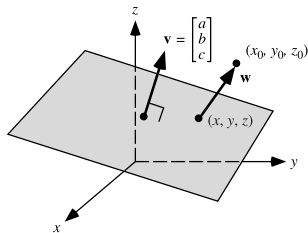


Préliminaires géométriques

Guillaume TOCHON

Laboratoire de Recherche de l'EPITA



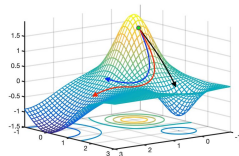
Motivation

La résolution d'un problème classique d'optimisation $(\arg) \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ pour une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ nécessite la capacité à se repérer et se déplacer dans \mathbb{R}^n :

Motivation

La résolution d'un problème classique d'optimisation $(\arg) \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ pour une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ nécessite la capacité à se repérer et se déplacer dans \mathbb{R}^n :

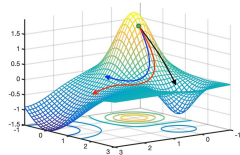
- comment décrire une direction de déplacement ?



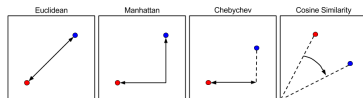
Motivation

La résolution d'un problème classique d'optimisation $(\arg) \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ pour une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ nécessite la capacité à se repérer et se déplacer dans \mathbb{R}^n :

- comment décrire une direction de déplacement ?



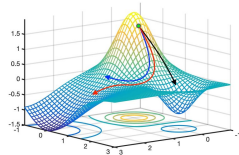
- comment évaluer des distances entre points ?



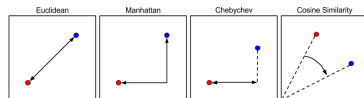
Motivation

La résolution d'un problème classique d'optimisation $(\arg) \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ pour une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ nécessite la capacité à se repérer et se déplacer dans \mathbb{R}^n :

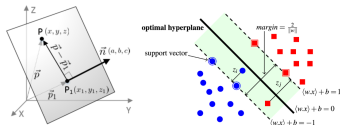
- comment décrire une direction de déplacement ?



- comment évaluer des distances entre points ?



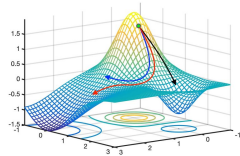
- comment décrire des sous parties "simples" de \mathbb{R}^n (droite, plan, etc) ?



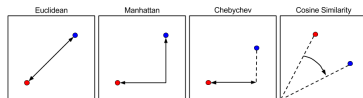
Motivation

La résolution d'un problème classique d'optimisation $(\arg) \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ pour une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ nécessite la capacité à se repérer et se déplacer dans \mathbb{R}^n :

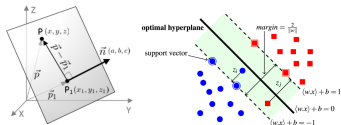
- comment décrire une direction de déplacement ?



- comment évaluer des distances entre points ?



- comment décrire des sous parties "simples" de \mathbb{R}^n (droite, plan, etc) ?



⇒ Besoin d'éléments de base de géométrie euclidienne (produit scalaire, norme, etc).

Notations

On utilisera dans ce cours la convention standard anglo-saxonne pour les notations :

Les scalaires (éléments de \mathbb{R}) en minuscule classique : $x \in \mathbb{R}$.

Notations

On utilisera dans ce cours la convention standard anglo-saxonne pour les notations :

Les scalaires (éléments de \mathbb{R}) en minuscule classique : $x \in \mathbb{R}$.

Les vecteurs (éléments de \mathbb{R}^n) en minuscule et gras : $\mathbf{x} \in \mathbb{R}^n$
→ par défaut, un vecteur est écrit en colonne : $\mathbf{x} \in \mathbb{R}^n \leftrightarrow \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

→ on écrit aussi \mathbf{x} comme transposé d'un vecteur ligne : $\mathbf{x} \in \mathbb{R}^n \leftrightarrow \mathbf{x} = (x_1, \dots, x_n)^T$

Notations

On utilisera dans ce cours la convention standard anglo-saxonne pour les notations :

Les scalaires (éléments de \mathbb{R}) en minuscule classique : $x \in \mathbb{R}$.

