

Scribes

- Document à me rendre au plus tard une semaine après la séance (rendre un seul document par groupe)
- 2/09 Clara Karkach, Jules Danielou
- 9/09 Mohamed Alloui, Cherifi Aymen
- 16/09 Raissa Mezine, Sajjad Ghasemian
- 24/09 Joe Hubi, Mohamed Alloui
- 25/09 Idir Bellili, Saal Racim
- 30/09 Anlaoudine Moindze Boina
- 7/10 Amal Bourahma Atchory Rameaux Mel

Correction exercice 7, 8, 11

Différentes visions des opérations matricielles : application

LoRA: Low-Rank Adaptation of Large Language Models

Edward Hu* Yelong Shen* Phillip Wallis
Zeyuan Allen-Zhu Yuanzhi Li Shean Wang Weizhu Chen
Microsoft Corporation
{edwardhu, yeshe, phwallis, zeyuana, yuanzhil, swang, wzchen}@microsoft.com

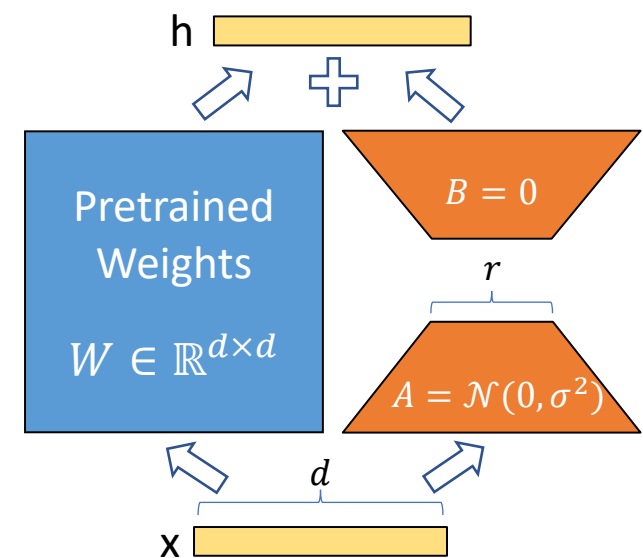


Figure 1: Our reparametrization. We only train A and B .

ICLR 2022, 18179 citations au 9/9/2025

Adaptation *de rang faible* ? Que peut-on dire du rang de $\delta W = AB$?

Rappel: le rang d'une matrice est le plus grand nombre de vecteurs linéairement indépendants qu'on peut trouver au sein de l'ensemble des colonnes de la matrice (vues comme des vecteurs).

Algèbre linéaire

- Plusieurs interprétations des opérations matricielles
- Produit scalaire, orthogonalité, normes
- Décomposition en valeurs singulières
- Théorème spectral, positivité, déterminant, trace

Décomposition en valeurs singulières

$$AV = U\Sigma \quad A \begin{bmatrix} v_1 & \dots & v_r & \dots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \dots & u_r & \dots & u_m \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array} \right]$$

