1 Aproximação dos mínimos quadrados

- Forma geral do problema

$$\text{minimizar} \sum_{j=1}^{m} (f(x_j) - M(x_j; c_i))^2.$$

- Modelo polinomial

$$p_n(x) = c_0 P_0(x) + c_1 P_1(x) + c_2 P_2(x) + \cdots + c_n P_n(x).$$

  - Polinómios ortogonais

$$P_{i+1}(x) = (x - B_i)\, P_i(x) \;-\; \mathbb{C}_i\, P_{i-1}(x),\; i = 0, \ldots, n-1,\; P_{-1}(x) = 0,\; P_0(x) = 1$$

$$B_i = \frac{\displaystyle\sum_{j=1}^{m} x_j\, P_i^2(x_j)}{\displaystyle\sum_{j=1}^{m} P_i^2(x_j)}, \quad \text{para todo o} \quad i \qquad \mathbb{C}_0 = 0 \quad \text{e} \quad \mathbb{C}_i = \frac{\displaystyle\sum_{j=1}^{m} P_i^2(x_j)}{\displaystyle\sum_{j=1}^{m} P_{i-1}^2(x_j)} \quad \text{para} \quad i > 0.$$

  - Coeficientes do polinómio

$$c_i = \frac{\displaystyle\sum_{j=1}^{m} f_j\, P_i(x_j)}{\displaystyle\sum_{j=1}^{m} P_i^2(x_j)}, \quad i = 0, 1, \ldots, n.$$

- Modelo linear não polinomial

$$M(x; c_1, \ldots, c_n) = c_1 \Phi_1(x) + c_2 \Phi_2(x) + \cdots + c_n \Phi_n(x).$$

  - Sistema das equações normais

$$\begin{pmatrix} \displaystyle\sum_{j=1}^{m} \Phi_1^2(x_j) & \displaystyle\sum_{j=1}^{m} \Phi_2(x_j)\Phi_1(x_j) & \cdots & \displaystyle\sum_{j=1}^{m} \Phi_n(x_j)\Phi_1(x_j) \\ \displaystyle\sum_{j=1}^{m} \Phi_1(x_j)\Phi_2(x_j) & \displaystyle\sum_{j=1}^{m} \Phi_2^2(x_j) & \cdots & \displaystyle\sum_{j=1}^{m} \Phi_n(x_j)\Phi_2(x_j) \\ \cdots & \cdots & \ddots & \cdots \\ \displaystyle\sum_{j=1}^{m} \Phi_1(x_j)\Phi_n(x_j) & \displaystyle\sum_{j=1}^{m} \Phi_2(x_j)\Phi_n(x_j) & \cdots & \displaystyle\sum_{j=1}^{m} \Phi_n^2(x_j) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \cdots \\ c_n \end{pmatrix} = \begin{pmatrix} \displaystyle\sum_{j=1}^{m} f_j \Phi_1(x_j) \\ \displaystyle\sum_{j=1}^{m} f_j \Phi_2(x_j) \\ \cdots \\ \displaystyle\sum_{j=1}^{m} f_j \Phi_n(x_j) \end{pmatrix}.$$

2 Mínimos *vs máximos*

$$\max f(x) = -\min(-f(x)), \quad x^* = \underbrace{\arg\max\left(f\left(x\right)\right)}_{\text{maximizante}} = \underbrace{\arg\min\left(-f\left(x\right)\right)}_{\text{minimizante}}$$

