Otimização Não Linear
Introdução

Fernanda Costa

Departamento de Matemática
Universidade do Minho

Outline

Introdução

Otimização

- Área da Matemática que se dedica ao estudo de problemas de minimização ou maximização de uma função de várias variáveis, dentro de um dado conjunto admissível.
- Os problemas de otimização surgem quando se pretende encontrar a melhor solução de um conjunto de soluções alternativas possíveis.

Os problemas de otimização surgem em diversas áreas:

- Ciências: determinação de parâmetros ótimos em modelos estatísticos, físicos ou químicos; ...
- Engenharia: dimensionamento ótimo de estruturas (pontes, edifícios, aeronaves) para minimizar o peso mantendo a resistência; ...

- Finanças: estratégias de investimento que maximizem o retorno esperado para um dado nível de risco; ...
- Medicina: planeamento ótimo de tratamentos de radioterapia (dose, tempo e localização); ...
- Economia: definição de políticas fiscais ótimas; ...
- Big-data: machine learning ... problemas de classificação; reconhecimento de padrões e imagens; treino de redes neuronais; otimização de hiperparâmetros; ...

Formulação matemática

Os ingredientes do problema de otimização são:

- **variáveis de decisão** (também designadas por incógnitas ou parâmetros), $w \in \mathbb{R}^d$;
- **restrições** - definem valores aceitáveis para as variáveis w;
- **função objetivo** (ou medida de desempenho) - que mede a qualidade das potenciais soluções.

Os problemas de otimização podem ser classificados em dois grandes grupos:

- Problemas de otimização sem restrições.
- Problemas de otimização com restrições.

Problema de otimização sem restrições

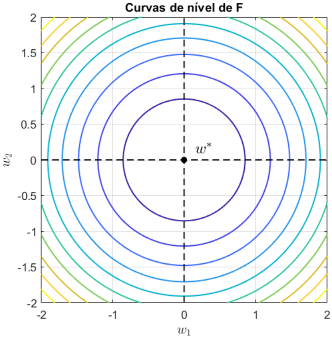
$$\underset{w \in \mathbb{R}^d}{\text{minimizar}}\ F(w) \qquad (P_{SR})$$

- $w = (w_1, w_2, \ldots, w_d)$ são as variáveis;
- $F : \mathbb{R}^d \to \mathbb{R}$ é a função objetivo (medida de desempenho).
  (loss or cost function in ML)

Exemplo:

- $\underset{w_1, w_2}{\text{minimizar}}\ w_1^2 + w_2^2$

  neste problema: $w = (w_1, w_2)$ e
  $F(w) = w_1^2 + w_2^2$



Recordar: Chama-se curva de nível de F com valor k, ao conjunto de pontos $w \in \text{dom } F$ cuja imagem é k. Formalmente, é o conjunto dado por $CN_k = \{w \in \text{dom} F : F(w) = k\}$

Problema de otimização com restrições (na sua forma mais geral)

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

- $c_n : \mathbb{R}^d \to \mathbb{R}$ 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 (ponto admissível)

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

Definição (conjunto admissível)

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

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

Exemplo:

$$\begin{array}{ll} \underset{w_1, w_2}{\text{minimizar}} & (w_1 - 1)^2 + (w_2 - 1)^2 \\ \text{sujeito a} & w_1^2 - w_2 \leq 0 \\ & w_1 + w_2 \leq 2 \end{array}$$

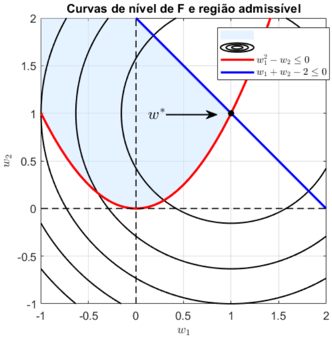


Fig. Representação geométrica do problema.

