Otimização Com Restrições

M. Fernanda P. Costa

Departamento de Matemática
Universidade do Minho

Outline

Otimização com restrições

Vamos agora estudar o problema geral de minimização de uma função sujeita a restrições nas variáveis. Uma formulação geral para estes problemas é:

$$\begin{array}{lll} \underset{w\in\mathbb{R}^d}{\text{minimizar}} & F(w) & \\ \text{sujeito a} & c_n(w) = 0, & n \in \mathcal{E} = \{1,\ldots,j\} \\ & c_n(w) \geq 0, & n \in \mathcal{I} = \{j+1,\ldots,N\} \end{array} \qquad (\mathrm{P}_{CR})$$

- $w = (w_1, w_2, \ldots, w_d)$ são as variáveis de decisão
- $F : \mathbb{R}^d \to \mathbb{R}$ é a função objetivo (medida de desempenho)
- $c_n : \mathbb{R}^d \to \mathbb{R}$ com $n \in \mathcal{E}$, são as funções de restrição de igualdade
- $c_n : \mathbb{R}^d \to \mathbb{R}$ com $n \in \mathcal{I}$, são as funções de restrição de desigualdade

Definição: Chama-se **ponto admissível** para (P_{CR}) a um ponto que verifica todas as restrições.

Definição: Ao conjunto de todos os pontos admissíveis para (P_{CR}), chama-se **conjunto admissível** e será denotado por $\mathcal{D}$.

$$\mathcal{D} = \{w \in \mathbb{R}^d : c_n(w) = 0, n \in \mathcal{E}; c_n(w) \geq 0, n \in \mathcal{I}\}$$

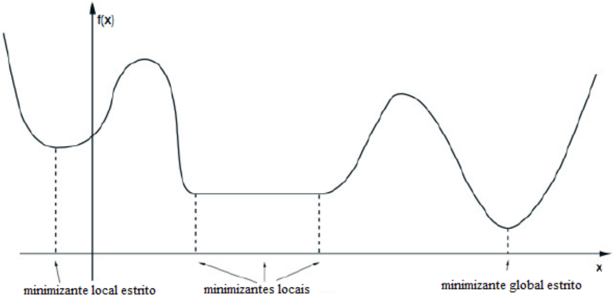
▷ Minimizantes Local e Global:

- $w^* \in \mathbb{R}^d$ é um minimizante global sse $w^* \in \mathcal{D}$ e satisfaz a condição:

$$F(w^*) \leq F(w), \text{ para todo } w \in \mathcal{D}$$

- $w^* \in \mathbb{R}^d$ é um minimizante local sse $w^* \in \mathcal{D}$ e existe uma vizinhança $B(w^*; \epsilon)$ de w^* de raio $\epsilon > 0$ que satisfaz a condição:

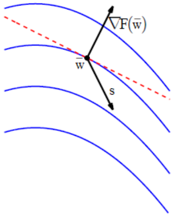
$$F(w^*) \leq F(w), \ \forall w \in B(w^*, \varepsilon) \cap \mathcal{D}$$



- A suavidade da função objetivo e das funções de restrição é uma questão importante na caracterização de soluções, tal como no caso da otimização sem restrições.
- Garante que a função objetivo e as funções restrições se comportam de maneira razoavelmente previsível e, portanto, permite que os algoritmos façam boas escolhas para as direções de procura.

$\triangleright$ Conjunto de direções de descida para F a partir do ponto $\overline{w}$ é o conjunto dos vetores $s \in \mathbb{R}^d$ que satisfazem a condição $\nabla F(\overline{w})^T s < 0$:

$$\{s \in \mathbb{R}^d : \nabla F(\overline{w})^T s < 0\}$$



($\nabla F(\overline{w})^T s < 0 \Rightarrow$ o declive de F em $\overline{w}$ na direção de s é negativo)

Definição: restrição ativa ou não ativa
Seja $\overline{w} \in \mathcal{D}$ um ponto admissível. Uma restrição de desigualdade, $c_n(w) \geq 0$, é dita ativa no ponto $\overline{w}$, se $c_n(\overline{w}) = 0$. Caso $c_n(\overline{w}) > 0$, diz-se que c_n é não ativa no ponto $\overline{w}$.

