-se a seguir duas estratégias para ultrapassar esta dificuldade.

- **Estratégia de skipping**

  Esta estratégia consiste em não atualizar B_k se a condição

  $$s_k^T y_k \geq \gamma s_k^T B_k^T s_k$$

  não é satisfeita, onde γ é um parâmetro positivo ($\gamma = 10^{-2}$, por exemplo).

▷ Esta estratégia pode, por vezes, produzir um desempenho fraco ou mesmo falhar, portanto não pode ser vista como adequada para algoritmos SQP de uso geral.

- **Estratégia damped**

  Esta estratégia consiste em modificar a definição de y_k de modo a garantir que a atualização BFGS esteja sempre bem definida.

Procedimento: Atualização BFGS damped

- Dar: matriz simétrica e definida positiva B_k
- Definir s_k e y_k como em (8) e fazer

$$r_k = \theta_k y_k + (1 - \theta_k) B_k s_k,$$

onde o escalar θ_k é definido por

$$\theta_k = \begin{cases} 1 & \text{se } s_k^T y_k \geq 0.2 s_k^T B_k s_k, \\ \dfrac{0.8 s_k^T B_k s_k}{s_k^T B_k s_k - s_k^T y_k} & \text{se } s_k^T y_k < 0.2 s_k^T B_k s_k. \end{cases} \tag{9}$$

- Atualizar B_k da seguinte forma:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{r_k r_k^T}{s_k^T r_k} \tag{10}$$

(**nota:** A fórmula (10) é simplesmente a fórmula BFGS standard, com y_k substituído por r_k. Garante que B_{k+1} é definida positiva, uma vez que é fácil mostrar que quando $\theta_k \neq 1$ temos $s_k^T r_k = 0.2 s_k^T B_k s_k > 0$)

- Para compreender melhor esta estratégia, notar que a escolha $\theta_k = 0$ dá $B_{k+1} = B_k$, enquanto $\theta_k = 1$ dá a matriz (possivelmente indefinida) produzida pela atualização BFGS não modificada.

▷ Um valor de $\theta_k \in (0, 1)$ produz assim uma matriz entre a aproximação atual B_k e a matriz produzida pela fórmula BFGS não modificada.

▷ A escolha de θ_k assegura que a nova aproximação B_{k+1} se mantém suficientemente próxima da aproximação atual B_k para garantir a positividade.

Nota:

- Todas as aproximações quasi-Newton B_k apresentadas são matrizes $\mathrm{I} \times \mathrm{I}$ densas que podem ser dispendiosas de armazenar e manipular no caso de problemas de grande dimensão. A fórmula de atualização de memória reduzida (do inglês *limited-memory update*) é útil neste contexto e é frequentemente implementada em pacotes de software.

Funções Mérito

- Para garantir que o método SQP com procura unidirecional convirja a partir de qualquer ponto inicial, é frequentemente usado no algoritmo SQP uma função mérito para controlar o tamanho do passo.

  O comprimento do passo, η_k, será aceite se o novo ponto $w^{(k+1)} = w^{(k)} + \eta_k s^k$ produzir uma redução significativa no valor da função mérito.

- Uma variedade de funções mérito têm sido usadas em métodos SQP, nomeadamente funções de penalidade exatas não suaves e Lagrangeanas aumentadas.

  Apenas apresentamos aqui as funções mérito exatas não suaves, tipificada pela função mérito ℓ_1.

- Na descrição que se segue, consideramos o problema com restrições de igualdade (P1), uma vez que as restrições de desigualdade podem ser convertidas em restrições de igualdade pela introdução de variáveis de folga.

  (nota: $c(w) \geq 0$ converter em $\overline{c}(w, v) = c(w) - v = 0$, onde $v \geq 0$ é o vetor das variáveis de folga; a condição $v \geq 0$ é tipicamente não monitorizada pela função mérito.)

Função de penalidade ℓ_1

A função mérito ℓ_1 para (P1) é definida por:

$$\phi_1(w;\mu) = F(w) + \mu||c(w)||_1 \tag{11}$$

Num método de procura unidirecional, o comprimento do passo η_k será aceite se a seguinte condição de redução significativa se verifica:

$$\phi_1(w^{(k)} + \eta_k s^k; \mu_k) \leq \phi_1(w^{(k)}; \mu_k) + \delta\eta_k D(\phi_1(w^{(k)}; \mu_k); s^k) \tag{12}$$

onde

- $D(\phi_1(w^{(k)}; \mu_k); s^k)$ denota a derivada direcional de ϕ_1 na direção de s^k
- $\delta \in]0, 1[$.