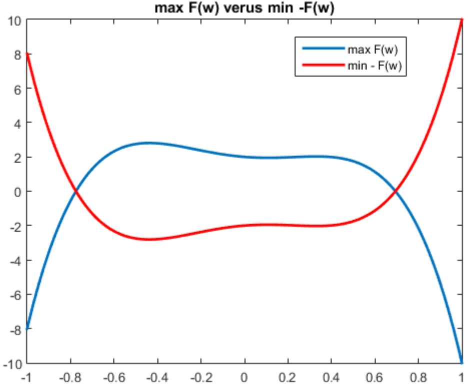
Podemos escrever este problema na forma (P_{CR}), fazendo:

$$\begin{array}{ll} \underset{w=(w_1, w_2)}{\text{minimizar}} & F(w) = (w_1 - 1)^2 + (w_2 - 1)^2 \\ \text{sujeito a} & c_1(w) = -w_1^2 + w_2 \geq 0 \\ & c_2(w) = -w_1 - w_2 + 2 \geq 0 \end{array}$$

onde $\mathcal{I} = \{1, 2\}$ e $\mathcal{E} = \{\}$.

Maximização versus Minimização

- Qualquer problema de *maximização* pode ser reformulado como um problema de minimização: maximizar $F(w) = -$minimizar $-F(w)$.



$\triangleright$ o ponto w^* onde F atinge o seu máximo $F(w^*)$ é o mesmo onde $-F$ atinge o seu mínimo $-F(w^*)$.

Minimizantes locais, globais e estritos

Considere o **problema de otimização sem restrições** (P_{SR}), e $w^* \in \mathcal{D} = \mathbb{R}^d$.

- w^* é um **minimizante global** se

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

- w^* é um **minimizante global estrito** se

$$F(w^*) < F(w), \ \ \forall w \in \mathcal{D}, \ w \neq w^*;$$

- w^* é um **minimizante local** se existir $\epsilon > 0$:

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

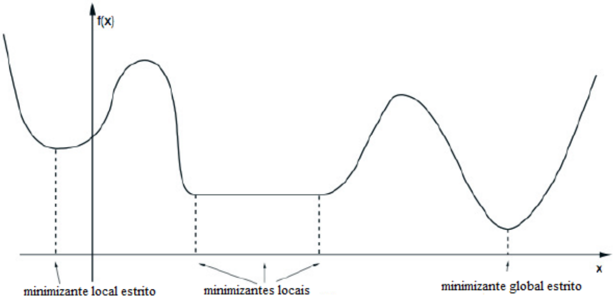
- w^* é um **minimizante local estrito** se existir $\epsilon > 0$:

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

Nota: Todo o **minimizante global** é um **minimizante local.**

Recordar: $B(w^*; \epsilon)$ é a bola centrada em w^* de raio ϵ: $B(w^*; \epsilon) := \{w \in \mathbb{R}^d : \|w^* - w\| \leq \epsilon\}$.

Exemplo:



obs: Não confundir minimizante com mínimo. Se w^* é um minimizante de F então $F(w^*)$ é o respetivo mínimo: o minimizante é um vetor em $\mathbb{R}^d$ que minimiza a função F; o mínimo será $F(w^*)$ (um valor real).

Para o **problema de otimização com restrições** (P_{CR}), onde $w^* \in \mathcal{D} \subset \mathbb{R}^d$, modifica-se ligeiramente a terminologia para usar a palavra "solução" (ou "solução ótima") em vez de "minimizante".

- w^* é uma **solução global** se

$$F(w^*) \leq F(w), \quad \forall\, w \in \mathcal{D};$$

- w^* é uma **solução global estrita** se

$$F(w^*) < F(w), \quad \forall\, w \in \mathcal{D},\, w \neq w^*;$$

- w^* é uma **solução local** se existe $\varepsilon > 0$ tal que

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

- w^* é uma **solução local estrita** se existe $\varepsilon > 0$ tal que

$$F(w^*) < F(w), \quad \forall\, w \in B(w^*, \varepsilon) \cap \mathcal{D},\, w \neq w^*.$$

Nota: Todo a **solução ótima global** é uma **solução ótima local.**

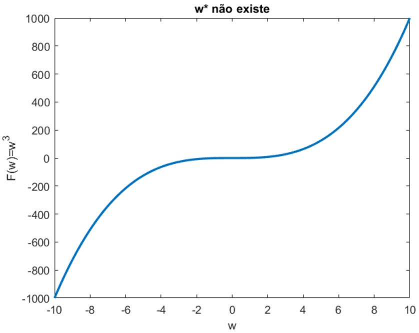
Se $w^* \in \mathcal{D}$ é uma "solução ótima" do problema com restrições então $F(w^*)$ é designado por valor ótimo (um valor real).