Definição: conjunto ativo num ponto admissível
É conjunto dos índices das restrições de igualdade e dos índices das restrições de desigualdade ativas no ponto admissível $\overline{w}$:

$$\mathcal{A}(\overline{w}) = \mathcal{E} \cup \{n : c_n(\overline{w}) = 0, n \in \mathcal{I}\}.$$

Derivadas das funções de restrição $c_n : \mathbb{R}^d \to \mathbb{R}$ para $n \in \mathcal{E} \cup \mathcal{I}$

- Vetores gradientes (1ª derivada) das restrições c_n ($n = 1, \ldots, N$):

$$\nabla c_n(w) = \begin{pmatrix} \dfrac{c_n}{\partial w_1} \\ \vdots \\ \dfrac{\partial c_n}{\partial w_d} \end{pmatrix} \in \mathbb{R}^d$$

- Matrizes hessianas (2ª derivada) das restrições c_n ($n = 1, \ldots, N$):

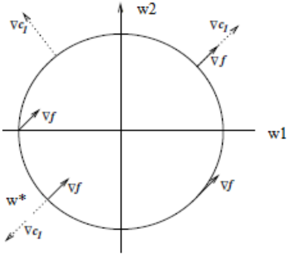
$$\nabla^2 c_n(w) = \begin{pmatrix} \dfrac{\partial^2 c_n}{\partial w_1^2} & \cdots & \dfrac{\partial^2 c_n}{\partial w_d \partial w_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 c_n}{\partial w_1 \partial w_d} & \cdots & \dfrac{\partial^2 c_n}{\partial w_d^2} \end{pmatrix} \in \mathbb{R}^{d \times d}(\text{matriz simétrica})$$

Existem N vetores gradientes e N matrizes hessianas das restrições.

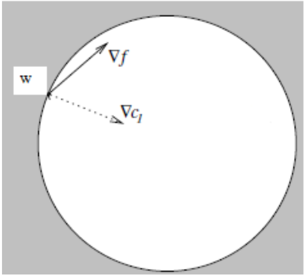
Propriedade do Vetor Gradiente das Restrições

Seja $\mathcal{A}(\overline{w})$ o conjunto ativo no ponto $\overline{w} \in \mathcal{D}$. O vetor gradiente $\nabla c_n(\overline{w})$ da restrição $c_n(w) = 0$ (ou $c_n(w) \geq 0$) é perpendicular à superfície da restrição no ponto $\overline{w}$ (e no caso da restrição de desigualdade o vetor aponta no sentido do lado admissível da região), para $n \in \mathcal{A}(\overline{w})$.

Exemplo: A figura seguinte mostra os vetores gradientes da função objetivo e da função da restrição em vários pontos admissíveis.



(a) $w_1^2 + w_2^2 - 2 = 0$



(b) $w_1^2 + w_2^2 - 2 \leq 0$

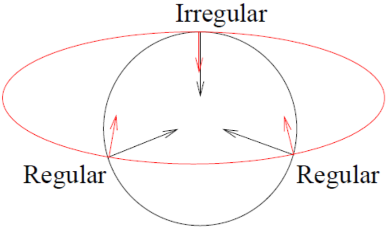
Definição (Ponto Regular)

Seja $\overline{w} \in \mathcal{D}$ um ponto admissível e $\mathcal{A}(\overline{w})$ o conjunto ativo em $\overline{w}$. O ponto admissível é designado por ponto regular se o conjunto dos gradientes das restrições ativas em $\overline{w}$,

$$\{\nabla c_n(\overline{w}) : n \in \mathcal{A}(\overline{w})\}$$

é linearmente independente.

Exemplo: Problema com duas restrições de igualdade. 3 pontos que verificam as restrições. 2 pontos são regulares e 1 é não regular.



Exemplo:
Considere as restrições definidas por: $c_1(w) = w_1^2 + w_2^2 + w_3^2 - 3 = 0$ e $c_2(w) = 2w_1 - 4w_2 + w_3^2 + 1 = 0$. Verifique se o ponto $\overline{w} = (1,1,1)^T$ é admissível e regular.

Resolução:
$\overline{w}$ é ponto admissível pois verifica as restrições: $c_1(1,1,1) = 0$ e $c_2(1,1,1) = 0$.
Os vetores gradientes das restrições no ponto $(1,1,1)^T$,

$$\nabla c_1(w) = \begin{pmatrix} 2w_1 \\ 2w_2 \\ 2w_3 \end{pmatrix} \rightarrow \nabla c_1(1,1,1) = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

$$\nabla c_2(w) = \begin{pmatrix} 2w_1 \\ -4w_2 \\ 2w_3 \end{pmatrix} \rightarrow \nabla c_2(1,1,1) = \begin{pmatrix} 2 \\ -4 \\ 2 \end{pmatrix}$$

são linearmente independentes, logo $\overline{w}$ é ponto regular.