▷ A condição (12) é análoga a condição de Armijo para otimização sem restrições desde que s^k seja uma direção de descida para ϕ_1, ou seja,

$$D(\phi_1(w^{(k)}; \mu); s^k) < 0$$

Teorema 2

Sejam s^k e $\lambda^{(k+1)}$ gerados pela iteração SQP (7). Então a derivada direcional de ϕ_1 na direção s^k satisfaz

$$D(\phi_1(w^{(k)};\mu);s^k) = \nabla F_k^T s^k - \mu||c_k||_1 \tag{13}$$

Além disso,

$$D(\phi_1(w^{(k)};\mu);s^k) = -(s^k)^T \nabla^2_{ww} L_k s^k - (\mu - ||\lambda^{(k+1)}||_\infty)||c_k||_1 \tag{14}$$

onde $\|\lambda^{k+1}\| = \max_i |\lambda_i^{k+1}|$

Demonstração: (Teorema 18.3) em [1]

De (14) conclui-se que s^k será uma direção de descida para ϕ_1, se $s^k \neq 0$, $\nabla^2_{ww} L_k$ for definida positiva e

$$\mu > \|\lambda_{k+1}\|_\infty \qquad (15)$$

(nota: uma análise mais cuidada mostra que a hipótese sobre $\nabla^2_{ww} L_k$ pode ser relaxada, sendo apenas necessário que a Hessiana reduzida $Z_k^T \nabla^2_{ww} L_k Z_k$ seja definida positiva.)

Uma estratégia alternativa, baseada em (13), é exigir que a derivada direcional seja suficientemente negativa no sentido de que

$$D(\phi_1(w^{(k)};\mu); s^k) = \nabla F_k^T s^k - \mu ||c_k||_1 \leq -\rho\mu ||c_k||_1$$

para $\rho \in]0,1[$. Esta desigualdade verifica-se se

$$\mu \geq \frac{\nabla F_k^T s^k}{(1-\rho)\|c_k\|_1} \qquad (16)$$

Esta escolha não depende dos multiplicadores de Lagrange e apresenta um desempenho adequado na prática.

▷ Na prática, é desejável que os parâmetros μ_k da função mérito permaneçam inalterados à medida que os iterandos convergem para a solução.

Uma estratégia para escolher o novo valor de μ é a seguinte. Se o valor de μ da iteração anterior satisfaz (16) (ou (15)), o valor de μ não é alterado. Caso contrário, μ é incrementado de modo a satisfazer a desigualdade (16) (ou (15)) com alguma margem. Esta regra é dada por:

$$\mu_k = \begin{cases} \mu_{k-1} & \text{se } \mu_{k-1} \geq \gamma_k + \delta \\ \gamma_k + 2\delta & \text{caso contrário} \end{cases} \tag{17}$$

onde $\delta > 0$ ($\delta = 10^{-2}$, por exemplo), e γ_k é dado por

$$\gamma_k = \begin{cases} \lambda^{(k+1)} & \text{se (15)} \\ \dfrac{\nabla F_k^T s^k}{(1-\rho)\|c_k\|_1} & \text{se (16)} \end{cases}$$

Algoritmo: Algoritmo SQP de procura unidirecional

- Dar: $\delta \in]0, 0.5[$, $\mu_0 > 0$, e o par inicial $\left(w^{(0)}, \lambda^{(0)}\right)$
- Dar: uma matriz B_0 simétrica definida positiva $\mathrm{I} \times \mathrm{I}$ (caso contrário $\nabla^2_{ww} L_0$)
- Calcular F_0, ∇F_0, c_0, A_0
- **Para** $k = 0, 1, \ldots$

1. Resolver o subproblema quadrático (Q2) (ou (Q1)) para obter (s^k, π^k)
2. Fazer $\lambda^{(k+1)} = \pi^k$
3. Escolher μ_k usando a regra (17)
4. Fazer $\alpha_k = 1$
5. **Enquanto** $\phi_1(w^{(k)} + \alpha_k s^k; \mu_k) > \phi_1(w^{(k)}; \mu_k) + \delta \alpha_k D(\phi_1(w^{(k)}; \mu_k); s^k)$
   Fazer $\alpha_k = \frac{\alpha_k}{2}$
   Se $\alpha_k \|s^k\| \leq 10^{-8}$ **parar com** $\alpha_k = 1$
   fim
6. Fazer $w^{(k+1)} = w^{(k)} + \alpha_k s^k$
7. **Se** $\left(w^{(k+1)}, \lambda^{(k+1)}\right)$ satisfaz o critério de paragem para (P2) (ou (P1))
   Parar com a solução $\left(w^{(k+1)}, \lambda^{(k+1)}\right)$
   fim

8. Calcular F_{k+1}, ∇F_{k+1}, c_{k+1}, A_{k+1} (e possivelmente $\nabla^2_{ww} L_{k+1}$)
8. Se é usada uma aproximação quasi-Newton, fazer $s_k = w^{(k+1)} - w^{(k)}$, $y_k = \nabla_w L(w^{(k+1)}, \lambda^{(k+1)}) - \nabla_w L(w^{(k)}, \lambda^{(k+1)})$ e obter B_{k+1} usando uma fórmula quasi-Newton.

J. Nocedal and S. Wright.
Numerical Optimization.
Springer, 2006.