Introdução aos Problemas de Otimização
Condições de Otimalidade

Departamento de Matemática
Universidade do Minho

Outline

Otimização - processo para encontrar a melhor solução para um problema, de um conjunto de soluções alternativas.

Os **problemas de otimização** surgem em diversas áreas:

- ciências,
- engenharias
- finanças
- medicina
- economia
- big-data
- machine learning
- entre outas

Problema de otimização

Os problemas de programação matemática podem ser escritos, na sua forma mais geral, na seguinte forma:

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

- $w = (w_1, w_2, \ldots, w_d)^T$ é o vetor das variáveis de decisão
- $F : \mathbb{R}^d \to \mathbb{R}$ é a função objetivo (escalar) de w que se pretende minimizar.
- $c_n : \mathbb{R}^d \to \mathbb{R}$ são as funções de restrição de igualdade e desigualdade

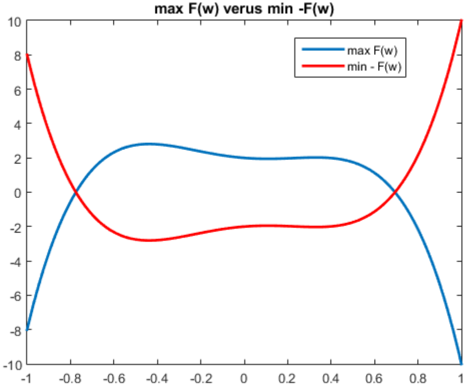
▷ Um ponto que verifique todas as restrições, chama-se **ponto admissível** do problema.

▷ Ao conjunto de todos os pontos admissíveis, chama-se **conjunto admissível**:

$$\mathcal{D} = \{w \in \mathbb{R}^d : c_n(w) = 0,\ n = 1, ..., j;\ c_n(w) \geq 0,\ n = j+1, \ldots, N\}$$

Maximização verus minimização

- Qualquer problema de *maximização* pode ser reformulado como um problema de minimização pois $\max F(w)$ é igual a $-\min -F(w)$.



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

Exemplos de problemas de otimização

Encontrar boas (ou melhores) ações

- w representa alguma ação, por exemplo:
  - alocação de recursos
  - trajeto escolhido para uma viagem
  - compra de ações
- restrições limitam ações ou impõem condições ao resultado
- quanto menor for a função objetiva, melhor é
  - custo total (ou lucro)
  - desvio do resultado desejado ao alvo
  - uso do combustível
  - o tempo da viagem
  - risco em finanças

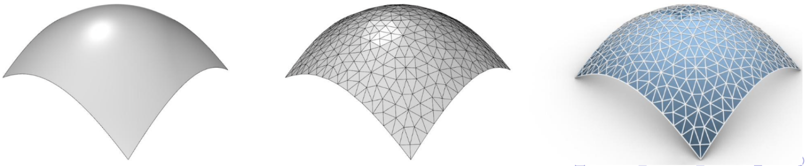
Engenharia de design

- w representa um projeto (de um circuito, dispositivo, estrutura, ...)
- restrições limitam ações ou impõem condições ao resultado
- restrições vêm de
  - processo de manufatura
  - requisitos de desempenho
- a função objetiva $F(w)$ é a combinação de custo, peso, potência,...

Engenharia de design

Exemplo:

- Está em estudo a construção uma cúpula de vidro esférica para cobrir um átrio central do edifício.
- A cúpula é composta por uma grelha triangular de perfis de suporte e painéis de vidro.
- Pretende-se otimizar o posicionamento dos nós da estrutura para reduzir o número de perfis de suporte metálicos e painéis de vidro diferentes para facilitar o processo de fabrico e reduzir os custos de construção sem sacrificar o desempenho estrutural da estrutura.



