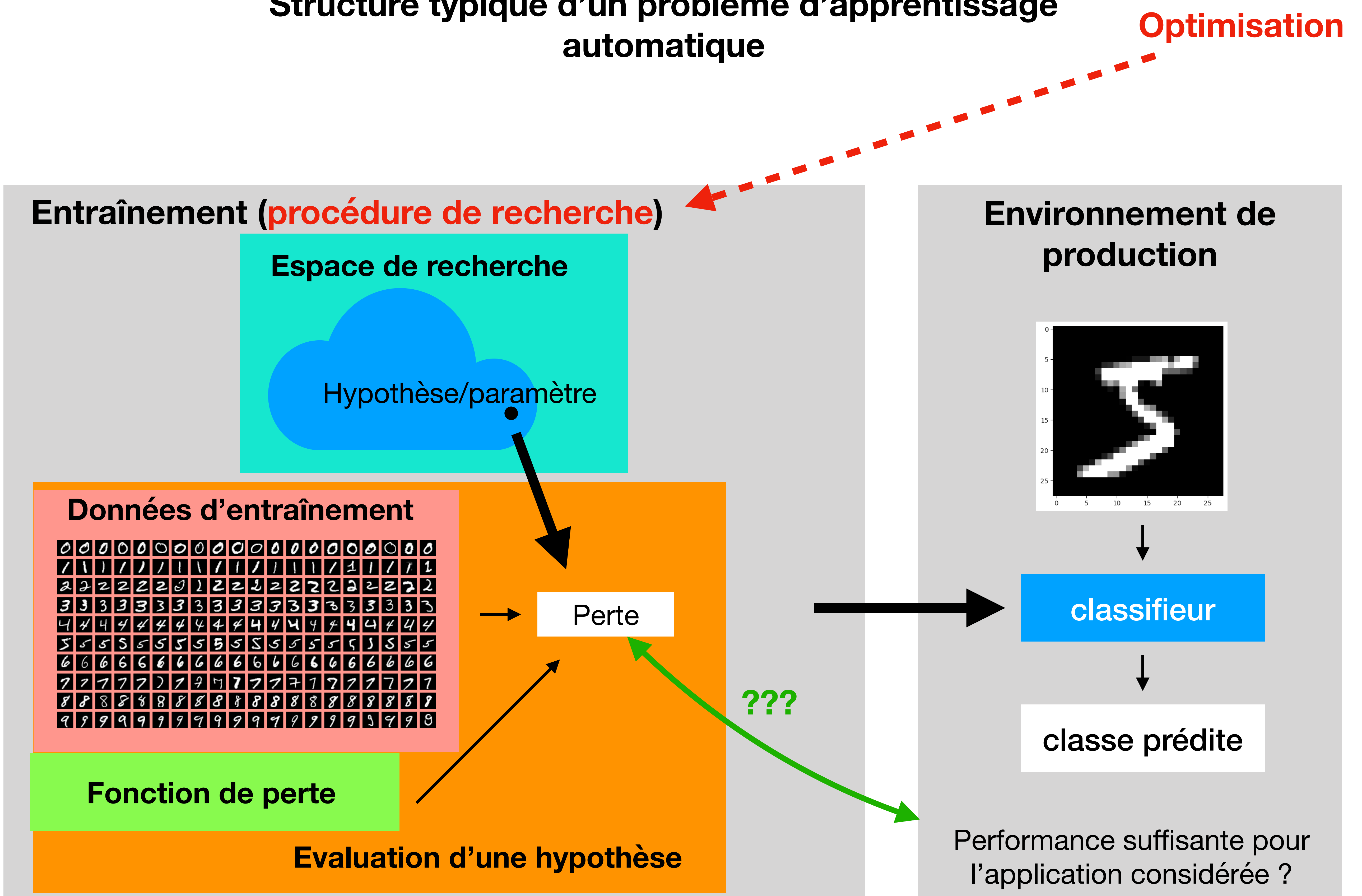
Optimisation

Structure typique d'un problème d'apprentissage automatique



Optimisation

$$f : \mathcal{D} = \mathrm{dom}(f) \subset \mathbf{R}^n \to \mathbf{R}$$

$$\min_x f(x) \ ?$$

- Minimisation d'une fonction sur un domaine
  - Domaine "ouvert" ou non ?
  - Domaine et fonction convexes ?
  - Fonction lisse ou non ?