Existência de minimizantes

Os exemplos seguintes ilustram alguns dos ingredientes fundamentais para garantir a existência de minimizantes:

Exemplo:

$$\underset{w\in\mathbb{R}}{\text{minimizar}} \quad F(w) = w^3$$

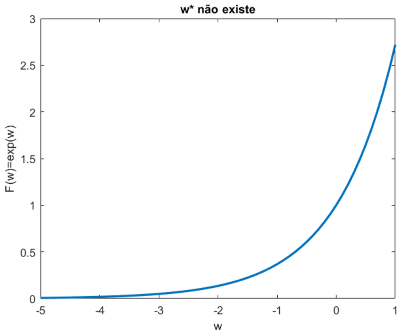


- $F(w)$ não tem minimizante nem mínimo, pois $F(w)$ é estritamente decrescente e $F(w)$ é ilimitada inferiormente ($F(w) \to -\infty$).

$\Rightarrow$ **F tem de ser limitada inferiormente**.

Exemplo:

$$\underset{w \in \mathbb{R}}{\text{minimizar}} \quad F(w) = e^w$$

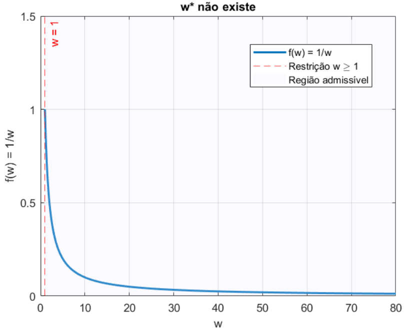


- $F(w)$ não tem minimizante nem mínimo, pois $F(w)$ é estritamente decrescente. $F(w) \to 0$ mas nunca atinge o valor 0 ($\inf F(w) = 0$). A região admissível $\mathbb{R}$ não é limitada inferiormente.

$\Rightarrow$ **A região admissível tem de ser limitada**.

Exemplo:

$$\begin{array}{ll} \underset{w \in \mathbb{R}}{\text{minimizar}} & \dfrac{1}{w} \\ \text{sujeito a} & w \geq 1 \end{array}$$



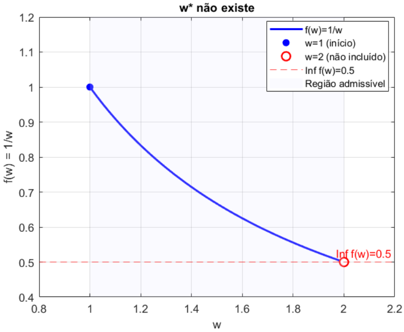
- Este problema não tem minimizante nem mínimo, pois $F(w)$ é estritamente decrescente e a região admissível $\mathcal{D} = \{w \in \mathbb{R} : w \geq 1\}$ não é limitada superiormente.

  ($F(w) \to 0$, mas nunca atinge 0; $\inf F(w) = 0$).

$\Rightarrow$ **A região admissível $\mathcal{D}$ tem de ser limitada**.

Exemplo:

$$\begin{array}{ll} \underset{w \in \mathbb{R}}{\text{minimizar}} & \dfrac{1}{w} \\ \text{sujeito a} & 1 \leq w < 2 \end{array}$$



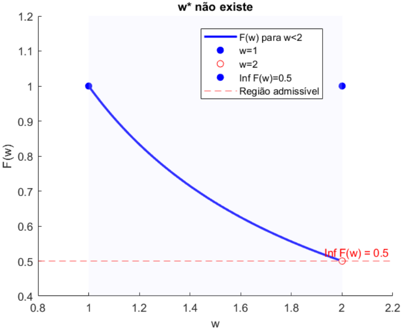
- Este problema não tem minimizante nem mínimo, pois $F(w)$ é estritamente decrescente e o ponto 2 não é admissível. ($F(w)$ nunca atinge o valor 0.5; $\inf F(w) = 0.5$).