Exemplos de problemas de otimização

Encontrar bons modelos

- w representa os parâmetros de um modelo
- as restrições impõem requisitos aos parâmetros do modelo (por exemplo, a não negatividade)
- a função objetiva $F(w)$ é o erro de previsão com alguns dados observados

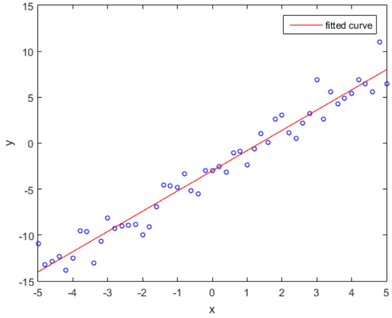
Exemplo: Dados N pares de pontos (x^n, y^n), o vetor dos parâmetros $w \in \mathbb{R}^d$ que ajustam melhor uma função (modelo) $\phi(w; x)$ aos dados, é dado por:

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

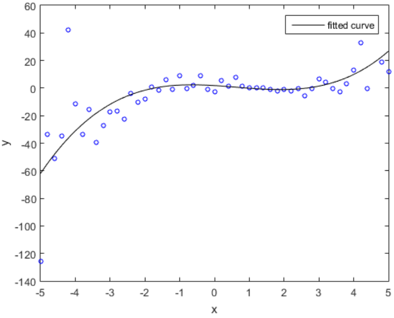
onde

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

▷ Exemplos de ajuste:



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



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

Exemplos de problemas de otimização

Análise do pior caso

- w variáveis são acções ou parâmetros fora do nosso controlo (e possivelmente sob o controlo de um adversário)
- restrições limitam os valores possíveis dos parâmetros ou ações
- minimizar $-F(w)$ estamos a encontrar os piores parâmetros/ ações
- Se o pior cenário possível é tolerável, então é bom saber que o pior cenário possível pode ocorrer

Modelos baseados em otimização

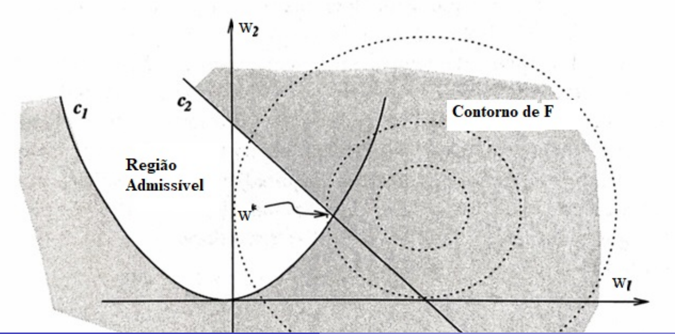
- modelam uma entidade realizando ações que permite resolver um problema de otimização
  - um organismo age para maximizar seu sucesso reprodutivo
  - as taxas de reação em uma célula maximizam o crescimento
  - correntes em um circuito elétrico minimizam a potência total

Exemplo de um problema académico de otimização

Minimizar o raio de uma circunferência centrada em $(2,1)$ sujeita às restrições $w_1^2 - w_2 \leq 0$ e $w_1 + w_2 \leq 2$.

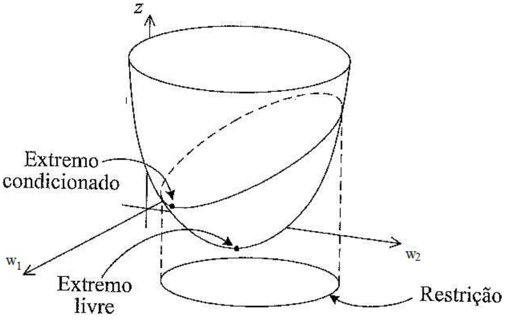
Podemos escrever o problema na forma (1) definindo

$$F(w) = (w_1 - 2)^2 + (w_2 - 1)^2, \quad w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$
$$c_1(x) = -w_1^2 + w_2 \geq 0$$
$$c_2(x) = -w_1 - w_2 + 2 \geq 0.$$

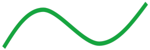


Diferentes tipos de problemas de otimização

- Alguns problemas têm restrições e outros não.