Optimisation

1. Optimisation lisse sans contraintes
2. Optimisation lisse avec contraintes
3. Algorithmes d'optimisation
4. Analyse convexe et dualité Lagrangienne
5. Ouverture vers l'optimisation non-lisse

Définitions

$$f : \mathcal{D} = \mathrm{dom}(f) \subset \mathbf{R}^n \to \mathbf{R}$$

$$\min_x f(x) \ ?$$

x^* est un minimum global de f ssi pour tout $x \in \mathrm{dom}(f)$, $f(x^*) \leq f(x)$

x^* est un minimum local de f ssi il existe un ensemble ouvert $U \subset$ dom(f) contenant x^* tel que pour tout $x \in U$, $f(x^*) \leq f(x)$

Notions de base de topologie

Dans $\mathbf{R}^n$, La boule fermée $B_f(x,\epsilon)$, centrée sur x et de rayon ϵ associée à la norme Euclidienne est l'ensemble des points y de $\mathbf{R}^n$ tels que $\|x-y\|_2 \leq \epsilon$

Dans $\mathbf{R}^n$, La boule ouverte $B_o(x,\epsilon)$, centrée sur x et de rayon ϵ associée à la norme Euclidienne est l'ensemble des points y de $\mathbf{R}^n$ tels que $\|x-y\|_2 < \epsilon$

U est un ouvert de $\mathbf{R}^n$ ssi $U \subset \mathbf{R}^n$ et pour tout x dans U, il existe $\epsilon > 0$, tel que $B_f(x,\epsilon) \subset U$.

U est un fermé de $\mathbf{R}^n$ ssi son complément dans $\mathbf{R}^n$ est un ouvert de $\mathbf{R}^n$

Jacobienne, gradient

$$f : U \subset \mathbf{R}^n \to \mathbf{R}^m, \text{ de classe } \mathrm{C}^1$$

Dérivée partielle $f : x \mapsto (f_1(x), f_2(x), \ldots, f_m(x))$

$$\frac{\partial f_i}{\partial x_j} : (a_1, \ldots, a_n) \mapsto \lim_{h \to 0} \frac{f_i(a_1, \ldots, a_{j-1}, a_j + h, a_{j+1}, \ldots a_n) - f_i(a_1, \ldots, a_j, \ldots, a_n)}{h}$$

Matrice Jacobienne (transposée du gradient si m=1)

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial \mathbf{f}}{\partial x_1} & \cdots & \dfrac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^{\mathrm{T}} f_1 \\ \vdots \\ \nabla^{\mathrm{T}} f_m \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}$$

“Chain rule”

$$J_{f \circ g}(\mathbf{a}) = J_f(g(\mathbf{a})) J_g(\mathbf{a})$$

Conditions d'optimalité

Conditions nécessaires, premier ordre

Si x^* est un minimum local et f est de classe $\mathcal{C}^1$ sur un ouvert $U \subset \mathbf{R}^n$ contenant x, alors $\nabla f(x^*) = 0$