3 Método de DSC

Dados: x_1, δ, ε

1. **Há sucesso na procura para a direita ($f(x_2) \leq f(x_1)$), com $x_2 = x_1 + \delta$:** Continuar a procura para a direita $(x_2, x_3, \ldots)$, dobrando sempre o tamanho do passo $(2\delta, 4\delta, 8\delta, \ldots)$, até o valor da função subir. Ir para 4.
2. **Não há sucesso na procura para a direita, mas há sucesso na procura para a esquerda ($f(x_{-1}) \leq f(x_1)$), com $x_{-1} = x_1 - \delta$:** Continuar a procura para a esquerda $(x_{-2}, x_{-3}, \ldots)$ dobrando sempre o tamanho do passo $(2\delta, 4\delta, 8\delta, \ldots)$, até o valor da função subir. Ir para 4.
3. **Não há sucesso na procura em nenhum dos sentidos:** Ordenar os pontos por ordem crescente, que passam a chamar-se $\mathbf{x_1}$, $\mathbf{x_2}$ e $\mathbf{x_3}$. Ir para 6.
4. x_m = média dos dois últimos pontos calculados. Ir para 5.
5. escolha dos 3 pontos para a interpolação quadrática:
   i. ordenar os quatro pontos por ordem crescente
   ii. comparar os valores da função dos pontos interiores
   iii. eliminar o ponto que estiver mais longe do valor mais baixo. Se os valores forem iguais elimina-se o ponto da direita. Os pontos selecionados, ordenados por ordem crescente, passam a chamar-se $\mathbf{x_1}$, $\mathbf{x_2}$ e $\mathbf{x_3}$.
   iv. $\Delta = \mathbf{x_3} - \mathbf{x_2} = \mathbf{x_2} - \mathbf{x_1}$.
   v. Ir para 6.
6. $x^*(q) = \mathbf{x_2} + \Delta \dfrac{f(\mathbf{x_1}) - f(\mathbf{x_3})}{2(f(\mathbf{x_3}) - 2f(\mathbf{x_2}) + f(\mathbf{x_1}))}$
7. Verificar o critério de paragem: $\Delta \leq \varepsilon$
   i. O critério de paragem não se verifica.
      $x_1 = x^*(q)$. Fazer $\delta = M\delta$ e ir para 1.
   ii. O critério de paragem verifica-se.
      $x^* \approx x^*(q)$. Terminar.

4 Condições de otimalidade

- Vetor gradiente de $f(x)$

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}.$$

- Matriz Hessiana de $f(x)$

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

- Condição necessária e suficiente de primeira ordem

  $\nabla f(x^*) = 0$

- Condição necessária de segunda ordem para que x^* seja minimizante/maximizante

  $\nabla^2 f(x^*)$ é semi-definida positiva/semi-definida negativa.

- Condição suficiente de segunda ordem para que x^* seja minimizante/maximizante $\nabla^2 f(x^*)$ é definida positiva/definida negativa

  - $\nabla^2 f(x^*)$ semi-definida positiva $\Rightarrow x^*$ é minimizante ou ponto sela de $f(x)$;
  - $\nabla^2 f(x^*)$ é semi-definida negativa $\Rightarrow x^*$ é maximizante ou ponto sela de $f(x)$
  - $\nabla^2 f(x^*)$ é indefinida $\Rightarrow x^*$ é ponto sela de $f(x)$.

- Uma matriz diz-se

  - **definida positiva** se todos os determinantes das submatrizes principais dessa matriz são positivos,
  - **definida negativa** se os determinantes das submatrizes principais dessa matriz são alternadamente negativos e positivos, sendo o determinante da primeira submatriz negativo,
  - **semi-definida positiva** se os determinantes das submatrizes principais dessa matriz são positivos ou iguais a zero,
  - **semi-definida negativa** se os determinantes das submatrizes principais dessa matriz são alternadamente negativos e positivos, sendo o determinante da primeira submatriz negativo, ou iguais a zero,
  - **indefinida** nos restantes casos.

5 Métodos do gradiente

- Algoritmo genérico

ler: x^1 e ε, $k \leftarrow 0$

repetir

$k \leftarrow k + 1$

calcular d^k

calcular α^k

$x^{k+1} \leftarrow x^k + \alpha^k d^k$

até $\left\|\nabla f(x^{k+1})\right\|_2 \leq \varepsilon$

$x^* \leftarrow x^{k+1}$, $f(x^*) \leftarrow f(x^{k+1})$

- Critério de Armijo

ler: x^k, d^k, $f(x^k)$, $\nabla f(x^k)$ e μ

$\alpha \leftarrow 2$

repetir

$\alpha \leftarrow 0.5 \times \alpha$

$x^{\text{aux}} \leftarrow x^k + \alpha d^k$

até $f(x^{\text{aux}}) \leq f(x^k) + \mu\alpha\nabla f(x^k)^T d^k$

$\alpha^k \leftarrow \alpha$

- Segurança-Newton

ler: x^k e η

resolver o sistema linear Newton $\nabla^2 f(x^k) d_N^k = -\nabla f(x^k)$ por EGPP