Direções admissíveis

Definição (Direção admissível)
Seja $\overline{w} \in \mathcal{D}$ um ponto admissível. Uma direção $s \in \mathbb{R}^d \setminus \{0\}$ é uma direção admissível a partir de $\overline{w}$, se existir um escalar suficientemente pequeno $\eta_\varepsilon > 0$ tal que

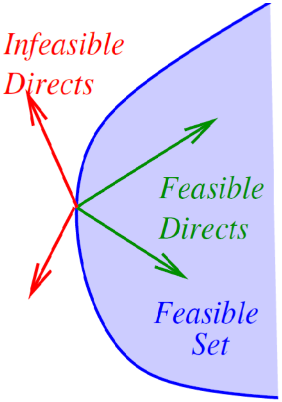
$$\overline{w} + \eta s \in \mathcal{D}, \quad \text{para todo } \eta \in (0, \eta_\varepsilon)$$

Uma condição necessária de otimalidade geral é: Se $w^* \in \mathcal{D}$ é minimizante local do problema de otimização, então não existem direções admissíveis de descida para F a partir de w^*:

$$\{s \in \mathbb{R}^d : \nabla F(w^*)^T s < 0, \forall s \text{ direções admissíveis}\} = \{\}$$

As direções admissíveis são fundamentais para a Otimalidade!
Vamos considerar dois casos distintos:

1. Problemas com restrições de igualdade.
2. Problemas com restrições de desigualdade.



Problemas com restrições de igualdade

$$\begin{array}{ll} \underset{w \in \mathbb{R}^d}{\text{minimizar}} & F(w) \\ \text{sujeito a} & c_n(w) = 0, \quad n = 1, \ldots, N \end{array} \tag{1}$$

Ideia:

Seja $\overline{w} \in \mathcal{D}$ um ponto admissível. $\overline{w}$ não é ponto ótimo se existir um passo infinitesimal $\delta := \eta s$ $(\eta \in (0, \eta_\varepsilon))$ que mantém a admissibilidade e diminui o valor de F. δ diminui o valor F se

$$\nabla F(\overline{w})^T \delta < 0.$$

δ mantém a admissibilidade se $c_n(\overline{w} + \delta) = 0$. Fazendo a expansão da série de Taylor de 1ª ordem em torno de $\overline{w}$, tem-se

$$0 = c_n(\overline{w} + \delta) \approx c_n(\overline{w}) + \nabla c_n(\overline{w})^T \delta = \nabla c_n(\overline{w})^T \delta.$$

A existência de tal δ significa que s tem as propriedades:

$$\nabla c_n(\overline{w})^T s = 0 \quad \text{e} \quad \nabla F(\overline{w})^T s < 0$$

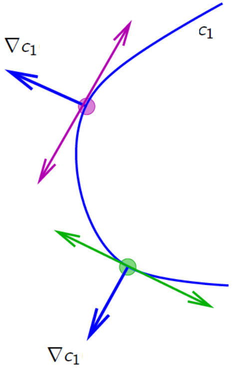
Interpretação geométrica das direções admissíveis

Condição suficiente para direção admissível s (n=1,...,N)

$$\nabla c_n(\overline{w})^T s = 0.$$

Direções admissíveis são direções tangentes às curvas $c_n(w) = 0, \forall n \in \mathcal{E}$.

Exemplo: Problema com uma restrição de igualdade.

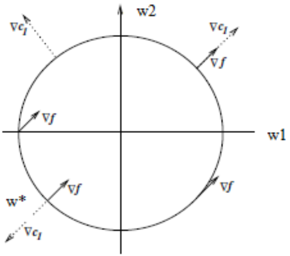


(a) Direções admissíveis em dois pontos diferentes.

Exemplo:

$$\begin{array}{ll} \underset{w \in \mathbb{R}^2}{\text{minimizar}} & F(w) = w_1 + w_2 \\ \text{sujeito a} & c_1(w) = w_1^2 + w_2^2 - 2 = 0 \end{array}$$

- gradiente de F e c_1: $\nabla F(w) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \nabla c_1(w) = \begin{pmatrix} 2w_1 \\ 2w_2 \end{pmatrix}$
- conjunto admissível: circunferência centrada na origem de raio $\sqrt{2}$.
- solução ótima: $w* = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$



- da figura vemos que na solução w^*, ambos os vetores $\nabla c_1(w^*)$ e $\nabla F(w^*)$ são perpendiculares à restrição no ponto w^*, e portanto são paralelos um ao outro. Portanto, existe uma constante λ_1^* tal que

$$\nabla F(w^*) = \lambda_1^* \nabla c_1(w^*).$$

Neste caso, $\lambda_1^* = -1/2$.