- Pode se ter uma única variável ou um conjunto de variáveis.
- As variáveis podem ser discretas (p.e. $x_i \in \mathbb{Z}, x_i \in 0, 1$) ou continuas.



- os problemas podem ser determinísticos ou estocásticos
  - nos problemas determinísticos os modelos são conhecidos.
  - quando alguns dos parâmetros do modelo determinístico são incertos e este modelo apresenta-se sensível a alterações destes parâmetros, então é apropriado considerar programação estocástica para solução desse problema.

O problema determinístico permite calcular a solução ótima para cada um dos cenários separadamente, enquanto que o problema estocástico considera o conjunto de todos os cenários simultaneamente, cada um com uma probabilidade de ocorrência associada.

Casos particulares de programação matemática

Definição 1

O problema de otimização (1) é **linear** se as funções F e c_n são lineares para todo $n \in \{1, \ldots, N\}$ em relação a w.

Nota: O problema (1) é um problema de programação não linear se pelo menos uma das restrições C_n ou a função objetivo F é não linear.

Definição 2

O problema de otimização (1) é **convexo** se a função F é convexa, c_n é lineares para todo $n \in \{1, \ldots, j\}$ e c_n é côncava para todo $n \in \{j+1, \ldots, N\}$ em relação a w.

Recordar: A funções linear $F(w) = a^T w$, $a \in \mathbb{R}^d$ é convexa e côncava. A função $F(w) = w^T Q w + a^T w + b$ com Q uma matriz simétrica $d \times d$ é convexa se e só se Q é semi-definida positiva;

Propriedade do Gradiente

Teorema 3

Seja $F : \mathbb{R}^d \to \mathbb{R}$ *diferenciável em* w^*. *Então, se* $\nabla F(w^*) \neq (0, 0)$, $\nabla F(w^*)$ *é um vetor perpendicular à curva de nível em* $F(w^*)$.

Teorema 4

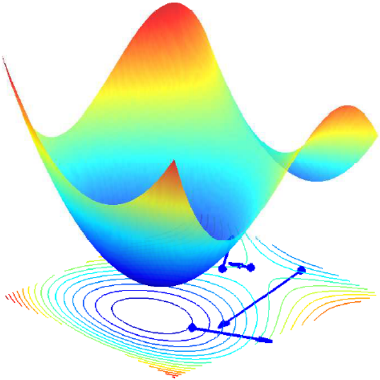
Se uma função $F : \mathbb{R}^d \to \mathbb{R}$ *é diferenciável em* w^*, *então existe derivada de* F *no ponto* w^*, *segundo qualquer direção de* $v \in \mathbb{R}^d$ *e tem-se*

$$D_v F(w^*) = \nabla F(w^*)^T v = \|\nabla F(w^*)\| \|v\| \cos \alpha,$$

onde α *é o ângulo formado por* v *e por* $\nabla F(w^*)$.

A derivada de F num dado ponto w^*, segundo a direção do vector v admite:

- Um valor máximo que ocorre quando v tem a direção e sentido do vetor $\nabla f(w^*)$.
- Um valor mínimo que ocorre quando v tem a mesma direção mas sentido contrário ao do vetor $\nabla f(w^*)$.
- O valor 0 quando v é perpendicular ao vetor $\nabla F(w^*)$.



Exercício :

Uma caldeira de um vulcão tem uma forma definida pelo gráfico de

$$F(w_1, w_2) = \frac{1}{5}w_1^2 + \frac{1}{5}w_2^2 - 400$$

a) Representa a superfície $z = F(w_1, w_2)$.

b) Representa as curvas de nível de F.

c) Mostre que o ponto $A = (30, 30, -40)$ pertence ao gráfico de F.

d) Se um explorador está no ponto A, que direção é que ele deve tomar para descer pela parte mais íngrema da caldeira.

e) Se o explorador está no ponto A e se mover na direção do vetor $v = (4, 3)$, ele está a subir ou a descer?

f) Se um explorador está no ponto A, em que direção é que ele se deve mover para percorrer um caminho plano?

