# Métodos de Penalidade
# Método de Penalidade Quadrática

M. Fernanda P. Costa

Departamento de Matemática
Universidade do Minho

# Outline

# Métodos de Penalidade

Considere a seguinte formulação geral do problema do otimização com restrições:

$$\begin{array}{rll} \underset{w \in \mathbb{R}^{\mathrm{I}}}{\text{minimizar}} & F(w) & \\ \text{sujeito a} & c_n(w) = 0, & n \in \mathcal{E} \\ & c_n(w) \geq 0, & n \in \mathcal{I} \end{array} \qquad (1)$$

- $w = (w_1, w_2, \ldots, w_{\mathrm{I}})^T$ é o vetor das variáveis de decisão
- $F : \mathbb{R}^{\mathrm{I}} \to \mathbb{R}$ é a função objetivo (medida de desempenho)
  (loss or cost function in ML)
- $c_n : \mathbb{R}^{\mathrm{I}} \to \mathbb{R}$ com $n \in \mathcal{E}$, são as funções de restrição de igualdade
- $c_n : \mathbb{R}^{\mathrm{I}} \to \mathbb{R}$ com $n \in \mathcal{I}$, são as funções de restrição de desigualdade

- Um classe importante de métodos para resolver problemas de otimização com restrições é os Métodos de Penalidade.

- Estes métodos substituem o problema original por uma sucessão de subproblemas sem restrições, adicionado um termo de penalidade, por restrição, à função objetivo. O termo penaliza, ou seja aumenta o valor da função objetivo, sempre que a restrição é violada.

- Vamos aqui apresentar três abordagens:

  - *Método de Penalidade Quadrática*
  - *Métodos de Penalidade Exata Não Suaves*
  - *Método da Lagrangeana Aumentada* (ou *Método dos Multiplicadores*)

# Método de Penalidade Quadrática

## Motivação

- O problema de otimização com restrições (1) é substituído por uma função de penalidade, definida pela:
  - função objetivo do problema original ***mais***
  - um termo de penalidade, por restrição, que é positivo quando o ponto $w$ viola a restrição e zero caso contrário.
- A maioria das abordagens define uma sucessão de funções de penalidade, em que os termos de penalidade são multiplicados por um coeficiente positivo $\mu$, chamado *parâmetro de penalidade*.
- A função de penalidade mais simples deste tipo é a **função de penalidade quadrática**.

## Método de Penalidade Quadrática

- Considere o problema com restrições (1).
- Define a seguinte função de penalidade quadrática $Q(w;\mu)$, com $\mu > 0$:

$$Q(w;\mu) = \underbrace{F(w)}_{\text{função objetivo}} + \underbrace{\frac{\mu}{2}\sum_{n\in\mathcal{E}}(c_n(w))^2 + \frac{\mu}{2}\sum_{n\in\mathcal{I}}(\max(0,-c_n(w))^2}_{\text{um termo de penalidade por restrição, definido pelo quadrado da violação da restrição}}$$

- Define uma sucessão de subproblemas de minimização sem restrições da forma:

$$\underset{w\in\mathbb{R}^{\mathrm{I}}}{\text{minimizar}} \quad Q(w,\mu_k)$$

  para uma sucessão crescente dos parâmetros de penalidade $\mu$, com $\mu_k \uparrow \infty$ quando $k \to \infty$.
- Ao fazer o valor de $\mu_k \uparrow \infty$, penaliza as violações das restrições de forma mais severa, forçando o minimizante de $Q$ a ficar mais próximo da região admissível do problema original (1).

**Algoritmo1:** Método de Penalidade Quadrática

1. Dar: $\mu_0$, um ponto inicial $w_s^{(0)}$, uma sequência não negativa $\{\tau_k\}$ com $\tau_k \to 0$
2. Para $k = 0, 1, \ldots$
3. Encontrar um minimizante (aproximado) $w^{(k)}$ de $Q(w; \mu_k)$, iniciando em $w_s^{(k)}$ e parar quando $\|\nabla_w Q(w; \mu_k)\| \leq \tau_k$
4. **Se** $w^{(k)}$ satisfaz o critério de paragem final
   **Parar** com a solução (aproximada) $w^{(k)}$
   **fim se**
5. Escolher novo parâmetro de penalidade $\mu_{k+1} > \mu_k$
6. Escolher novo ponto inicial $w_s^{(k+1)}$
7. Escolher nova tolerância $\tau_{k+1} < \tau_k$ (com $\tau_{k+1} > 0$)