Correction exercice 15

Flashback: interpretation des opérations matricielles

$$c_{ij} = \tilde{a}_i^T b_j = \langle \tilde{a}_i, b_j \rangle$$

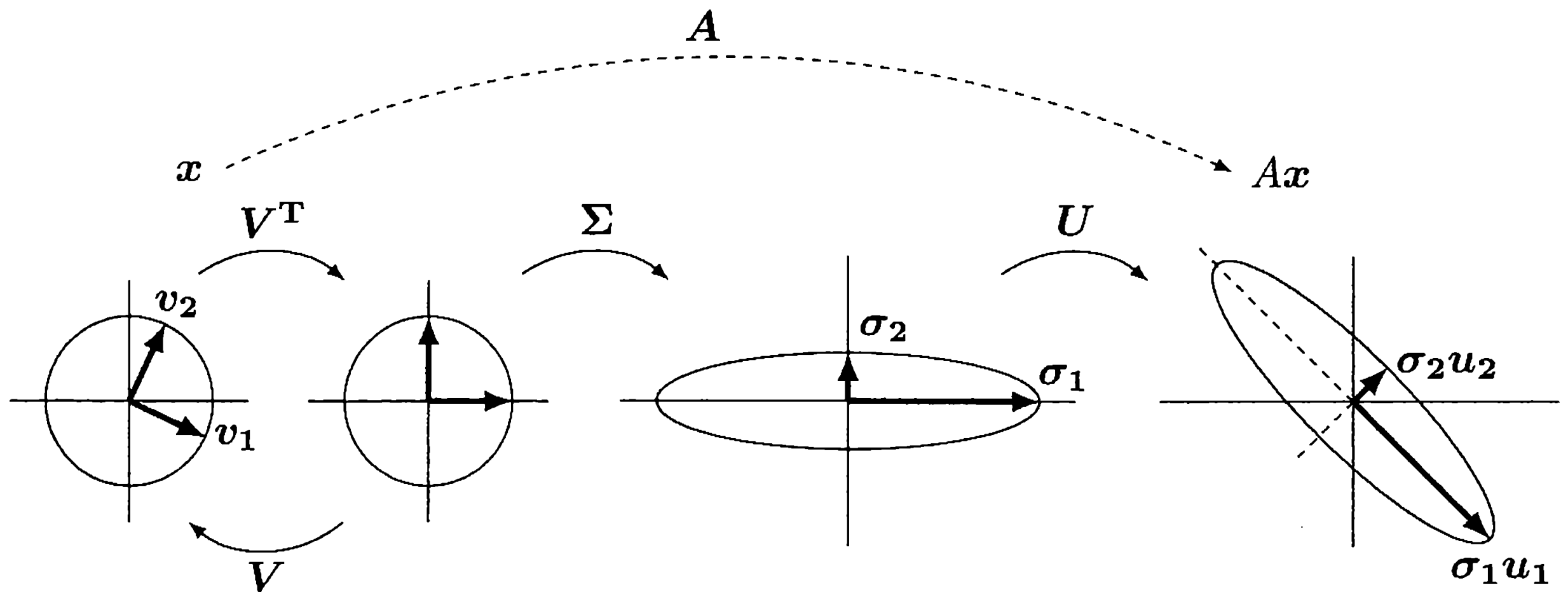
$$C = \begin{bmatrix} c_1 & \cdots & c_p \end{bmatrix} = AB = \begin{bmatrix} Ab_1 & \cdots & Ab_p \end{bmatrix}$$

$$C = \begin{bmatrix} \tilde{c}_1^T \\ \vdots \\ \tilde{c}_m^T \end{bmatrix} = AB = \begin{bmatrix} \tilde{a}_1^T B \\ \vdots \\ \tilde{a}_m^T B \end{bmatrix}$$