Definição 5

Considere o problema de otimização (1), e $w^* \in \mathcal{D}$.

- w^* é um minimizante global e a sua imagem é um mínimo global se

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

- w^* é um minimizante global estrito e a sua imagem é um mínimo global estrito se

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

- w^* é um minimizante local e a sua imagem é um mínimo local se

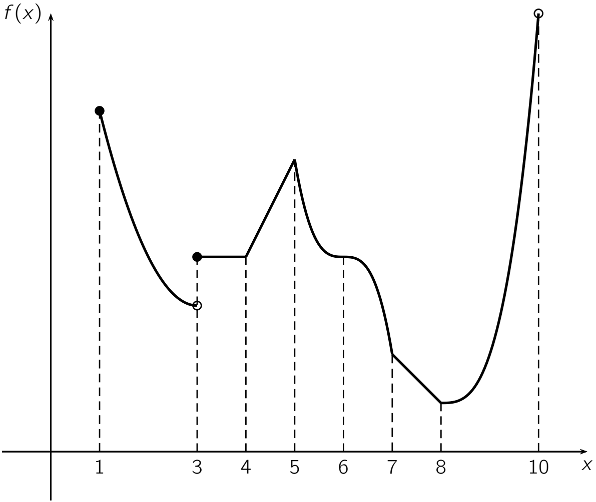
$$F(w^*) \leq F(w), \; \forall w \in B(w^*, \varepsilon) \cap \mathcal{D}$$

- w^*é um minimizante local estrito e a sua imagem é um mínimo local estrito se

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

Exercício: Considere a função $F : [1, 10[\subseteq \mathbb{R} \longrightarrow \mathbb{R}$ tal que

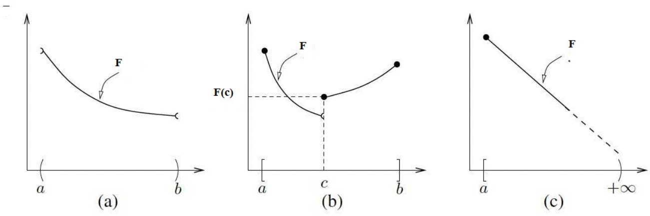
$$F(w) = \begin{cases} (w-3)^2+3 & \text{se } w \in [1,3[\\ 4 & \text{se } w \in [3,4[\\ 2w-4 & \text{se } w \in [4,5[\\ 2(6-w)^3+4 & \text{se } w \in [5,7[\\ 9-w & \text{se } w \in [7,8[\\ (w-8)^3+1 & \text{se } w \in [8,10[. \end{cases}$$



f(x)
1
3
4
5
6
7
8
10
x

(a) Indique os minimizantes globais de f.
(b) Indique o mínimo global de f.
(c) Indique os minimizantes globais estrito de f.
(d) Indique o mínimo global estrito de f.
(e) Indique os minimizantes locais de f.
(f) Indique os mínimos locais de f.
(g) Indique os minimizantes locais estrito de f.
(h) Indique os mínimos locais estrito de f.

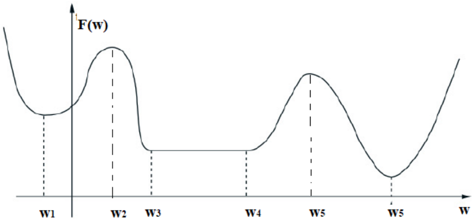
Condições para a existência de solução garantem que o problema tem pelo menos uma solução.



Teorema 6

Teorema de Weierstrass Uma função $F : \mathbb{R}^d \longrightarrow \mathbb{R}$, contínua definida num conjunto não vazio, limitado e fechado (compacto) tem máximo e mínimo nesse conjunto.

Condições Necessárias de Otimalidade



As condições necessárias de otimalidade identificam um conjunto de candidatos a minimizantes. Os minimizantes, caso existem, estão neste conjunto. Poderão também estar elementos que não são minimizantes.