Hessienne

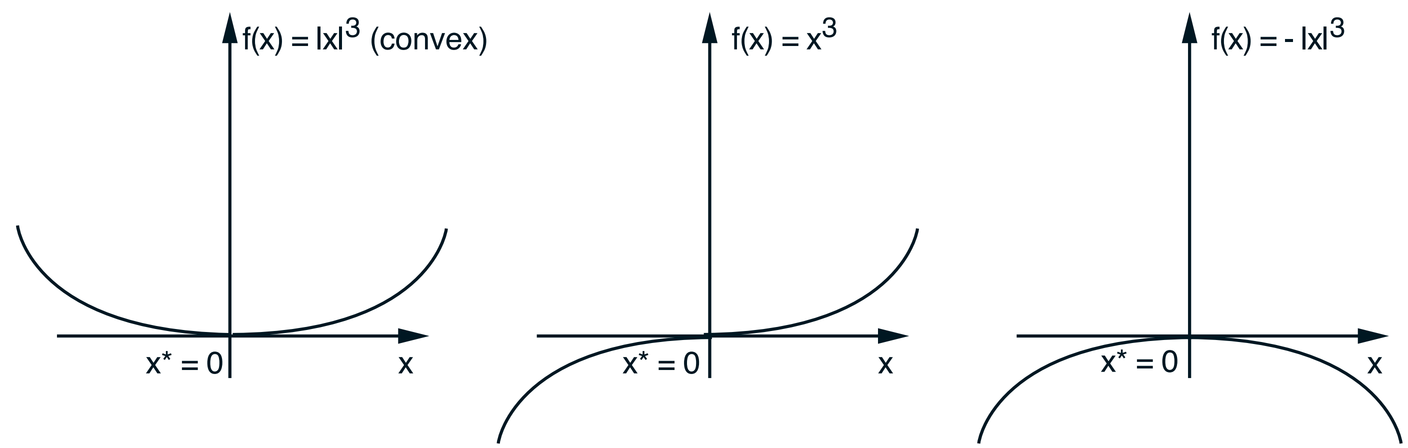
$$f : U \subset \mathbf{R}^n \mapsto \mathbf{R}$$

$$\nabla^2 f = H(f) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial {x_1}^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial {x_2}^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial {x_n}^2} \end{bmatrix}$$

Conditions d'optimalité

Conditions nécessaires, premier ordre

Si x^* est un minimum local et f est de classe $\mathcal{C}^1$ sur un ouvert $U \subset \mathbf{R}^n$ contenant x, alors $\nabla f(x^*) = 0$



Conditions nécessaires, second ordre

Si x^* est un minimum local et f est de classe $\mathcal{C}^2$ sur un ouvert $U \subset \mathbf{R}^n$ contenant x, alors $\nabla f(x^*) = 0$ et $\nabla^2 f(x^*)$ est semi-définie positive

Conditions suffisantes, second ordre

Si f est de classe $\mathcal{C}^2$ sur un ouvert $U \subset \mathbf{R}^n$ contenant x, si $\nabla f(x^*) = 0$ et si $\nabla^2 f(x^*)$ est définie positive, alors x^* est un minimum local.

Existence de minimum

Théorème de Weierstrass

Si f est définie et continue sur un sous-ensemble fermé et borné de $\mathbf{R}^n$, alors f admet un minimum global.

Théorème

Si f est définie et continue sur $\mathbf{R}^n$ et *coercive* (i.e. $f(x) \to +\infty$ when $\|x\| \to +\infty$, alors f admet un minimum global sur tout sous-ensemble fermé de $\mathbf{R}^n$.

Théorème

Si f est convexe et minorée, alors f admet un minimum global.

Optimisation

1. Optimisation lisse sans contraintes
2. Optimisation lisse avec contraintes
3. Analyse convexe et dualité Lagrangienne
4. Ouverture vers l’optimisation non-lisse
5. Algorithmes d’optimisation

Optimisation sous contraintes

$$\min_{x \in \mathbf{R}^n} f(x) \text{ subject to } \begin{cases} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \leq 0, & i \in \mathcal{I} \end{cases}$$

$$\Omega = \{x \in \mathbf{R}^n \mid \text{pour tout } i \in \mathcal{E},\ c_i(x) = 0,\ \text{pour tout } j \in \mathcal{I},\ c_j(x) \leq 0\}$$

x^* est une solution locale du problème ssi $x^* \in \Omega$ et il existe un ensemble ouvert $U \subset \mathbf{R}^n$ contenant x^* tel que pour tout $x \in U \cap \Omega$, $f(x^*) \leq f(x)$