$$C = \sum_i a_i \tilde{b}_i^T$$

Décomposition en valeurs singulières

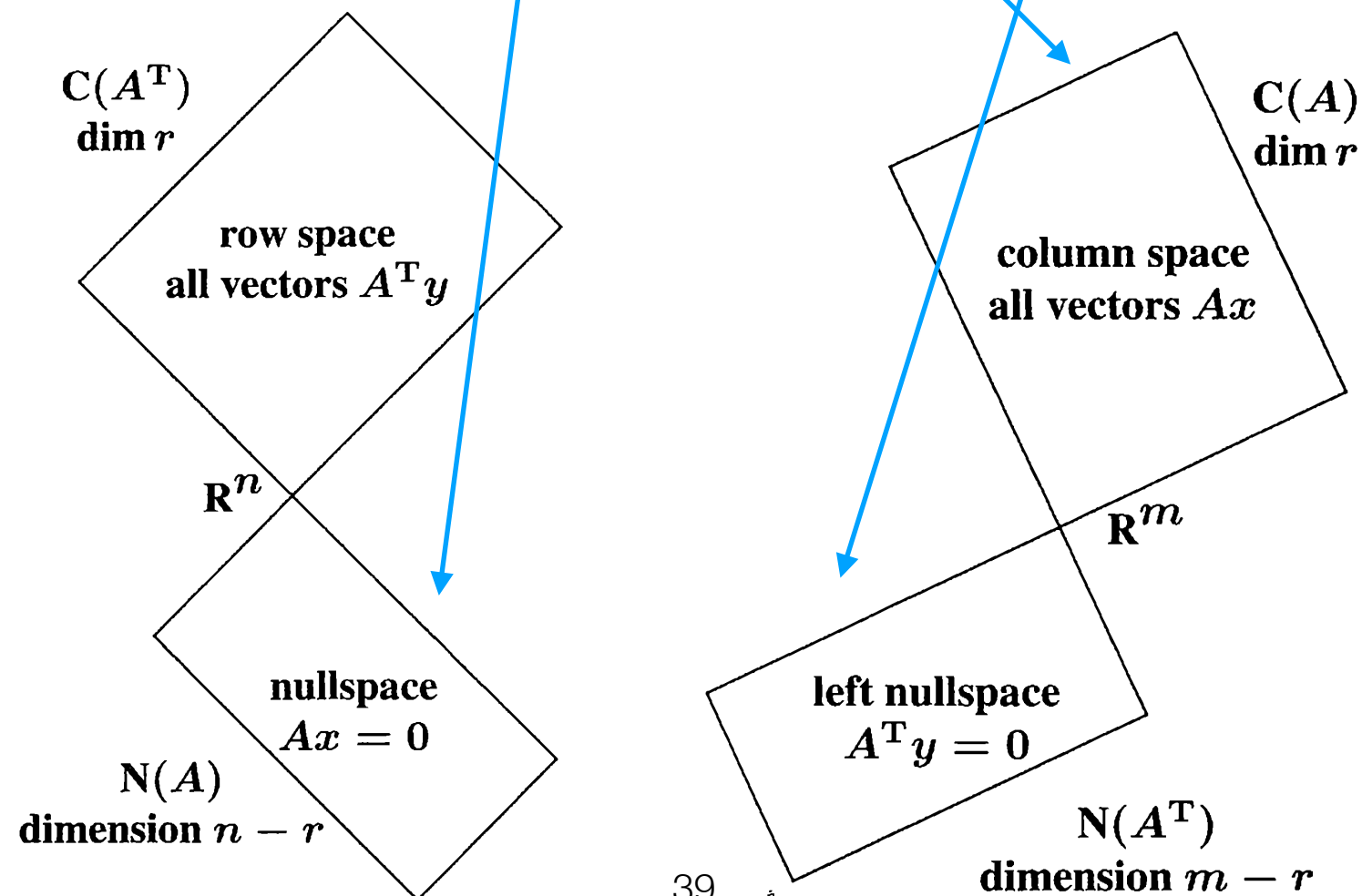
$$AV = U\Sigma \quad A \begin{bmatrix} v_1 & \dots & v_r & \dots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \dots & u_r & \dots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ \hline & & & 0 & \\ & & & & 0 \end{bmatrix}$$



Espaces associés à une matrice

Bases orthonormées

$$AV = U\Sigma \quad A \begin{bmatrix} \underline{v_1 \dots v_r} & \underline{v_{r+1} \dots v_n} \end{bmatrix} = \begin{bmatrix} \underline{u_1 \dots u_r} & \underline{u_{r+1} \dots u_m} \end{bmatrix} \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ \hline & & 0 & 0 \end{bmatrix}$$



Espaces associés à une matrice

$$AV = U\Sigma \quad A \begin{bmatrix} v_1 & \dots & v_r & \dots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \dots & u_r & \dots & u_m \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ \hline & & 0 & 0 \end{array} \right]$$

Par exemple : $M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$

$$U \approx \begin{pmatrix} -.21 & -.89 & .41 \\ -.52 & -.25 & -.82 \\ -.83 & .39 & .41 \end{pmatrix} \quad V \approx \begin{pmatrix} -.48 & -.57 & -.66 \\ .78 & .08 & -.62 \\ .41 & -.82 & .41 \end{pmatrix}$$

$$(\sigma_1 \quad \sigma_2 \quad \sigma_3) \approx (16.85 \quad 1.07 \quad 0)$$

Algèbre linéaire

- Plusieurs interprétations des opérations matricielles
- Produit scalaire, orthogonalité, normes
- Décomposition en valeurs singulières et applications
- Théorème spectral et applications

Théorème spectral

Si S est une matrice symétrique, réelle de taille $m \times m$, alors il existe une matrice orthogonale réelle Q de taille $m \times m$ et une matrice diagonale réelle Λ de taille $m \times m$ telles que $S = Q\Lambda Q^T$.

Positivité

Une matrice symétrique réelle est dite définie positive, noté $S \succ 0$ ssi pour toute matrice colonne u , $u^T S u > 0$.

Une matrice symétrique réelle est dite semi-définie positive, $S \succeq 0$, ssi pour toute matrice colonne u , $u^T S u \geq 0$.