As condições necessárias de otimalidade são do tipo:
Se w^* é um minimizante. Então w^* satisfaz as seguintes condições

Exemplo: Condições necessárias de otimalidade de 1º ordem
Se w^* é um minimizante. Então $F'(w^*) = 0$.

Quais são pontos que satisfazem condições necessárias de otimalidade de 1º ordem?

As condições necessárias de otimalidade são tanto mais interesantes quanto mais forte, i.e. quando mais reduzido for o conjunto de pontos que as satisfazem.

Exemplo: Condições necessárias de otimalidade de 2º ordem
Se w^* é um minimizante. Então $F'(w^*) = 0$ e $F''(w^*) \geq 0$.

Quais são pontos que satisfazem condições necessárias de otimalidade de 2º ordem?

Identifica o conjunto, mais reduzido, de candidatos

As condições suficientes de otimalidade são tipo:
Se w^* satisfaz as seguintes condições. Então w^* é minimizante.

Do conjunto de pontos que satisfaz as condições, todos são minimizantes.
Poderá haver minimizantes que não satisfazem as condições.

Exemplo: Se w^* satisfaz $F'(w^*) = 0$ e $F''(w^*) > 0$.
Então w^* é um minimizante local.

Quais são pontos que satisfazem condições suficientes de otimalidade?

Problema de otimização sem restrições:

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

Exemplo: A procura do vetor dos parâmetros que ajustam melhor um modelo. (Ver exemplos no secção "Encontrar bons modelos")

Nota: Nesta secção, assume-se que F é continuamente diferenciável até à 2ª ordem.

Definição 7

Um ponto w^* que satisfaça a condição $\nabla F(w^*) = 0$ é designado por ponto estacionário de F.

Teorema 8 (Condição necessária de 1ª ordem)

Se w^* é um minimizante local de F então $\nabla F(w^*) = 0$.

Demonstração.

Considere $z \in \mathbb{R}^d \setminus \{0\}$ arbitrário. Como w^* é um minimizante local, existe $\delta > 0$ tal que

$$F(w^*) \leq F(w^* + tz), \tag{2}$$

para todo $t \in (0, \delta)$. Pela expansão de Taylor,

$$F(w^* + \alpha z) = F(w^*) + t\nabla F(w^*)^T z + r(t),$$

com $\lim_{t\to 0} \frac{r(t)}{t} = 0$. Usando (2) e dividindo por t, obtemos

$$\nabla F(w^*)^T z + \frac{r(t)}{t} \geq 0.$$

Demonstração.

Passando o limite quando $t \to 0$, obtemos

$$\nabla F(w^*)^T z \geq 0.$$

Como z é arbitrário, se $\nabla F(w^*)$ não fosse nulo, poderíamos escolher $z = -\nabla F(w^*)$, resultando $\|\nabla F(w^*)\|^2 \leq 0$, o que é uma contradição. Logo, $\nabla F(w^*) = 0$. $\square$

Exercício: Seja $F : \mathbb{R}^3 \to \mathbb{R}$, dada por

$$F(w_1, w_2, w_3) = \sin(3w_1^2 + w_2^2) + \cos(w_1^2 - w_2^2) + 5w_3.$$

Verifique se F pode ter minimizantes em $\mathbb{R}^3$. E no conjunto $B = \{(w_1, w_2.w_3) \in \mathbb{R}^3 : w_1^2 + \frac{w_2^2}{4} + \frac{w_3^2}{9} \leq 1\}$ terá minimizante?.

Teorema 9 (Condição necessária de 2ª ordem para minimizante)

Se w^* *é um minimizante local de* F *então* $\nabla F(w^*) = 0$ *e* $\nabla^2 F(w^*)$ *é semi-definida positiva, isto é*

$$z^T \nabla^2 F(w^*) z \geq 0,$$

para todo $z \in \mathbb{R}^d$.

Demonstração.