**(nota:** $\nabla_w Q(w; \mu_k) = \nabla_w F(w) + \mu_k \sum_{n \in \mathcal{E}} c_n(w) \nabla_w c_n(w) - \mu_k \sum_{n \in \mathcal{I}} \max(0, -c_n(w)) \nabla_w c_n(w)$**)**

**(nota: se $\mathcal{I} \neq \{\}$, o vetor gradiente não está definido nos pontos $w$ para os quais $c_n(w) = 0$ com $n \in \mathcal{I}$)**

**Suavidade dos termos de penalidade:**

- Restrições de igualdade: os termos $c_n^2$ são suaves.

  $\Rightarrow$ quando apenas existem restrições de igualdade, podemos usar técnicas baseados em derivadas para otimização sem restrições para encontrar a solução $w^{(k)}$ de $Q(w;\mu_k)$;

- Restrições de desigualdade: os termos $(\max(0,-c_n))^2$ podem ser menos suaves do que $c_n$.

  **ex:** se uma restrição é $w_1 \geq 0$, então a função $(\max(0,-c_n))^2$ tem uma 2ª derivada descontínua, e portanto $Q$ não é mais duas vezes continuamente diferenciável.

- A minimização de $Q(w;\mu_k)$ torna-se mais difícil de realizar à medida que $\mu_k$ aumenta.

**Escolha do ponto inicial $w_s^k$ para minimizar $Q(w;\mu_k)$:**

- usar um dos minimizantes anteriores de $Q(w;\mu)$: $w^{(k-1)}, w^{(k-2)}, \ldots$ ("warm start": $w_s^k = w^{(k-1)}$)
- ...

**Escolha da sequência $\{\mu_k\}$:** adaptativa, por exemplo, se a minimização $Q(w;\mu_k)$ foi de custo computacional:

- elevado: fazer um aumento modesto, por ex., $\mu_{k+1} = 1.5\mu_k$
- baixo: fazer um aumento maior, por ex., $\mu_{k+1} = 10\mu_k$

**Escolha da sequência $\{\tau_k\}$:** $\tau_k \longrightarrow 0$
(a minimização é realizada progressivamente com mais precisão)

**Critério de paragem $\|\nabla_w Q(w;\mu_k)\| \leq \tau_k$:**

$\Rightarrow$ não há garantias que seja satisfeito, os iterandos podem afastar-se da região admissível quando o $\mu_k$ não é grande o suficiente;

$\Rightarrow$ uma implementação prática, deve incluir salvaguardas que incremente o parâmetro de penalidade (e possivelmente restaure o ponto inicial) quando a violação das restrições não está a decrescer rapidamente o suficiente , ou quando os iterandos parecem ser divergentes.

**nota:** Para escolhas adequadas de $\mu_k$ e do ponto inicial $w_s^k$, a minimização de $Q(w;\mu_k)$ é efetuada em poucos passos/iterações.

**Exercício1:** Considere o problema

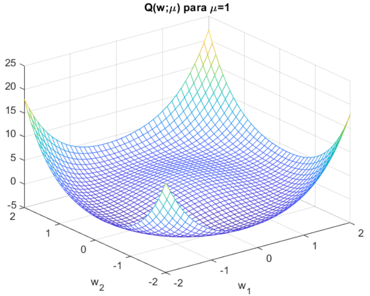
$$\underset{w \in \mathbb{R}^2}{\text{minimizar}} \quad F(w) = w_1 + w_2 \text{ sujeito a } w_1^2 + w_2^2 - 2 = 0$$

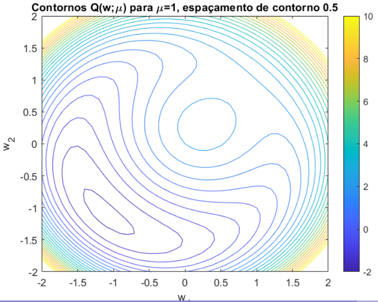
cuja solução é $(-1, -1)^T$. Minimize a função de penalidade quadrática $Q(w; \mu)$ do problema, com $\mu = 1$ e $\mu = 10$, usando um algoritmo de otimização sem restrições. Faça os gráficos dos contornos de $Q$.