As condições de otimalidade podem ser estabelecidas em termos da **Função Lagrangiana**.

A Função Lagrangiana é definida pela função objetivo F e pelas funções de restrição c_n. A Função Lagrangiana associada ao problema (1) é

$$L(w, \lambda) = F(w) - \sum_{n=1}^{N} \lambda_n c_n(w)$$

onde λ_n é designado por multiplicador de Lagrange associada à restrição $c_n(w) = 0$ $(n = 1, \ldots, N)$.

Notação:
$\lambda = (\lambda_1, \ldots, \lambda_N)^T$, é vetor dos multiplicadores de Lagrange (variáveis duais)

Condições de Karush-Kuhn-Tucker (KKT)

Teorema (Condição necessária de 1ª ordem)
Seja w^* um minimizante local do problema (1). Se w^* é um ponto regular das restrições então existe um vetor de multiplicadores de Lagrange λ^* tal que as seguintes condições são satisfeitas em (w^*, λ^*):

$$\nabla_w L(w^*, \lambda^*) = 0 \qquad \text{condição de 1ª ordem}$$

$$c_n(w^*) = 0 \qquad \text{admissibilidade } (n = 1, \ldots, N)$$

- $\nabla_w L(w^*, \lambda^*) = 0 \Leftrightarrow \nabla F(w^*) = \sum_{n=1}^{N} \lambda_n^* \nabla c_n(w^*)$
- (w^*, λ^*) é ponto estacionário da Lagrangiana e chama-se ponto KKT.
- encontrar pontos estacionários do problema (1) $\Leftrightarrow$ encontrar pontos estacionários da Lagrangiana

Teorema (Condições necessárias de 2ª ordem)
Seja w^* um minimizante local do problema (1) que é ponto regular. Seja λ^* o vetor de multiplicadores de Lagrange que verifica as condições KKT. Então

$$s^T \nabla^2_{w\,w} L(w^*, \lambda^*) s \geq 0, \quad \forall s \in \mathcal{C}$$

onde $\mathcal{C}(w^*, \lambda^*) = \{s \in \mathbb{R}^d : \nabla c_n(w^*)^T s = 0\}$.
(todos os vetores tangentes às superfícies das restrições)

Observação:

- A matriz hessiana da função Lagrangiana em ordem a w é

$$\nabla^2_{w\,w} L(w^*, \lambda^*) = \nabla^2 F(w^*) - \sum_{n=1}^{N} \lambda_n^* \nabla^2 c_n(w^*)$$

Teorema (Condições suficientes de 2ª ordem)
Suponha que para algum ponto admissível w^* do problema (1) existe um vetor de multiplicadores de Lagrange λ^* que verifica as condições KKT, e que

$$s^T \nabla^2_{w\,w} L(w^*, \lambda^*) s > 0, \quad s \in \mathcal{C}$$

onde $\mathcal{C}(w^*, \lambda^*) = \{s \in \mathbb{R}^d \setminus \{0\} : \nabla c_n(w^*)^T s = 0\}$.
Então w^* é **minimizante local estrito** do problema (1).

Observação:
$\mathcal{C}$ é o núcleo dos gradientes das restrições em w^*:

$$\mathcal{C}(w^*, \lambda^*) = null[\nabla c_n(w^*)^T]$$

Podemos então definir uma matriz Z em que as colunas são a base do espaço núcleo. Assim, as condições de 2ª ordem podem ser substituídas por:

$$Z^T \nabla^2_{w\,w} L(w^*, \lambda^*) Z \geq 0 \text{ é semi-definida positiva.}$$

$$Z^T \nabla^2_{w\,w} L(w^*, \lambda^*) Z > 0 \text{ é definida positiva.}$$

Problema com restrições de desigualdade

$$\begin{array}{ll} \underset{w \in \mathbb{R}^I}{\text{minimizar}} & F(w) \\ \text{sujeito a} & c_n(w) \geq 0, \quad n = 1, \ldots, N \end{array} \tag{2}$$

Ideia:

Seja $\overline{w} \in \mathcal{D}$ um ponto admissível. $\overline{w}$ não é ponto ótimo se existe um passo infinitesimal $\delta := \eta s$ $(\eta \in (0, \eta_\varepsilon))$ que mantém a admissibilidade e diminui o valor de F. δ diminui o valor de F se $\nabla F(\overline{w})^T \delta < 0$. Por outro lado, δ mantém a admissibilidade se

$$0 \leq c_n(\overline{w} + \delta) \approx c_n(\overline{w}) + \nabla c_n(\overline{w})^T \delta$$

A existência de tal δ significa que s satisfaz

$$c_n(\overline{w}) + \nabla c_n(\overline{w})^T s \geq 0 \ \text{ e } \ \nabla F(\overline{w})^T s < 0$$

Condição suficiente para direção admissível s (n=1,...,N):

- **Caso das restrições não ativas:** $c_n(\overline{w}) > 0$.

$$c_n(\overline{w}) + \nabla c_n(\overline{w})^T s \geq 0$$

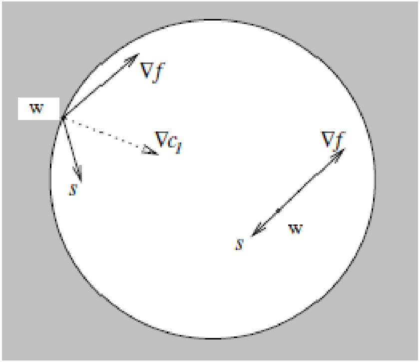
Notar que, s satisfaz a condição para $\eta > 0$ suficientemente pequeno.

- **Caso das restrições ativas:** $c_n(\overline{w}) = 0$.

$$\nabla c_n(\overline{w})^T s \geq 0$$

Interpretação geométrica das direções admissíveis

Exemplo: Problema com uma restrição de desigualdade.



(a) $w_1^2 + w_2^2 - 2 \leq 0$

As condições de otimalidade novamente podem ser estabelecidas em termos da **Função Lagrangiana**.

A Função Lagrangiana associada ao problema (2), é definida por

$$L(w, \alpha) = F(w) - \sum_{n=1}^{N} \alpha_n c_n(w)$$

onde α_n é o multiplicador de Lagrange associada à restrição $c_n(w) \geq 0$ $(n = 1, \dots, N)$.

Notação:
$\alpha = (\alpha_1, \dots, \alpha_N)^T$, é vetor dos multiplicadores de Lagrange (variáveis duais)

Condições de Karush-Kuhn-Tucker (KKT)

Teorema (Condição necessária de 1ª ordem)

Seja w^* um minimizante local do problema (2). Se w^* é um ponto regular das restrições então existe um vetor de multiplicadores de Lagrange α^* tal que as seguintes condições são satisfeitas em (w^*, α^*)

$$\nabla_w L(w^*, \alpha^*) = 0 \quad \text{condição 1ª ordem}$$

$$c_n(w^*) \geq 0 \quad \text{admissibilidade } (n = 1, \ldots, N)$$

$$\alpha_n^* \geq 0 \quad \text{admissibilidade dual } (n = 1, \ldots, N)$$

$$\alpha_n^* c_n(w^*) = 0 \quad \text{complementaridade } (n = 1, \ldots, N)$$

- $\nabla_w L(w^*, \alpha^*) = 0 \Leftrightarrow \nabla F(w^*) = \sum_{n=1}^{N} \alpha_n^* \nabla c_n(w^*)$

A condição

$$\alpha_n^* c_n(w^*) = 0 \ (n = 1, \dots, N)$$

, chama-se **condição de complementaridade** e significa que:

- ou a restrição n está **ativa** na solução ($c_n(w^*) = 0$)
- ou o multiplicador α_n^* que lhe está associado é **nulo** ($\alpha_n^* = 0$)

- Se os multiplicadores que correspondem a restrições **ativas** são **todos positivos**, a **complementaridade** diz-se **estrita.**
- Qualquer restrição que **não esteja ativa** em w^* ($c_n(w^*) > 0$) tem multiplicador nulo.

- Se o multiplicador que corresponde a uma restrição ativa é positivo, a restrição diz-se **não degenerada.**
- Se um multiplicador que corresponde a uma restrição ativa é nulo, a restrição diz-se **degenerada**.