Condition nécessaire d'optimalité: intuition

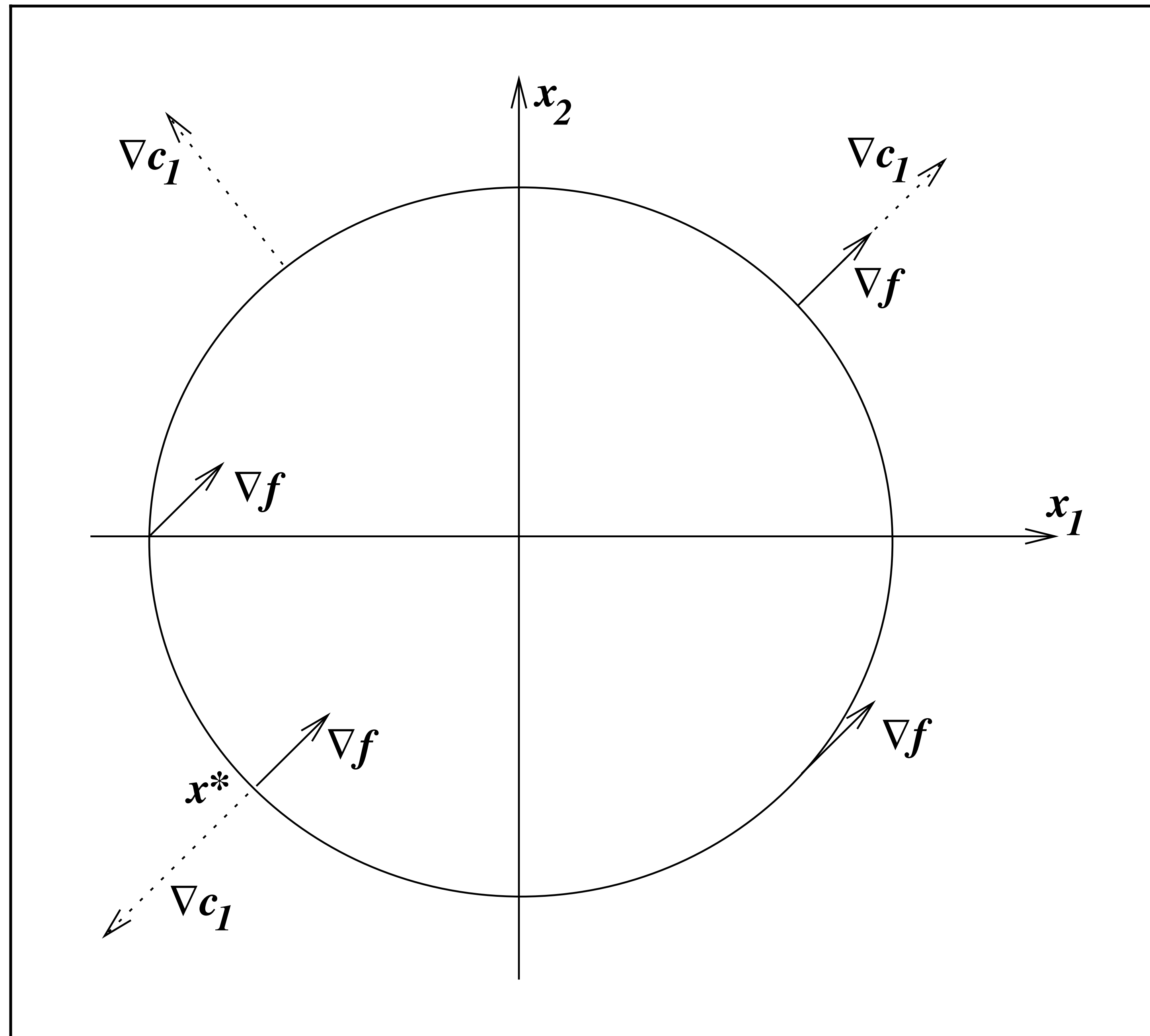
$$f : x_1, x_2 \mapsto x_1 + x_2$$

Cas d'une seule contrainte d'égalité

$$x_1^2 + x_2^2 = 2$$

Condition nécessaire d'optimalité

$$\nabla f(x^*) = \lambda_1^* \nabla c_1(x^*).$$



Condition nécessaire d'optimalité: intuition

$$f : x_1, x_2 \mapsto x_1 + x_2$$

Cas d'une seule contrainte d'inégalité