(nota: use a função *fminunc* do MatLab para calcular os minimizantes/maximizantes de $Q$)

**Solução:** $Q(w; \mu) = w_1 + w_2 + \frac{\mu}{2}(w_1^2 + w_2^2 - 2)^2$

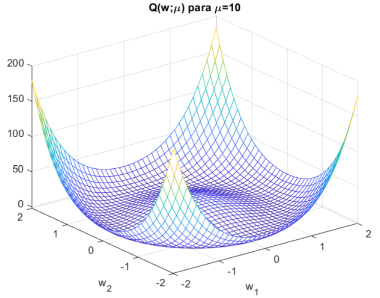
- Para $\mu = 1$, a função $Q$ tem um minimizante no ponto $(-1.1072, -1.1072)^T$ e um maximizante no ponto $(0.2696, 0.2696)^T$.

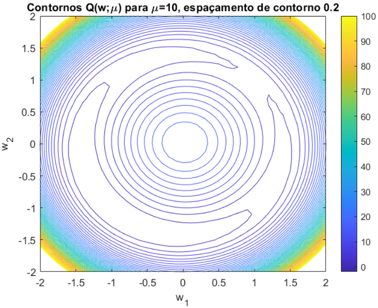


- Para $\mu = 10$, os pontos que não satisfazem a restrição sofrem uma maior penalização - o "vale" de valores baixos de $Q$ é evidente. A função $Q$ tem um minimizante no ponto $(-1.0123, -1.0123)^T$, que está mais perto da solução $(-1, -1)^T$.

  $Q$ tem um maximizante no ponto $(0.025, 0.025)^T$, e $Q$ vai rapidamente para $\infty$ para pontos fora da circunferência $w_1^2 + w_2^2 = 2$.



A situação nem sempre é tão *favorável* como a reportada no Exercício1. Para um dado valor do parâmetro de penalidade $\mu$, a função de penalidade pode ser ilimitada inferiormente ainda que o problema com restrições original tenha uma única solução. Esta deficiência é, infelizmente, comum a todas as funções de penalidade aqui apresentadas.

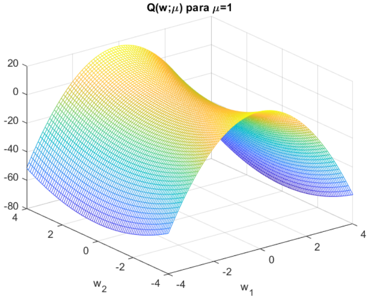
**Exercício2:** Considere o problema

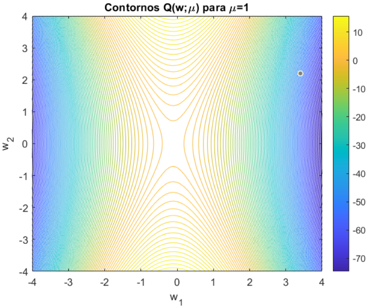
$$\underset{w \in \mathbb{R}^2}{\text{minimizar}} \quad F(w) = -5w_1^2 + w_2^2 \text{ sujeito a } w_1 = 1$$

cuja solução é $(1,0)^T$. Minimize a função de penalidade quadrática $Q(w;\mu)$ do problema, com $\mu = 1$ e $\mu = 100$, usando um algoritmo de otimização sem restrições. Faça os gráficos dos contornos de $Q$.

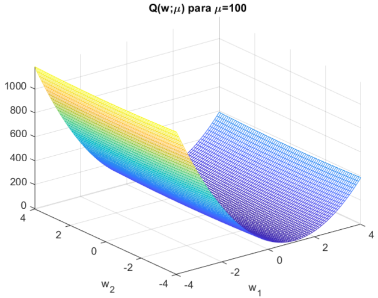
**Solução:** a função de penalidade $Q(w;\mu) = -5w_1^2 + w_2^2 + \frac{\mu}{2}(w_1 - 1)^2$ é ilimitada para qualquer $\mu < 10$.