Considere $z \in \mathbb{R}^d \setminus \{0\}$ arbitrário. Pela expansão de Taylor,

$$F(w^* + \alpha z) = F(w^*) + t \nabla F(w^*)^T z + \frac{t^2}{2} z^T \nabla F(w^*)^T z + r(t),$$

com $\lim_{t \to 0} \frac{r(t)}{t^2} = 0$. Como w^* é um minimizante local, pelo Teorema 8 garante que $\nabla F(w^*) = 0$. Portanto, para t suficientemente pequeno,

$$0 \leq F(w^* + \alpha z) - F(w^*) = \frac{t^2}{2} z^T \nabla F(w^*)^T z + r(t)$$

Demonstração.

Dividindo por t^2 e passando o limite quando $t \to 0$, obtemos $z^T \nabla F(w^*)^T z \geq 0$. $\square$

Exercício: Seja $F : \mathbb{R}^2 \to \mathbb{R}$, dada por

$$F(w_1, w_2) = (w_1 - w_2^2)(w_1 - \frac{1}{2} w_2^2).$$

Verifique que $(w_1^*, w_2^*) = (0, 0)$ satisfaz as condições de otimalidade de segunda ordem, contudo não é um minimizante.

Teorema 10 (Condições suficientes de 2ª ordem)

Se $\nabla F(w^*) = 0$ e $\nabla^2 F(w^*)$ é definida positiva então w^* é um minimizante local de F.

Teorema 11

Se F é convexa, então qualquer minimizante local w^* é um minimizante global de F. Mais ainda, se F é diferenciável, então qualquer ponto estacionário é minimizante global de F.

Demonstração.

Suponhamos que w^* é uma minimizante local, mas não global de F. Então podemos encontrar um ponto $z \in \mathrm{R}^d$ com $F(z) < F(w^*)$. Considerando a linha segmento de reta que junto w^* a z, isto é

$$w = tz + (1-t)w^*, \text{para algum } t \in (0,1].$$

Por convexidade de F, temos

$$F(w) \leq tF(z) + (1-t)F(w^*) < F(w^*).$$

Logo, w^* não é um minimizante local. □

Demonstração.

Para a segunda parte do Teorema, o resultada é imediato pela Proposição 6 da aula 1, que garante que

$$F(w) \geq F(\bar{w}) + \nabla F(\bar{w})^T (w - \bar{w}).$$

Como $\nabla F(\bar{w}) = 0$, então $\bar{w}$ minimiza F em $\mathbb{R}^d$.
Logo é um minimizante global. $\square$

Exercício: Determine, caso existe, o(s) minimizantes das seguintes funções:

a) $F(w_1, w_2) = w_1^2 + w_1 w_2 + w_2^2 + w_1 - w_2 + 1$.
b) $F(w) = 2w_1^3 - 3w_1^2 - 6w_1 w_2(w_1 - w_2 - 1)$.
c) $F(w_1, w_2) = w_1^2 w_2^2 - 2w_1 w_2$.
d) $F(w_1, w_2) = \frac{w_1}{w_2} + \frac{8}{w_1} - w_2$.
e) $F(w_1, w_2) = \frac{1}{3}w_1^3 + \frac{1}{2}w_1^2 + 2w_1 w_2 + \frac{1}{2}w_2^2 - w_2 + 9$.

Exercício Calcule os pontos estacionários e verifique se o(s) minimizante(s), caso existem, são minimizantes globais da seguinte função de $F : \mathbb{R}^2 \to \mathbb{R}$:

$$F(w_1, w_2) = 8w_1^2 + 3w_1 w_2 + 7w_2^2 - 25w_1 + 31w_2 - 29.$$

Exercício : Deduza as condições de otimalidade para resolver o seguinte problema: $\underset{w \in \mathbb{R}^d}{\text{maximizar}}\ G(w)$.

Os apontamentos foram baseados na seguinte bibliografia: [1] e [2].

J. Nocedal and S. J. Wright.
Numerical optimization.
Springer, 1999.

G. Smirnov and V. Bushenkov.
Curso de optimização: programação matemática: cálculo de variações: controlo óptimo.
2005.