Teorema (Condições necessárias de 2ª ordem)
Seja w^* um minimizante local do problema (2) que é ponto regular. Seja α^* o vetor de multiplicadores de Lagrange que verifica as condições KKT e a condição de **complementaridade estrita**. Então

$$s^T \nabla^2_{w\,w} L(w^*, \lambda^*) s \geq 0, \quad \forall s \in \mathcal{C}$$

onde $\mathcal{C}(w^*, \lambda^*) = \{s \in \mathbb{R}^d : \nabla c_n(w^*)^T s = 0, \forall n \in \mathcal{A}(w^*) \text{ com } \alpha_n^* > 0\}$.
(vetores tangentes às superfícies das **restrições ativas e não degeneradas**)

Observação:

- A matriz hessiana da função Lagrangiana em ordem a w é

$$\nabla^2_{w\,w} L(w^*, \alpha^*) = \nabla^2 F(w^*) - \sum_{n=1}^{N} \alpha_n^* \nabla^2 c_n(w^*)$$

Teorema (Condições suficiente de 2ª ordem)
Suponha que para algum ponto admissível w^* do problema (2) existe um vetor de multiplicadores de Lagrange α^* que verifica as condições KKT e a condição de **complementaridade estrita**, e que

$$s^T \nabla^2_{w\,w} L(w^*, \lambda^*) s > 0, \quad \forall s \in \mathcal{C}$$

onde
$\mathcal{C}(w^*, \lambda^*) = \{s \in \mathbb{R}^d \setminus \{0\} : \nabla c_n(w^*)^T s = 0, \forall n \in \mathcal{A}(w^*) \text{ com } \alpha_n^* > 0\}$.
Então w^* é **minimizante local estrito** do problema (1).

Observação: $\mathcal{C}$ é o núcleo dos gradientes das restrições ativas em w^* e não degeneradas:

$$\mathcal{C}(w^*, \lambda^*) = null[\nabla c_n(w^*)^T]$$

Podemos então definir uma matriz Z em que as colunas são a base do espaço núcleo. Assim, as condições de 2ª ordem podem ser substituídas por:

$$Z^T \nabla^2_{w\,w} L(w^*, \lambda^*) Z \geq 0 \text{ é semi-definida positiva.}$$

$$Z^T \nabla^2_{w\,w} L(w^*, \lambda^*) Z > 0 \text{ é definida positiva.}$$

Exemplo:

$$\begin{array}{ll} \underset{w\in\mathbb{R}^2}{\text{minimizar}} & F(w) = w_1 \\ \text{sujeito a} & c_1(w) = (w_1+1)^2 + w_2^2 \geq 1 \\ & c_2(w) = w_1^2 + w_2^2 \leq 2 \end{array}$$

Verifique se os pontos $\overline{w}^1 = (0,0)^T$, $\overline{w}^2 = (-1,-1)^T$, $\overline{w}^3 = (0,\sqrt{2})^T$ satisfazem as condições de otimalidade de 1ªordem.

Resolução:

- Neste problema $(d = 2, \mathcal{E} = \{\}, \mathcal{I} = \{2\})$ temos

$$\begin{array}{l} F(w_1,w_2) = w_1 \\ c_1(w_1,w_2) = (w_1+1)^2 + w_2^2 - 1 \geq 0 \\ c_2(w_1,w_2) = 2 - w_1^2 - w_2^2 \geq 0 \end{array}$$

  pelo que a Função Lagrangiana é dada por:

$$L(w,\alpha) = w_1 - \alpha_1((w_1+1)^2 + w_2^2 - 1) - \alpha_2(2 - w_1^2 - w_2^2)$$

- Calcular:

$$\nabla_w L(w_1,w_2,\alpha) = \begin{bmatrix} \dfrac{\partial L}{\partial w_1} \\ \dfrac{\partial L}{\partial w_2} \end{bmatrix} = \begin{bmatrix} 1 - 2\alpha_1(w_1+1) + 2\alpha_2 w_1 \\ -2\alpha_1 w_2 + 2\alpha_2 w_2 \end{bmatrix}$$

- $$\nabla c_1(w_1, w_2) = \begin{bmatrix} \dfrac{\partial c_1}{\partial w_1} \\ \dfrac{\partial c_1}{\partial w_2} \end{bmatrix} = \begin{bmatrix} 2(w_1 + 1) \\ 2w_2 \end{bmatrix},$$
  $$\nabla c_2(w_1, w_2) = \begin{bmatrix} \dfrac{\partial c_2}{\partial w_1} \\ \dfrac{\partial c_2}{\partial w_2} \end{bmatrix} = \begin{bmatrix} -2w_1 \\ -2w_2 \end{bmatrix},$$
- Verificar se $\overline{w}^1 = (0,0)^T$ satisfaz as condições de otimalidade de 1ªordem:
  $c_1(0,0) = 0$ e $c_2(0,0) = 2 > 0$, então $\overline{w}^1$ é ponto admissível e c_1 está ativa.
  $\nabla c_1(0,0) = (2,0)^T \neq (0,0)$, então $\overline{w}^1$ é ponto regular.
  Como c_2 não está ativa, então $\alpha_2 = 0$.
  Resolvendo $\nabla_w L(\overline{w}^1, \alpha) = 0$ tem-se que: $\left\{ \begin{array}{rl} 1 - 2\alpha_1 & = 0 \\ 0 & = 0 \end{array} \right. \Rightarrow \alpha_1 = \frac{1}{2} \geq 0.$
  Portanto $(0, 0, \frac{1}{2}, 0)$ satisfaz as condições de otimalidade de 1ªordem.