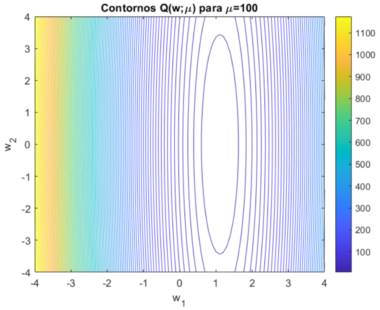
- Para $\mu = 1 < 10$, a função $Q(w;1)$ é ilimitada inferiormente. Usando o fminunc, o algoritmo pára ao fim de 1ª iteração com exitflag=-3 *'objective function at current iteration went below ObjectiveLimit'*.



- Para $\mu = 100$ e usando o fminunc, o algoritmo pára ao fim de 2ª iteração com exitflag=1 *'magnitude of gradient is smaller than the OptimalityTolerance tolerance'*, e $Q(w; 10)$ tem um minimizante no ponto $(1.1111, 0)^T$, que está perto da solução $(1, 0)^T$.



**Exercício3:** Considere novamente o problema

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

a) Minimize a função de penalidade quadrática $Q$ do problema, para $\mu_k = 1, 10, 100, 1000$ usando um algoritmo de otimização sem restrições.

b) Implemente o Método de Penalidade Quadrática (Algoritmo1) para resolver o problema, com $w^{(0)} = (0,0)^T$ e $\mu_0 = 1$. Faça $\mu_{k+1} = 10\mu_k$, $\tau_k = 1/\mu_k$ e escolha para ponto inicial $w_s^{(k+1)}$ a solução de $Q(w;\mu_k)$. Para critério de paragem final use as condições:

$$\|c_{\mathcal{E}}(w^{(k)})\|_\infty \leq \varepsilon_1 \text{ (medida da violação)}$$

$$\frac{|F(w^{(k)}) - F(w^{(k-1)})|}{|Fw^{(k)}|} \leq \varepsilon_2, \quad \frac{\|w^{(k)} - w^{(k-1)}\|_\infty}{\|w^{(k)}\|_\infty} \leq \varepsilon_3$$

com $\varepsilon_1 = 10^{-6}, \varepsilon_2 = \varepsilon_3 = 10^{-4}$.
Indique a solução (aproximada) de cada função de penalidade $Q(w;\mu_k)$.

c) Compare a solução ótima obtida em b) com a solução obtida usando a função fmincon do MatLab.

**Exercício4:** Considere agora o problema

$$\underset{w \in \mathbb{R}^2}{\text{minimizar}}\ F(w) = \frac{1}{2} w_1^2 + w_2^2 - w_1 w_2 - 2w_1 - 6w_2$$

$$\text{sujeito a } c(w) = \begin{cases} -w_1 - w_2 + 2 & \geq 0 \\ w_1 - 2w_2 + 2 & \geq 0 \\ -2w_1 - w_2 + 3 & \geq 0 \\ w_1, w_2 & \geq 0 \end{cases}$$

a) Utilize o Método de Penalidade Quadrática para resolver o problema, com $w^{(0)} = (0,0)^T$ e $\mu_0 = 1.5$. Faça $\mu_{k+1} = 2\mu_k$, $\tau_k = 1/\mu_k$ e escolha para ponto inicial $w_s^{(k+1)}$ a solução de $Q(w; \mu_k)$. Para critério de paragem final use as condições:

$$\|\max(0, -c_{\mathcal{I}}(w^{(k)}))\|_\infty \leq \varepsilon_1 \quad \text{(medida da violação)}$$

$$\frac{|F(w^{(k)}) - F(w^{(k-1)})|}{|Fw^{(k)}|} \leq \varepsilon_2, \quad \frac{\|w^{(k)} - w^{(k-1)}\|_\infty}{\|w^{(k)}\|_\infty} \leq \varepsilon_3$$

com $\varepsilon_1 = 10^{-6}, \varepsilon_2 = \varepsilon_3 = 10^{-6}$.

b) Compare a solução ótima obtida em a) com a solução obtida usando a função fmincon do MatLab.