$\Rightarrow$ **A região admissível $\mathcal{D}$ tem de ser limitada e fechada.**

Exemplo:

$$\begin{array}{ll} \underset{w\in\mathbb{R}}{\text{minimizar}} & F(w) \\ \text{sujeito a} & 1 \leq w \leq 2 \end{array}$$

onde $F(w) = \begin{cases} \dfrac{1}{w}, & \text{se } w < 2, \\ 1, & \text{se } w = 2. \end{cases}$



- Este problema não tem minimizante nem mínimo devido à descontinuidade de F. $F(w)$ é estritamente decrescente e $F(w) \to 0.5$, mas nunca atinge 0.5 ($\inf F(w) = 0.5$).

$\Rightarrow$ **A região admissível $\mathcal{D}$ tem de ser limitada e fechada, e F suficientemente suave.**

Recordamos agora duas formulações do Teorema de Weierstrass — um para sucessões e outro para funções. O Teorema 2 sintetiza os ingredientes identificados nos exemplos anteriores como sendo importantes para garantir a existência de um minimizante.

Teorema 1 (Teorema de Weierstrass para sucessões)

Seja $\{w_k\}$, $k \to \infty$, uma sucessão de pontos num conjunto compacto $\mathcal{D}$ (i.e., num conjunto fechado e limitado). Então existe uma subsucessão $\{w_{k_j}\}$ de $\{w_k\}$ que converge para um ponto em $\mathcal{D}$.

Teorema 2 (Teorema de Weierstrass para funções)

Seja F uma função contínua de valor real, definida num conjunto compacto $\mathcal{D} \subset \mathbb{R}^d$ e não vazio. Então $\mathcal{D}$ contém um minimizante (global) de F no conjunto $\mathcal{D}$.

Obs: As hipóteses do Teorema 2 podem ser relaxadas, por exemplo:

- que o conjunto $\{w \in \mathcal{D} : F(w) \leq F(u)\}$ é compacto para algum $u \in \mathcal{D}$ (em vez de assumirmos que D é compacto);
- que F é semi-contínua inferior, i.e., para toda a constante c, o conjunto $\{w \in \mathcal{D} : F(w) \leq F(c)\}$ é fechado (em vez de se assumirmos que F é contínua).

Classificação dos problemas de otimização

Classificação pelo tipo de restrições:

- **Problemas sem restrições:** todos os pontos $w \in \mathbb{R}^d$ são admissíveis

$$\underset{w \in \mathbb{R}^d}{\text{minimizar}} \; F(w)$$

Caso especial importante

- **Problemas de mínimos quadrados** (*Least-Squares Problems*):

$$\underset{w \in \mathbb{R}^d}{\text{minimizar}} \; F(w) := \sum_{n=1}^{N} (f_n(w))^2$$

No contexto de "ajuste de dados", ou seja, "regressão", cada função $f_n(w)$ define um resíduo/erro.

- Regressão: dados N pares de pontos $\{(x^n, y^n)\}_{n=1}^N$, o vetor dos parâmetros $w \in \mathbb{R}^d$ que melhor ajustam a função $\phi(w;x)$ (modelo matemático) aos dados, é determinado pelo problema de mínimos quadrados:

$$\underset{w\in\mathbb{R}^d}{\text{minimizar}}\ F(w) = \sum_{n=1}^{N} \underbrace{\left(\phi(w;x^n) - y^n\right)^2}_{f_n(w)}$$

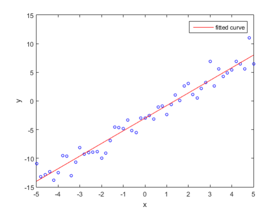
  - $\widetilde{y}^{\,n} = \Phi(w;x^n)$ é o valor previsto pelo modelo;
  - y^n - é o valor conhecido associado a x^n ;

- É um **problema de mínimos quadrados linear** se ϕ é linear nas variáveis $w = (w_1, w_2, \ldots, w_d)$:

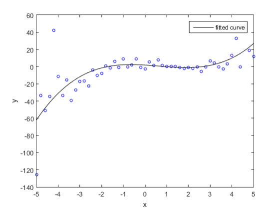
$$\phi(w;x^n) = w_1 g_1(x^n) + w_2 g_2(x^n) + \cdots + w_d g_d(x^n)$$

onde cada função g_i depende apenas dos dados x^n. Caso contrário, é um **problema mínimos quadrados não-linear**.