- Verificar se $\overline{w}^2 = (-1,-1)$ satisfaz as condições de otimalidade de 1ªordem:
  $c_1(-1,-1) = 0$ e $c_2(-1,-1) = 0$, então $\overline{w}^2$ é ponto admissível e ambas as restrições estão ativas.
  $\nabla c_1(-1,-1) = (0,-2)^T$ e $\nabla c_2(-1,-1) = (2,2)^T$ são linearmente independentes, então $\overline{w}^2$ é ponto regular.
  Resolvendo $\nabla_w L(\overline{w}^2, \alpha) = 0$ tem-se que: $\left\{ \begin{array}{rl} 1 - 2\alpha_2 & = 0 \\ 2\alpha_1 - 2\alpha_2 & = 0 \end{array} \right.$
  $\Rightarrow \alpha_1 = \alpha_2 = \frac{1}{2} \geq 0.$
  Portanto $(-1,-1,\frac{1}{2},\frac{1}{2})$ satisfaz as condições de otimalidade de 1ªordem.

- Verificar se $\overline{w}^3 = (0, \sqrt{2})$ satisfaz as condições de otimalidade de 1ªordem:
  $c_1(0, \sqrt{2}) = 2 > 0$ e $c_2(0, 0) = 0$, então $\overline{w}^3$ é ponto admissível e c_2 está ativa.
  $\nabla c_2(0, \sqrt{2}) = (0, -2\sqrt{2})^T \neq (0, 0)$, então $\overline{w}^3$ é ponto regular.
  Como c_1 não está ativa, então $\alpha_1 = 0$.
  Resolvendo $\nabla_w L(\overline{w}^3, \alpha) = 0$ vem: $\left\{ \begin{array}{cc} 1 & = 0 \\ 2\alpha_2\sqrt{2} & = 0 \end{array} \right. \Rightarrow$ Sistema impossível.
  Portanto, $\overline{w}^3$ não satisfaz as condições de otimalidade de 1ªordem.

Dualidade

Ideias gerais:

- A teoria da dualidade mostra como podemos construir um problema alternativo (problema dual) a partir do problema de otimização original (problema primal).
- Em alguns casos, o problema dual é computacionalmente mais fácil de resolver do que o problema primal.
- Noutros casos, o problema dual pode ser usado para obter facilmente um limite inferior para o valor ótimo F^* da função objetivo do problema primal.
- A dualidade tem também sido usada para desenvolver algoritmos para resolver o problema primal. Como por exemplo, o método da Lagrangiana aumentada.

Função dual de Lagrange

Para simplificar a exposição, considera-se o caso especial do problema de otimização com restrições de desigualdade:

$$\begin{array}{ll} \underset{w \in \mathbb{R}^d}{\text{minimizar}} & F(w) \\ \text{sujeito a} & c_n(w) \geq 0, \quad n = 1, \ldots, N \end{array} \qquad (\mathrm{P}_{primal})$$

A Função Lagrangiana associada a este problema é

$$L(w, \alpha) = F(w) - \sum_{n=1}^{N} \alpha_n c_n(w)$$

$\triangleright$ $\alpha = (\alpha_1, \ldots, \alpha_N)^T$ é vetor dos multiplicadores de Lagrange associadas às restrições $c_n(w) \geq 0$.