Relation d'ordre sur les matrices symétriques réelles:

$$S_1 \prec S_2 \text{ ssi } S_2 - S_1 \succ 0$$

et

$$S_1 \preceq S_2 \text{ ssi } S_2 - S_1 \succeq 0$$

Déterminant et trace

A matrice carrée $n \times n$

$$\det(A) = \sum_{\sigma \in S_n} \left(\operatorname{sgn}(\sigma) \prod_{i=1}^n a_{i, \sigma_i} \right)$$

$\operatorname{sgn}(\sigma)$ est la parité du nombre d'éléments dans une décomposition de σ en une séquence de transpositions (échange de deux éléments).

$$\det(AB) = \det(A) \det(B)$$

$$\operatorname{Tr}(A) = \sum_{i=1}^n a_{i,i} \quad \operatorname{Tr}(AB) = \operatorname{Tr}(BA)$$

$$\operatorname{Tr}(A_1 A_2 \dots A_k) = \operatorname{Tr}(A_2 A_3 \dots A_k A_1) = \dots = \operatorname{Tr}(A_k A_1 A_2 \dots A_{k-1})$$

Optimisation

Structure typique d'un problème d'apprentissage automatique

Optimisation

Entraînement (**procédure de recherche**)

Espace de recherche

Hypothèse/paramètre

Données d'entraînement

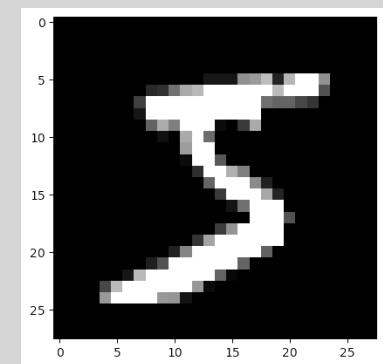
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

Fonction de perte

Perte

Evaluation d'une hypothèse

Environnement de production



classifieur

classe prédite

Performance suffisante pour l'application considérée ?

Optimisation

$$f : \mathcal{D} = \text{dom}(f) \subset \mathbf{R}^n \rightarrow \mathbf{R}$$
$$\min_x f(x) ?$$

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

Optimisation

1. Optimisation sans contraintes
2. Optimisation sous contraintes
3. Analyse convexe et dualité

Définitions

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

$$\min_x f(x) \text{ ?}$$

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

x^* est un minimum local de f ssi il existe un ensemble ouvert $U \subset \text{dom}(f)$ 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

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

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

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

Matrice Jacobienne (transposée du gradient si m=1)

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

“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

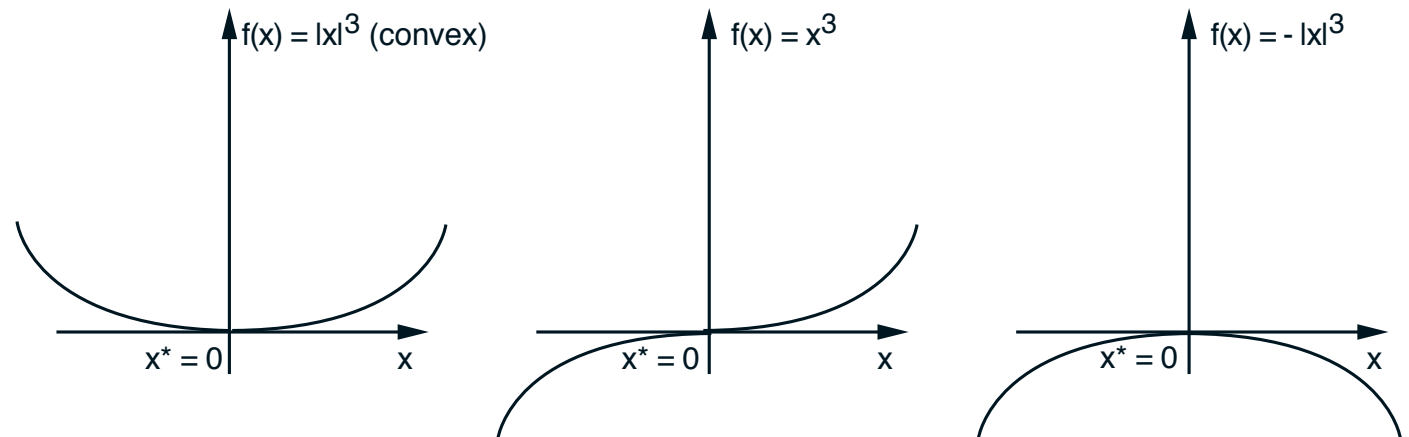
$$f : U \subset \mathbf{R}^n \mapsto \mathbf{R}$$