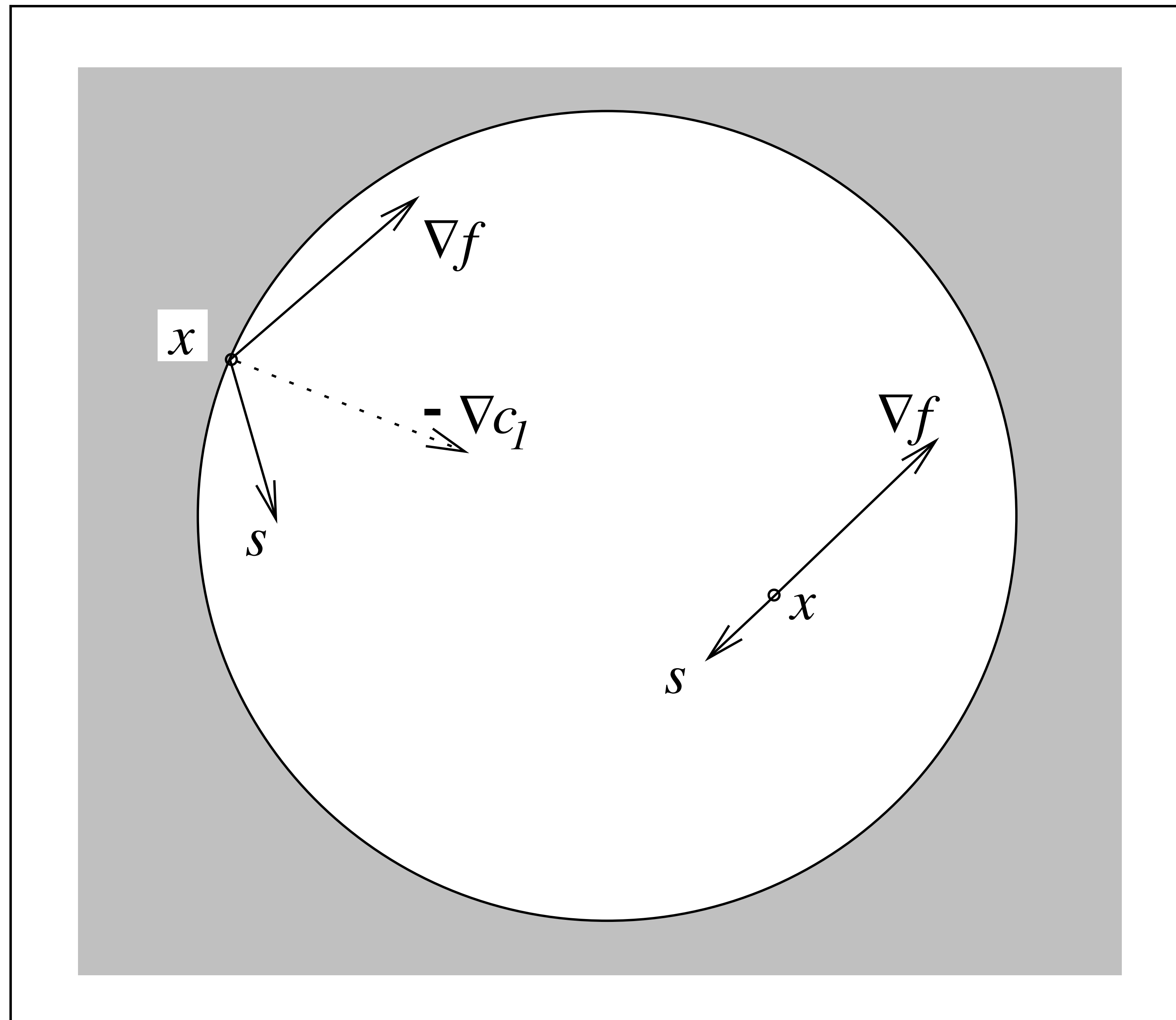
$$x_1^2 + x_2^2 \leq 2$$

Condition nécessaire d'optimalité

$$\nabla f(x^*) = -\lambda_1^* \nabla c_1(x^*)$$

$$\lambda_1^* \geq 0$$

$$\lambda_1^* c_1(x^*) = 0.$$



Cas général : KKT conditions

Soit f, $(e_i)_{i=1}^{n_E}$ et $(c_i)_{i=1}^{n_I}$ des fonctions de $\mathbf{R}^d$ vers $\mathbf{R}$ de classe $\mathcal{C}^1$.