Função dual de Lagrange

A função dual $F_D : \mathbb{R}^N \to \mathbb{R}$ é definida pelo ínfimo (valor mínimo) da função Lagrangiana sobre w: para $\alpha \in \mathbb{R}^N$

$$F_D(\alpha) = \inf_{w \in \mathbb{R}^d} L(w, \alpha) \qquad (\mathrm{F}_{dual})$$

- Se a função Lagrangeana é ilimitada inferiormente em w, para alguns valores de α, então a função dual toma o valor $-\infty$.
- Considera-se para domínio de F_D o conjunto dos valores de $\alpha \in \mathbb{R}^N$ para os quais F_D é finita, ou seja,

$$dom\, F_D = \{\alpha \in \mathbb{R}^N : F_D(\alpha) > -\infty\}$$

- A função dual produz limites inferiores no valor ótimo F^* do problema primal (P_{primal}). Para qualquer $\alpha \geq 0$, tem-se que

$$F_D(\alpha) \leq F^*.$$

Problema dual

O problema dual para o problema (P_{primal}) é definido da forma:

$$\begin{array}{rl} \underset{\alpha \in \mathbb{R}^N}{\text{maximizar}} & F_D \equiv \inf_{w \in \mathbb{R}^d} L(w, \alpha) \\ \text{sujeito a} & \alpha \geq 0 \end{array}$$

- Calcular o ínfimo de (F_{dual}) implica encontrar o minimizante global da função $L(., \alpha)$ para o α dado, o que pode ser extremamente difícil na prática.
- **Porém,** quando F e $-c_n$ são funções convexas e $\alpha \geq 0$ (o caso em que estamos interessados), a função Lagrangiana $L(., \alpha)$ é também convexa. Neste caso, todos minimizantes locais são minimizantes globais, e o problema dual pode ser reescrito na forma:

Definição: (Problema dual)

$$\begin{array}{rl} \underset{w \in \mathbb{R}^d, \alpha \in \mathbb{R}^N}{\text{maximizar}} & L(w, \alpha) \\ \text{sujeito a} & \nabla_w L(w, \alpha) = 0 \\ & \alpha \geq 0 \end{array} \qquad (P_{dual})$$

Exemplo:

Definir o problema dual para o seguinte problema

$$\begin{array}{ll} \underset{w \in \mathbb{R}^2}{\text{minimizar}} & 0.5(w_1^2 + w_2^2) \\ \text{sujeito a} & w_1 - 1 \geq 0. \end{array}$$

Resolução: Como o problema é convexo, a função Lagrangiana associada ao problema é convexa:

$$L(w, \alpha_1) = 0.5(w_1^2 + w_2^2) - \alpha_1(w_1 - 1),$$

e o problema dual tem a seguinte forma:

$$\begin{array}{ll} \underset{w \in \mathbb{R}^2, \alpha_1 \in \mathbb{R}}{\text{maximizar}} & L(w, \alpha_1) \equiv 0.5(w_1^2 + w_2^2) - \alpha_1(w_1 - 1) \\ \text{sujeito a} & \nabla_w L(w, \alpha_1) = 0 \\ & \alpha \geq 0 \end{array}$$

Pela condição de 1ª ordem: $\nabla_w L(w, \alpha_1) = 0$ tem-se:

- $\nabla_{w_1} L = 0 \Leftrightarrow w_1 - \alpha_1 = 0 \Leftrightarrow w_1 = \alpha_1$
- $\nabla_{w_2} L = 0 \Leftrightarrow w_2 = 0$

Substituindo na função Lagrangiana:

$$L(w, \alpha_1) = 0.5(w_1^2 + w_2^2) - \alpha_1(w_1 - 1) = 0.5(\alpha_1^2 + 0) - \alpha_1(\alpha_1 - 1) = -0.5\alpha_1^2 + \alpha_1.$$

Portanto, o problema dual é:

$$\begin{array}{rl} \underset{\alpha_1 \in \mathbb{R}}{\text{maximizar}} & -0.5\alpha_1^2 + \alpha_1 \\ \text{sujeito a} & \alpha_1 \geq 0 \end{array}$$

o qual tem a solução $\alpha_1^* = 1$.

Teorema: (Dual de Wolfe)

Se (w^*, α^*) é um par solução do problema (P_{primal}), se F e $-c_n$, $n = 1, \ldots, N$, são funções convexas e continuamente diferenciáveis, e se w^* é ponto regular, então (w^*, α^*) é solução do problema dual:

$$\begin{array}{rr} \underset{w \in \mathbb{R}^d, \alpha \in \mathbb{R}^N}{\text{maximizar}} & L(w, \alpha) \\ \text{sujeito a} & \nabla_w L(w, \alpha) = 0 \\ & \alpha \geq 0. \end{array}$$

Além disso $F(w^*) = L(w^*, \alpha^*)$.

Bibliografia

- Jorge Nocedal and Stephen Wright. **Numerical Optimization**, Second Edition, Springer 2006