**Exercício5:** (HS6) Considere o problema

$$\underset{w\in\mathbb{R}^2}{\text{minimizar}}\ F(w) = (1-w_1)^2 \ \text{ sujeito a } \ 10(w_2 - w_1^2) = 0$$

a) Utilize o Método de Penalidade Quadrática para resolver o problema, com $w^{(0)} = (-1.2, 1)^T$ e $\mu_0 = 1$. Faça $\mu_{k+1} = 2\mu_k$, $\tau_k = 1/\mu_k$ e escolha para ponto inicial $w_s^{(k+1)}$ a solução de $Q(w;\mu_k)$. Para critério de paragem final use as condições:

$$\|c_{\mathcal{E}}(w^{(k)})\|_\infty \leq \varepsilon_1 \ \text{(medida da violação)}$$

$$\frac{|F(w^{(k)}) - F(w^{(k-1)})|}{|Fw^{(k)}|} \leq \varepsilon_2, \quad \frac{\|w^{(k)} - w^{(k-1)}\|_\infty}{\|w^{(k)}\|_\infty} \leq \varepsilon_3$$

com $\varepsilon_1 = 10^{-6}, \varepsilon_2 = \varepsilon_3 = 10^{-6}$.

b) Compare a solução ótima obtida em a) com a solução obtida usando a função fmincon do MatLab.

**Exercício6:** (HS32) Considere o problema

$$\underset{w\in\mathbb{R}^3}{\text{minimizar}}\ F(w) = (w_1 + 3w_2 + w_3)^2 + 4(w_1 - w_2)^2$$

$$\text{sujeito a} \left\{ \begin{array}{rl} 6w_2 + 4w_3 - w_1^3 - 3 & \geq 0 \\ 1 - w_1 - w_2 - w_3 & = 0 \\ w_1, w_2, w_3 & \geq 0 \end{array} \right.$$

a) Utilize o Método de Penalidade Quadrática para resolver o problema, com $w^{(0)} = (0.1, 0.7, 0.2)^T$ e $\mu_0 = 1.5$. Faça $\mu_{k+1} = 2\mu_k$, $\tau_k = 1/\mu_k$ e escolha para ponto inicial $w_s^{(k+1)}$ a solução de $Q(w; \mu_k)$. Para critério de paragem final use as condições:

$$\|c_{\mathcal{E}}(w^{(k)}), \max(0, -c_{\mathcal{I}}(w^{(k)}))\|_\infty \leq \varepsilon_1 \ \text{(medida da violação)}$$

$$\frac{|F(w^{(k)}) - F(w^{(k-1)})|}{|Fw^{(k)}|} \leq \varepsilon_2, \quad \frac{\|w^{(k)} - w^{(k-1)}\|_\infty}{\|w^{(k)}\|_\infty} \leq \varepsilon_3$$

com $\varepsilon_1 = 10^{-6}, \varepsilon_2 = \varepsilon_3 = 10^{-6}$.

b) Compare a solução ótima obtida em a) com a solução obtida usando a função fmincon do MatLab.

# Convergência do método de penalidade quadrática

São descritas algumas das propriedades do método em dois teoremas. Será apenas considerado o problema com restrições de igualdade

$$\begin{array}{ll} \underset{w\in\mathbb{R}^{\mathrm{I}}}{\text{minimizar}} & F(w) \\ \text{sujeito a} & c_n(w)=0, \quad n\in\mathcal{E} \end{array} \tag{2}$$

para o qual a função de penalização quadrática é definida por

$$Q(w;\mu)=F(w)+\frac{\mu}{2}\sum_{n\in\mathcal{E}}(c_n(w))^2 \tag{3}$$

e assume-se que $\mu_k\uparrow\infty$ (i.e., $\mu_{k+1}>\mu_k$ para todo $k$ e $\lim_{k\to+\infty}\mu_k=+\infty$).

Para o primeiro resultado, assume-se que a função de penalidade $Q(w; \mu_k)$ tem um minimizante global para cada valor de $\mu_k$.