Exemplo:



$\phi(w;x) = w_1 + w_2 x$



$\phi(w;x) = w_1 + w_2 x + w_3 x^2 + w_4 x^3$

Classificação dos problemas de otimização

Classificação pelo tipo de restrições:

- **Problemas com restrições de limites:** apenas existem restrições de limites simples nas variáveis

$$\begin{array}{ll} \underset{w \in \mathbb{R}^d}{\text{minimizar}} & F(w) \\ \text{sujeito a} & l \leq w \leq u \end{array}$$

  onde $l, u \in \mathbb{R}^d$ são os vetores dos limites inferiores e superiores de w; podendo alguns elementos ser ∞.

- **Problemas com restrições lineares:** A função objetivo é não linear e as restrições são lineares.

$$\begin{array}{rl} \underset{w \in \mathbb{R}^d}{\text{minimizar}} & F(w) \\ \text{sujeito a} & A_1 w = b_1 \\ & A_2 w \geq b_2 \end{array}$$

onde $A_1 \in \mathbb{R}^{m_1 \times d}$ e $A_2 \in \mathbb{R}^{m_2 \times d}$ são as matrizes dos coeficientes das restrições; e $b_1 \in \mathbb{R}^{m_1}$, $b_2 \in \mathbb{R}^{m_2}$ são os vetores dos termos independentes.

Casos especiais importantes:

- **Programação Linear (LP):** A função objetivo é linear e as restrição são lineares.

$$\begin{array}{rl} \underset{w \in \mathbb{R}^d}{\text{minimizar}} & F(w) = c^T w \\ \text{sujeito a} & A_1 w = b_1 \\ & A_2 w \geq b_2 \end{array}$$

onde $c \in \mathbb{R}^d$ é o vetor dos coeficiente da F.

- **Programação Quadrática (QP):** A função objetivo é quadrática e as restrição são lineares.

$$\begin{array}{rl} \underset{w \in \mathbb{R}^d}{\text{minimizar}} & F(w) = w^T Q w + c^T w + a \\ \text{sujeito a} & A_1 w = b_1 \\ & A_2 w \geq b_2 \end{array}$$

onde $Q \in \mathbb{R}^{d \times d}$, $c \in \mathbb{R}^d$, $a \in \mathbb{R}$ são os coeficientes de F.

- **Problemas com restrições de igualdade:** todas as restrições do problema são restrições de igualdade

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

Caso especial: problemas com restrições lineares de igualdade:

$$\begin{array}{rl} \underset{w\in\mathbb{R}^d}{\text{minimizar}} & F(w) \\ \text{sujeito a} & Aw = b \end{array}$$

- **Problemas com restrições não lineares:** pelo menos uma das funções (função objetivo ou das restrições) do problema (P_{CR}) é não linear.

Classificação pelo tipo de variáveis:

- **Variáveis contínuas:** $w \in \mathbb{R}^d$

  $\Rightarrow$ **Problemas de otimização contínua** — classe de problemas estudada neste curso.

- **Variáveis inteiras:** $w \in \mathcal{D} = \mathbb{Z}^d$

  $\Rightarrow$ **Problemas de programação inteira (PI)** (*integer programming problems*).

- **Variáveis mistas:** existem variáveis inteiras (ou binárias) e variáveis contínuas.

  $\Rightarrow$ **Problemas de programação inteira-mista (PIM)** (*mixed-integer programming problems*).

$\Rightarrow$ Os problemas de **PI** e **PIM** pertencem à classe de **problemas de otimização discreta**. Esses problemas podem envolver não apenas variáveis inteiras ou binárias, mas também variáveis discretas mais gerais, por exemplo: $\mathcal{D} = \left\{\frac{1}{4}, \frac{1}{2}, 2, 4, \ldots\right\}$.