Exercices cours optimisation

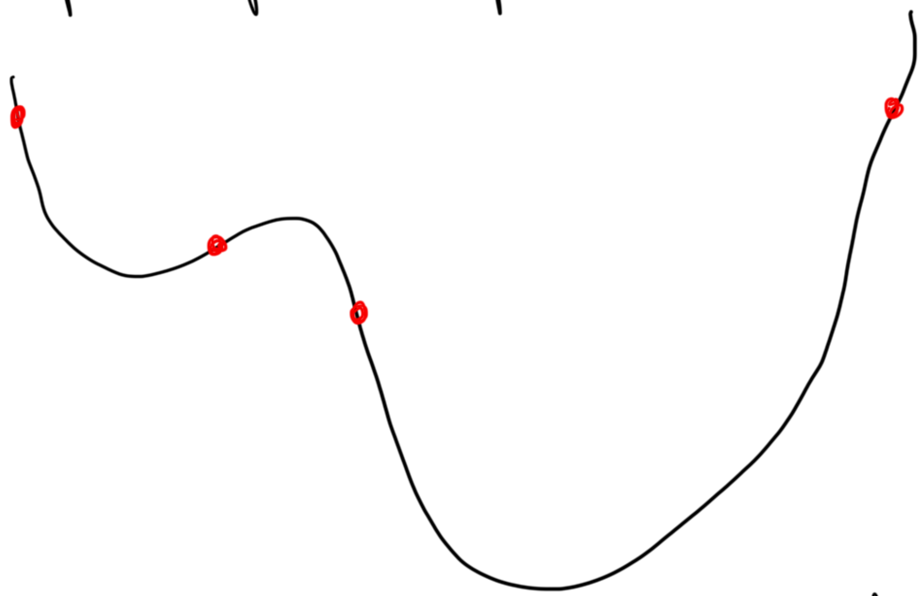
Exercice n°1 :

L'analogue en "temps continu" de la descente de gradient est l'équation du "gradient flow"

$$\frac{d}{dt}\theta(t) = -\nabla f(\theta(t))$$

1. Calculer $\frac{d}{dt} f(\theta(t))$ et expliquer comment le gradient flow minimise la fonction f.

2. Dessiner les trajectoires du gradient flow sur l'exemple suivant en partant des points de départ • :



3. Comparer avec l'équation continue de la méthode de Newton

$$\frac{d}{dt}\theta(t) = -\left(\nabla^2 f(\theta(t))\right)^{-1}\nabla f(\theta(t))$$

en cherchant l'équation différentielle satisfaite par $\nabla f(\theta(t))$.

Exercice n°2 :

Soit $A \in S_n(\mathbb{R})$, $b \in \mathbb{R}^n$.

On considère $f(x) = x^T A x + b^T x$.

1. Écrire l'itération correspondant à un pas de descente de gradient avec pas η.

2. Montrer que le point fixe de l'algorithme est l'unique minimiseur.

3. Expliquer pourquoi le choix de η optimal dépend des valeurs propres de A.

4. Que se passe-t-il si $\eta > 2/\lambda_{max}(A)$.

Exercice n°3 :

On veut minimiser $F(\theta) = E_z(\ell(\theta, z))$, et on dispose d'un estimateur $g(\theta, z)$ tq $E_z(g(\theta, z)) = \nabla F(\theta)$.

1. Écrire l'algorithme de SGD.

2. Mq $E(\theta_{t+1} | \theta_t) = \theta_t - \eta \nabla F(\theta)$.

3. Pourquoi SGD ne suit-il donc pas exactement les trajectoires du GD.

4. À quoi ressemblent les trajectoires si le pas est constant ? S'il décroît ?