**Teorema 1**

*Suponha que cada $w^{(k)}$ é o minimizante global exato de $Q(w; \mu_k)$ em (3) no Algoritmo1, e que $\mu_k \uparrow \infty$. Então todo o ponto limite $w^*$ da sucessão de soluções $\{w^{(k)}\}$, dos respetivos problemas sem restrições $Q(w; \mu_k)$, é uma solução global do problema com restrições (2).*

(**nota:** em geral, impraticável pois exige a minimização global de cada subproblema $Q(w; \mu_k)$ (ou um problema convexo))

## Demonstração

Seja $\overline{w}$ a solução global de (2), isto é,

$$F(\overline{w}) \leq F(w) \ \text{ para todo } w \text{ com } \ c_n(w) = 0, \ \ n \in \mathcal{E}$$

Como $w^{(k)}$ minimiza $Q(w;\mu_k)$ para cada $k$, tem-se que $Q(w^{(k)};\mu_k) \leq Q(\overline{w};\mu_k)$, donde segue a desigualdade

$$F(w^{(k)}) + \frac{\mu_k}{2}\sum_{n\in\mathcal{E}}(c_n(w^{(k)}))^2 \leq F(\overline{w}) + \frac{\mu_k}{2}\sum_{n\in\mathcal{E}}(c_n(\overline{w}))^2 = F(\overline{w}) \tag{4}$$

Reorganizando esta expressão, obtém-se

$$\sum_{n\in\mathcal{E}}(c_n((w^{(k)}))^2 \leq \frac{2}{\mu_k}[F(\overline{w}) - F(w^{(k)})]. \tag{5}$$

Suponha que $w^*$ é o limite da sucessão $\{w^{(k)}\}$, então existe uma subsucessão infinita $\mathcal{K}$ tal com

$$\lim_{k\in\mathcal{K}} w^{(k)} = w^*$$

Fazendo o limite com $k \to \infty$ para $k \in \mathcal{K}$, em ambos os lados de (5), obtém-se

$$\sum_{n\in\mathcal{E}}(c_n(w^*))^2 = \lim_{k\in\mathcal{K}}\sum_{n\in\mathcal{E}}(c_n((w^{(k)}))^2 \leq \lim_{k\in\mathcal{K}}\frac{2}{\mu_k}[F(\overline{w}) - F(w^{(k)})] = 0$$

onde a última igualdade vem de $\mu_k \uparrow \infty$. Portanto, tem-se que $c_n(w^*) = 0$ para todo $n \in \mathcal{E}$, pelo que $w^*$ é ponto admissível. Além disso, fazendo o limite com $k \to \infty$ para $k \in \mathcal{K}$ em (4), tem-se pela não negatividade de $\mu_k$ e de cada $(c_n((w^{(k)}))^2$ que

$$F(w^*) \leq F(w^*) + \lim_{k\in\mathcal{K}}\frac{\mu_k}{2}\sum_{n\in\mathcal{E}}(c_n(w^{(k)}))^2 \leq F(\overline{w})$$

Como $w^*$ é um ponto admissível cujo valor da função objetivo não é maior do que na solução global $\overline{w}$, conclui-se que $w^*$, também, é uma solução global do problema (2).

$\square$

- O segundo resultado diz respeito às propriedades de convergência da sucessão $\{w^{(k)}\}$ quando são efetuadas minimizações não exatas (mas cada vez mais exatas/precisas) de $Q(w; \mu_k)$.

- Em contraste com o Teorema 1, mostra que a sucessão pode convergir:

  ▷ para pontos não-admissíveis, ou

  ▷ para um ponto KKT (um ponto que satisfaz as condições necessárias de 1ª ordem), em vez de um minimizante. Também mostra que as quantidades $\mu_k \, c_n(w^{(k)})$ podem ser usadas como estimativas dos multiplicadores de Lagrange $\lambda_n^*$.

Para estabelecer os resultados do Teorema 2, assume-se que o critério de paragem $\|\nabla_w Q(w;\mu_k)\| \leq \tau_k$ é satisfeito para todo $k$.