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

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. $f(x) \rightarrow +\infty$ when $\|x\| \rightarrow +\infty$), 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 sans contraintes
2. Optimisation sous contraintes
3. Analyse convexe et dualité

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

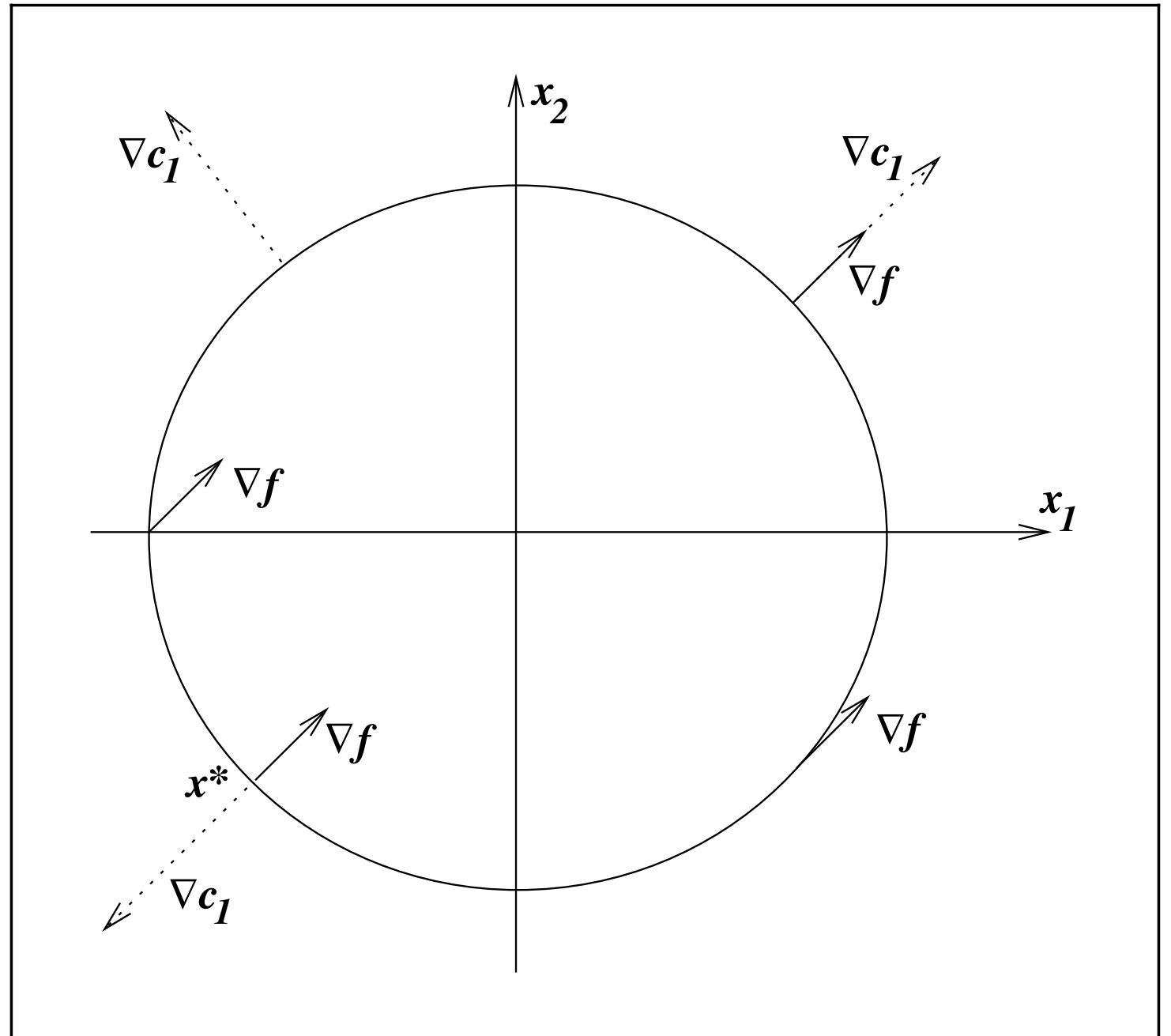
$$f : x_1, x_2 \mapsto x_1 + x_2$$

**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**

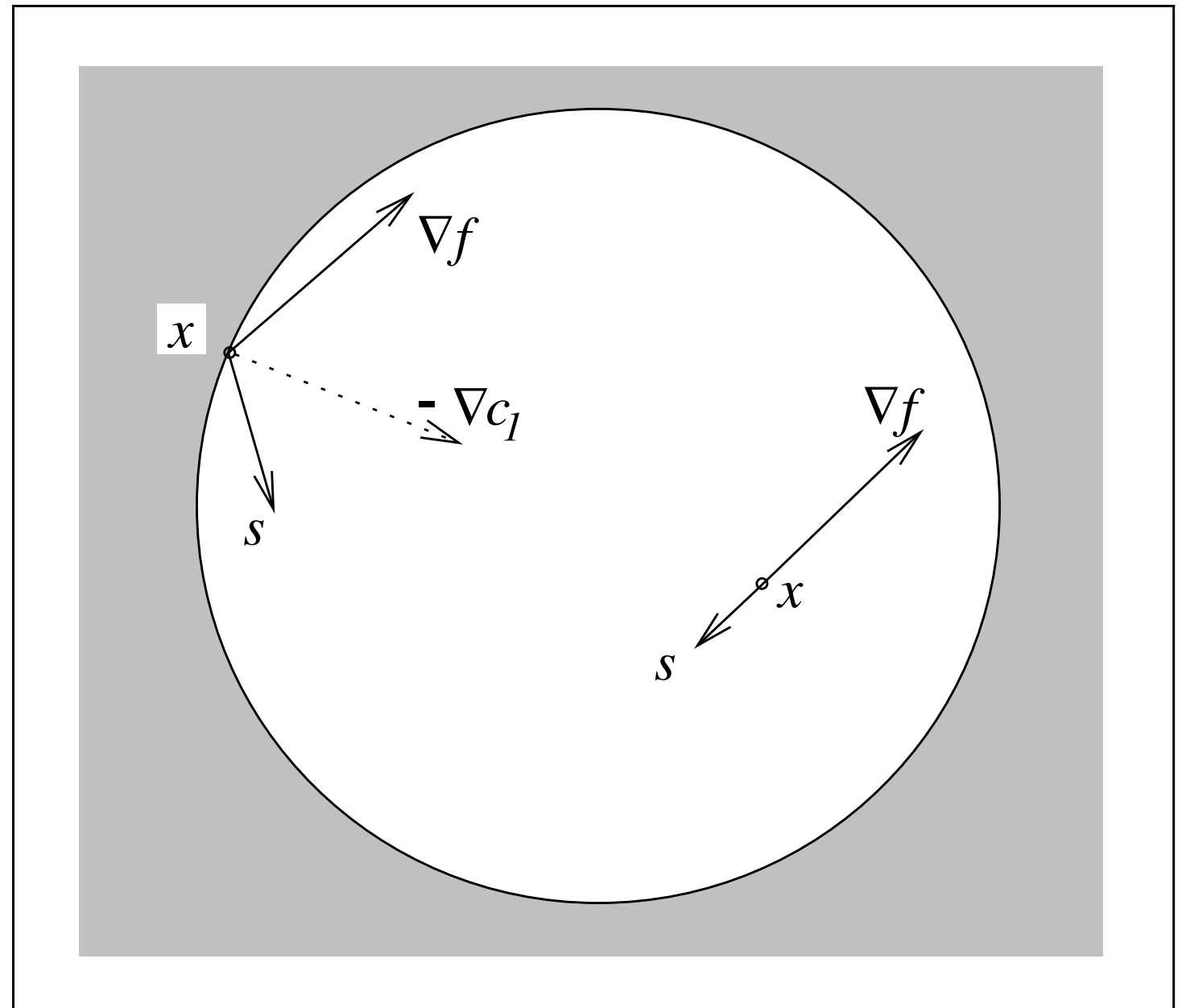
$$x_1^2 + x_2^2 \leq 2$$

**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

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$$

Conditions de Karush-Kuhn-Tucker (KKT)

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0,$$

$$c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E},$$

$$c_i(x^*) \leq 0 \quad \text{for all } i \in \mathcal{I},$$

$$\lambda_i^* \geq 0, \quad \text{for all } i \in \mathcal{I},$$

$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}.$$

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)

The gradients of the equality constraints are linearly independent at x^* and there exists a vector $d \in \mathbf{R}^n$ such that $\nabla c_i(x^*)^T d < 0$ for all **active** inequality constraints and $\nabla e_j(x^*)^T d = 0$ for all equality constraints.

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^* .