se $\exists d_N^k$ (o sistema linear tem solução única) **então**

se $\left|\nabla f(x^k)^T d_N^k\right| \leq \eta$ **então**

$d_{SN}^k = -\nabla f(x^k)$

senão

se $\nabla f(x^k)^T d_N^k > \eta$ **então**

$d_{SN}^k = -d_N^k$

senão

$d_{SN}^k = d_N^k$

fim se

fim se

senão

$d_{SN}^k = -\nabla f(x^k)$

fim se

- Quasi-Newton

ler: x^k

se $k = 1$ **então**

$H^k = I$

senão

$s^{k-1} \leftarrow x^k - x^{k-1}$

$y^{k-1} \leftarrow \nabla f(x^k) - \nabla f(x^{k-1})$

atualizar H^k por DFP ou BFGS

fim se

$d_{QN}^k \leftarrow -H^k \nabla f(x^k)$

se $\nabla f(x^k)^T d_{QN}^k \geq 0$ **então**

$d_{QN}^k \leftarrow -\nabla f(x^k)$

fim se

- Fórmula DFP: $H^{(k)} = H^{(k-1)} - \dfrac{H^{(k-1)} y^{(k-1)} y^{(k-1)^T} H^{(k-1)}}{y^{(k-1)^T} H^{(k-1)} y^{(k-1)}} + \dfrac{s^{(k-1)} s^{(k-1)^T}}{s^{(k-1)^T} y^{(k-1)}}$

- Fórmula BFGS: $H^{(k)} = \left(I - \dfrac{s^{(k-1)} y^{(k-1)^T}}{s^{(k-1)^T} y^{(k-1)}}\right) H^{(k-1)} \left(I - \dfrac{y^{(k-1)} s^{(k-1)^T}}{s^{(k-1)^T} y^{(k-1)}}\right) + \dfrac{s^{(k-1)} s^{(k-1)^T}}{s^{(k-1)^T} y^{(k-1)}}$

6 Método de Nelder-Mead

1. Ordenar o simplex por ordem crescente dos valores da função.

- X_1 – o melhor vértice;
- X_n – o segundo pior vértice;
- X_{n+1} – o pior vértice.

$x_r = (1+\alpha)\overline{x} - \alpha X_{n+1}, \quad x_e = \gamma x_r + (1-\gamma)\overline{x}, \quad \hat{x}_c = \beta x_r + (1-\beta)\overline{x}, \quad x_c = \beta X_{n+1} + (1-\beta)\overline{x}.$

$\alpha = 1$, $\beta = 0.5$, $\gamma = 2$

Encolher o simplex: $x_i = \frac{X_i + X_1}{2}$

Em cada iteração fazer:

$$x_r \begin{cases} \underset{f(x_r)<f(X_1)}{\text{muito bom}} & \Rightarrow \quad x_e \begin{cases} \underset{f(x_e)<f(X_1)}{\text{muito bom}} & \Rightarrow \quad \text{aceita-se } x_e \\ \text{caso contrário} & \Rightarrow \quad \text{aceita-se } x_r \end{cases} \\ \underset{f(X_1)\leq f(x_r)<f(X_n)}{\text{bom}} & \Rightarrow \quad \text{aceita-se } x_r \\ \underset{f(X_n)\leq f(x_r)<f(X_{n+1})}{\text{fraco}} & \Rightarrow \quad \hat{x}_c \begin{cases} \underset{f(\hat{x}_c)<f(X_n)}{\text{bom}} & \Rightarrow \quad \text{aceita-se } \hat{x}_c \\ \text{caso contrário} & \Rightarrow \quad \text{encolhe-se o simplex} \end{cases} \\ \underset{f(x_r)\geq f(X_{n+1})}{\text{muito fraco}} & \Rightarrow \quad x_c \begin{cases} \underset{f(x_c)<f(X_n)}{\text{bom}} & \Rightarrow \quad \text{aceita-se } x_c \\ \text{caso contrário} & \Rightarrow \quad \text{encolhe-se o simplex} \end{cases} \end{cases}$$

Até se verificar critério de paragem: $\frac{1}{\Delta} \max_{2\leq i\leq n+1} \|X_i - X_1\|_2 \leq \varepsilon$, com $\Delta = \max(1, \|X_1\|_2)$.

Para verificar o critério de paragem é necessário que o simplex se encontre ordenado.

Quando se verificar o critério de paragem para-se e $x^* \approx X_1$.