## Teorema 2

*Suponha que as tolerâncias e os parâmetros de penalidade no Algoritmo1 satisfazem $\tau_k \to 0$ e $\mu_k \uparrow \infty$. Se um ponto limite $w^*$ da sucessão é não-admissível, então $w^*$ é um ponto estacionário da função $\|c(w)\|^2$. Por outro lado, se um ponto limite $w^*$ da sucessão é um ponto admissível e os gradientes $\nabla c_n(w^*)$ são linearmente independentes, então $w^*$ é um ponto KKT do problema com restrições (2). Para tais pontos, para qualquer subsucessão infinita tal que $\lim_{k\in\mathcal{K}} w^{(k)} = w^*$, tem-se que*

$$\lim_{k\in\mathcal{K}} -\mu_k c_n(w^{(k)}) = \lambda_n^*, \quad \textit{para todo } \; n \in \mathcal{E} \tag{6}$$

*onde $\lambda^*$ é o vector dos multiplicadores de Lagrange que satisfazem as condições KKT para o problema com restrições de igualdade (2).*

## Demonstração

Derivando $Q(w;\mu_k)$ em (3), obtém-se:

$$\nabla_w Q(w^{(k)};\mu_k) = \nabla F(w^{(k)}) + \sum_{n\in\mathcal{E}} \mu_k c_n(w^{(k)})\nabla c_n(w^{(k)}). \tag{7}$$

Pelo critério de paragem do Algoritmo1, tem-se que

$$\|\nabla F(w^{(k)}) + \mu_k \sum_{n\in\mathcal{E}} c_n(w^{(k)})\nabla c_n(w^{(k)})\| \leq \tau_k. \tag{8}$$

Usando a desigualdade $\|a\| - \|b\| \leq \|a+b\|$, obtém-se

$$\mu_k \| \sum_{n\in\mathcal{E}} c_n(w^{(k)})\nabla c_n(w^{(k)})\| - \|\nabla F(w^{(k)}\| \leq \tau_k \Leftrightarrow$$

$$\| \sum_{n\in\mathcal{E}} c_n(w^{(k)})\nabla c_n(w^{(k)})\| \leq \frac{1}{\mu_k}[\tau_k + \|\nabla F(w^{(k)})\|]. \tag{9}$$

Seja $w^*$ um ponto limite da sucessão $\{w^{(k)}\}$.

Então existe uma subsucessão $\mathcal{K}$ tal que $\lim_{k\in\mathcal{K}} w^{(k)} = w^*$.

Fazendo o limite com $k \to \infty$ para $k \in \mathcal{K}$, em ambos os lados de (9), o termo dentro do parêntesis do lado direito aproxima-se de $\|\nabla F(w^*)\|$. Como $\mu_k \uparrow \infty$, o lado direito aproxima-se de zero. Considerando o limite no lado esquerdo, obtém-se:

$$\sum_{n\in\mathcal{E}} c_n(w^*)\nabla c_n(w^*) = 0. \tag{10}$$

Podemos ter $c_n(w^*) \neq 0$ se os gradientes das restrições $\nabla c_n(w^*)$ são linearmente dependentes em $w^*$. Neste caso (10) implica que $w^*$ é um ponto estacionário da função $\|c(w)\|^2$.

Se, por outro lado, os gradientes das restrições $\nabla c_n(w^*)$ são linearmente independentes, tem-se de (10) que $c_n(w^*) = 0$ para todo $n \in \mathcal{E}$, portanto $w^*$ é um ponto admissível. Logo, a 2ª condição das condições KKT (condição de admissibilidade) é satisfeita. É necessário verificar também a 1ª condição KKT e mostrar que o limite (6) é válido.

Denotando por $A(w)$ a matriz dos gradientes das restrições

$$A(w)^T = [\nabla c_n]_{n\in\mathcal{E}},$$

e denotando por $\lambda^{(k)}$ o vetor $-\mu_k c(w^{(k)})$, tem-se por (7) e (8) que

$$A(w^{(k)})^T \lambda^{(k)} = \nabla F(w^{(k)}) - \nabla_w Q(w^{(k)};\mu_k), \quad \|\nabla_w Q(w^{(k)};\mu_k)\| \leq \tau_k. \tag{11}$$

Para todo $k \in \mathcal{K}$ suficientemente grande, a matriz $A(w^{(k)})$ tem característica completa por linhas, assim $A(w^{(k)})A(w^{(k)})^T$ é não singular. Multiplicando (11) por $A(w^{(k)})$ e reorganizando obtém-se:

$$\lambda^{(k)} = \left[A(w^{(k)})A(w^{(k)})^T\right]^{-1} A(w^{(k)}) \left[\nabla F(w^{(k)}) - \nabla_w Q(w^{(k)};\mu_k)\right]$$