on dit que le point $x^* \in \mathbf{R}^d$ vérifie les conditions de Karush-Kuhn-Tucker (KKT) pour le problème :

$$\min_{x \in \mathbf{R}^d} f(x) \text{ tel que } \begin{cases} e_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \leq 0, & i \in \mathcal{I} \end{cases}$$

si et seulement si il existe $\lambda_E^* \in \mathbf{R}^{n_E}$ et $\lambda_I^* \in \mathbf{R}^{n_I}$, tels que :

1. $\nabla_x \mathcal{L}(x^*, \lambda_E^*, \lambda_I^*) = 0$, où $\mathcal{L}(x, \lambda_E, \lambda_I) = f(x) + \sum_{i=1}^{n_E} \lambda_{E,i} e_i(x) + \sum_{i=1}^{n_I} \lambda_{I,i} c_i(x)$
2. $e_i(x^*) = 0$ pour tout i dans $\{1, 2, \ldots, n_E\}$
3. $c_i(x^*) \leq 0$ pour tout i dans $\{1, 2, \ldots, n_I\}$
4. $\lambda_{I,i} \geq 0$ pour tout i dans $\{1, 2, \ldots, n_I\}$
5. $\lambda_{I,i} c_i(x^*) = 0$ pour tout i dans $\{1, 2, \ldots, n_I\}$

Condition nécessaire d'optimalité: cas général

Mangasarian-Fromovitz Constraint Qualification (MFCQ)
(Plus d'exemples de qualifications des contraintes sur la feuille d'exercice numéro 1)

Les gradients des contraintes d'égalité sont linéairement indépendants aux point x^* et il existe un vecteur $d \in \mathbf{R}^n$ tel que $\nabla c_i(x^*)^T d < 0$ pour toutes les contraintes d'inégalité **actives** en x^* et tel que $\nabla e_j(x^*)^T d = 0$ pour toutes les contraintes d'égalité.

Théorème : Si x^* est une solution locale et que la qualification des contraintes de Mangasarian-Fromovitz est respectée en x^* alors les conditions KKT sont vérifiées en x^*.

Autres exemples de qualifications des contraintes

Theorem 1 *Conditions nécessaires de solution locale pour des contraintes linéaires. Si les fonctions e_i et c_j sont affines pour $i = 1 \dots n_E$ et $j = 1 \dots n_I$ alors les conditions KKT sont des conditions nécessaires de solution locale pour le problème considéré.*

Theorem 2 *Conditions nécessaires de solution locale pour des contraintes linéairement indépendantes.*

*Si au point $x \in \mathbf{R}^d$ la matrice formée par les gradients des contraintes d'égalité et des contraintes d'inégalité **actives** est de rang plein (c'est à dire que les n_E vecteurs gradients pour les contraintes d'égalité—qui doivent toujours toutes être actives—et les vecteurs gradients pour les contraintes d'inégalité actives en x—c'est à dire telles que $c_i(x) = 0$—sont linéairement indépendants), alors les conditions KKT sont des conditions nécessaires de solution locale en x pour le problème considéré.*

Theorem 3 *Conditions nécessaires de solution locale pour un problème convexe.*

Si f est convexe, c_i est convexe pour tout i in $1 \dots n_I$ et e_j est affine pour tout j in $1 \dots n_E$ et si la condition de Slater est vérifiée, c'est à dire s'il existe $x \in \mathbf{R}^d$ tel que $e_j(x) = 0$ pour tout j in $1 \dots n_E$ et $c_i(x) < 0$ pour tout i in $1 \dots n_I$, alors les conditions KKT sont des conditions nécessaires de solution locale pour le problème considéré.