Les vecteurs (éléments de \mathbb{R}^n) en minuscule et gras : $\mathbf{x} \in \mathbb{R}^n$
→ par défaut, un vecteur est écrit en colonne : $\mathbf{x} \in \mathbb{R}^n \leftrightarrow \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

→ on écrit aussi \mathbf{x} comme transposé d'un vecteur ligne : $\mathbf{x} \in \mathbb{R}^n \leftrightarrow \mathbf{x} = (x_1, \dots, x_n)^T$

Les matrices (éléments de $\mathbb{R}^{m \times n}$) en majuscule et gras : $\mathbf{A} \in \mathbb{R}^{m \times n}$

→ Écriture matricielle : $\mathbf{A} = [a_{ij}]_{m \times n} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$

→ Écriture "en ligne" : $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ comme n vecteurs colonne $\mathbf{a}_i \in \mathbb{R}^m$

→ Écriture "en colonne" : $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_m^T \end{pmatrix}$ comme m vecteurs ligne $\mathbf{a}_i \in \mathbb{R}^n$

Définition

Un produit scalaire sur \mathbb{R}^n est une application $\phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ qui vérifie les propriétés suivantes :

symétrie :

bilinéarité :

définition :

positivité :

Définition

Un produit scalaire sur \mathbb{R}^n est une application $\phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ qui vérifie les propriétés suivantes :

symétrie : $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \phi(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{y}, \mathbf{x})$

bilinéarité : $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ et $\forall \lambda \in \mathbb{R}, \phi(\mathbf{x}, \mathbf{y} + \lambda \mathbf{z}) = \phi(\mathbf{x}, \mathbf{y}) + \lambda \phi(\mathbf{x}, \mathbf{z})$

définition : $\phi(\mathbf{x}, \mathbf{x}) = 0 \Rightarrow \mathbf{x} = \mathbf{0}_n$

positivité : $\forall \mathbf{x} \in \mathbb{R}^n, \phi(\mathbf{x}, \mathbf{x}) \geq 0$

Produit scalaire

Définition

Un produit scalaire sur \mathbb{R}^n est une application $\phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ qui vérifie les propriétés suivantes :

symétrie : $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \phi(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{y}, \mathbf{x})$

bilinéarité : $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ et $\forall \lambda \in \mathbb{R}, \phi(\mathbf{x}, \mathbf{y} + \lambda \mathbf{z}) = \phi(\mathbf{x}, \mathbf{y}) + \lambda \phi(\mathbf{x}, \mathbf{z})$

définition : $\phi(\mathbf{x}, \mathbf{x}) = 0 \Rightarrow \mathbf{x} = \mathbf{0}_n$

positivité : $\forall \mathbf{x} \in \mathbb{R}^n, \phi(\mathbf{x}, \mathbf{x}) \geq 0$

On note en général par $\langle \mathbf{x}, \mathbf{y} \rangle$ le produit scalaire de \mathbf{x} et \mathbf{y} : $\langle \cdot, \cdot \rangle = \phi(\cdot, \cdot)$

Dans \mathbb{R}^n , le produit scalaire *canonique* est défini par

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i = \mathbf{x}^T \mathbf{y}$$

Produit scalaire

Définition

Un produit scalaire sur \mathbb{R}^n est une application $\phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ qui vérifie les propriétés suivantes :

symétrie : $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \phi(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{y}, \mathbf{x})$

bilinéarité : $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ et $\forall \lambda \in \mathbb{R}, \phi(\mathbf{x}, \mathbf{y} + \lambda \mathbf{z}) = \phi(\mathbf{x}, \mathbf{y}) + \lambda \phi(\mathbf{x}, \mathbf{z})$

définition : $\phi(\mathbf{x}, \mathbf{x}) = 0 \Rightarrow \mathbf{x} = \mathbf{0}_n$

positivité : $\forall \mathbf{x} \in \mathbb{R}^n, \phi(\mathbf{x}, \mathbf{x}) \geq 0$