Portanto, fazendo o limite $k \to \infty$ para $k \in \mathcal{K}$, obtém-se

$$\lim_{k\in\mathcal{K}} \lambda^{(k)} = \lambda^* = \left[A(w^*)A(w^*)^T\right]^{-1} A(w^*)\nabla F(w^*).$$

Fazendo o limite em (8), conclui-se que

$$\nabla F(w^*) - A(w^*)^T \lambda^* = 0,$$

assim $\lambda^*$ satisfaz a 1ª condição KKT para o problema (2). Portanto, $w^*$ é um ponto KKT para o problema (2), com único vetor de multiplicadores de Lagrange $\lambda^*$. □

## Comportamento da matriz Hessiana do método de penalidade quadrática

- Vamos agora analisar a natureza do mau-condicionamento na matriz Hessiana $\nabla^2_{ww} Q(w; \mu_k)$.
- A compreensão das propriedades desta matriz, e das Hessianas semelhantes que surgem noutros noutros métodos de penalização e de barreira, é importante para escolher algoritmos eficazes para o problema de minimização e para os cálculos de álgebra linear em cada iteração.

Considere apenas o problema com restrições de igualdade (2). A matriz Hessiana da função de penalidade quadrática é dada por:

$$\nabla^2_{ww} Q(w; \mu_k) = \nabla^2 F(w^*) + \sum_{n \in \mathcal{E}} \mu_k c_n(w) \nabla^2 c_n(w) + \mu_k A(w)^T A(w)$$

**nota:**

- $A(w)^T = [\nabla c_n(w)]_{\mathrm{I} \times |\mathcal{E}|}$ é a matriz dos gradientes das restrições (conhecida por matriz do Jacobiano; geralmente $rank(A) = |\mathcal{E}| < \mathrm{I}$);

Quando $w$ está próximo de um minimizante de $Q(;\mu_k)$, e as condições do Teorema 2 são satisfeitas, tem-se que $\mu_k c_n(w) \approx -\lambda_n^*$ e portanto

$$\begin{aligned}\nabla^2_{ww} Q(w;\mu_k) \approx & \quad \nabla^2 F(w^*) - \sum_{n\in\mathcal{E}} \lambda_n^* \nabla^2 c_n(w) + \mu_k A(w)^T A(w) \\ \approx & \quad \underbrace{\nabla^2_{ww} L(w,\lambda^*)}_{\text{independente de } \mu_k} \quad + \quad \underbrace{\mu_k A(w)^T A(w)}_{\text{rank } |\mathcal{E}| \text{ com } \textit{eignvalues} \text{ não nulos da ordem } \mu_k}\end{aligned}$$

- Em geral como o número de restrições $|\mathcal{E}| < \mathrm{I}$, $A(w)^T A(w)$ é singular.
- A matriz $\nabla^2_{ww} Q(w;\mu_k)$ tem alguns dos seus *eigenvalues* que se aproximam de uma constante, enquanto outros são da ordem $\mu_k$.
- Como $\mu_k \uparrow \infty$, o crescente mau-condicionamento da matriz $\nabla^2_{ww} Q(w;\mu_k)$ é evidente.
- Os métodos de otimização sem restrições (de 1ª e 2ª ordem) tem problemas com o mau-condicionamento, a menos que sejam acompanhados por uma estratégia de pré-condicionamento que elimine o mau-condicionamento sistemático.
- **nota:** $L(w,\lambda) = F(w) - \sum_{n\in\mathcal{E}} \lambda_n c_n(w)$ é a função Lagrangeana.

Uma consequência do mau-condicionamento é os possíveis erros ao aplicar o método de Newton para calcular a direção de procura $s$ para minimizar $Q(w; k)$, que se obtém resolvendo o seguinte sistema:

$$\nabla^2_{ww} Q(w; \mu_k) s = -\nabla_w Q(w; \mu_k)$$

Em geral, o mau-condicionamento da matriz Hessiana conduz a erros significativos no cálculo do valor de $s$, independentemente da técnica computacional utilizada para encontrar a solução deste sistema.