On note en général par $\langle \mathbf{x}, \mathbf{y} \rangle$ le produit scalaire de \mathbf{x} et \mathbf{y} : $\langle \cdot, \cdot \rangle = \phi(\cdot, \cdot)$

Dans \mathbb{R}^n , le produit scalaire *canonique* est défini par $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i = \mathbf{x}^T \mathbf{y}$

Tout produit scalaire permet de définir une norme par $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$, qui correspond intuitivement à la longueur du vecteur \mathbf{x}

Pour le produit scalaire canonique, la norme associée s'appelle norme euclidienne et se note $\|\cdot\|_2$: $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T \mathbf{x}} = \sqrt{\sum_{i=1}^n (x_i)^2}$



Angle et espace orthogonal

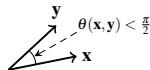
$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos(\theta(\mathbf{x}, \mathbf{y}))$ avec $\theta(\mathbf{x}, \mathbf{y})$ l'angle entre les vecteurs \mathbf{x} et \mathbf{y}
 $\Rightarrow \theta(\mathbf{x}, \mathbf{y}) = \cos^{-1} \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \right) \rightarrow$ le signe de $\langle \mathbf{x}, \mathbf{y} \rangle$ donne l'orientation entre \mathbf{x} et \mathbf{y}

Angle et espace orthogonal

$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos(\theta(\mathbf{x}, \mathbf{y}))$ avec $\theta(\mathbf{x}, \mathbf{y})$ l'angle entre les vecteurs \mathbf{x} et \mathbf{y}

$\Rightarrow \theta(\mathbf{x}, \mathbf{y}) = \cos^{-1} \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \right) \rightarrow$ le signe de $\langle \mathbf{x}, \mathbf{y} \rangle$ donne l'orientation entre \mathbf{x} et \mathbf{y}

- si $\langle \mathbf{x}, \mathbf{y} \rangle > 0$, $|\theta(\mathbf{x}, \mathbf{y})| < \frac{\pi}{2}$
 $\rightarrow \mathbf{x}$ et \mathbf{y} pointent dans la même direction.

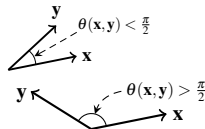


Angle et espace orthogonal

$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos(\theta(\mathbf{x}, \mathbf{y}))$ avec $\theta(\mathbf{x}, \mathbf{y})$ l'angle entre les vecteurs \mathbf{x} et \mathbf{y}

$\Rightarrow \theta(\mathbf{x}, \mathbf{y}) = \cos^{-1} \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \right) \rightarrow$ le signe de $\langle \mathbf{x}, \mathbf{y} \rangle$ donne l'orientation entre \mathbf{x} et \mathbf{y}

- si $\langle \mathbf{x}, \mathbf{y} \rangle > 0$, $|\theta(\mathbf{x}, \mathbf{y})| < \frac{\pi}{2}$
 $\rightarrow \mathbf{x}$ et \mathbf{y} pointent dans la même direction.
- si $\langle \mathbf{x}, \mathbf{y} \rangle < 0$, $|\theta(\mathbf{x}, \mathbf{y})| > \frac{\pi}{2}$
 $\rightarrow \mathbf{x}$ et \mathbf{y} pointent dans des directions opposées

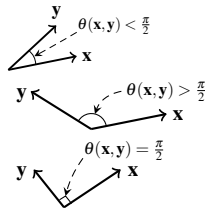


Angle et espace orthogonal

$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos(\theta(\mathbf{x}, \mathbf{y}))$ avec $\theta(\mathbf{x}, \mathbf{y})$ l'angle entre les vecteurs \mathbf{x} et \mathbf{y}

$\Rightarrow \theta(\mathbf{x}, \mathbf{y}) = \cos^{-1} \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \right) \rightarrow$ le signe de $\langle \mathbf{x}, \mathbf{y} \rangle$ donne l'orientation entre \mathbf{x} et \mathbf{y}

- si $\langle \mathbf{x}, \mathbf{y} \rangle > 0$, $|\theta(\mathbf{x}, \mathbf{y})| < \frac{\pi}{2}$
 $\rightarrow \mathbf{x}$ et \mathbf{y} pointent dans la même direction.
- si $\langle \mathbf{x}, \mathbf{y} \rangle < 0$, $|\theta(\mathbf{x}, \mathbf{y})| > \frac{\pi}{2}$
 $\rightarrow \mathbf{x}$ et \mathbf{y} pointent dans des directions opposées
- si $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, $\theta(\mathbf{x}, \mathbf{y}) = \frac{\pi}{2}$
 $\rightarrow \mathbf{x}$ et \mathbf{y} sont orthogonaux

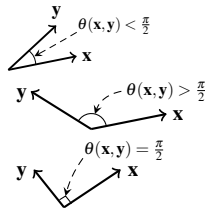


Angle et espace orthogonal

$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos(\theta(\mathbf{x}, \mathbf{y}))$ avec $\theta(\mathbf{x}, \mathbf{y})$ l'angle entre les vecteurs \mathbf{x} et \mathbf{y}

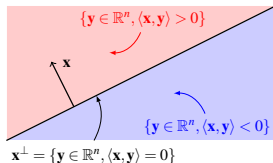
$\Rightarrow \theta(\mathbf{x}, \mathbf{y}) = \cos^{-1} \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \right) \rightarrow$ le signe de $\langle \mathbf{x}, \mathbf{y} \rangle$ donne l'orientation entre \mathbf{x} et \mathbf{y}

- si $\langle \mathbf{x}, \mathbf{y} \rangle > 0$, $|\theta(\mathbf{x}, \mathbf{y})| < \frac{\pi}{2}$
 $\rightarrow \mathbf{x}$ et \mathbf{y} pointent dans la même direction.
- si $\langle \mathbf{x}, \mathbf{y} \rangle < 0$, $|\theta(\mathbf{x}, \mathbf{y})| > \frac{\pi}{2}$
 $\rightarrow \mathbf{x}$ et \mathbf{y} pointent dans des directions opposées
- si $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, $\theta(\mathbf{x}, \mathbf{y}) = \frac{\pi}{2}$
 $\rightarrow \mathbf{x}$ et \mathbf{y} sont orthogonaux



$\mathbf{x}^\perp = \{\mathbf{y} \in \mathbb{R}^n, \langle \mathbf{x}, \mathbf{y} \rangle = 0\}$ définit l'espace orthogonal à \mathbf{x} et divise \mathbb{R}^n en deux demi-espaces :

- \rightarrow demi-espace **positif** (où pointe \mathbf{x})
- \rightarrow demi-espace **négatif** (opposé d'où pointe \mathbf{x})



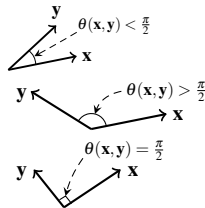
D'une manière générale pour un ensemble $E \subset \mathbb{R}^n$, on appelle orthogonal de E l'ensemble $E^\perp = \{\mathbf{y} \in \mathbb{R}^n \text{ tq } \forall \mathbf{x} \in E, \langle \mathbf{x}, \mathbf{y} \rangle = 0\}$

Angle et espace orthogonal

$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos(\theta(\mathbf{x}, \mathbf{y}))$ avec $\theta(\mathbf{x}, \mathbf{y})$ l'angle entre les vecteurs \mathbf{x} et \mathbf{y}

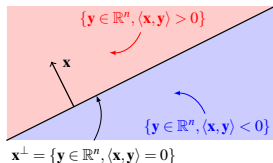
$\Rightarrow \theta(\mathbf{x}, \mathbf{y}) = \cos^{-1} \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \right) \rightarrow$ le signe de $\langle \mathbf{x}, \mathbf{y} \rangle$ donne l'orientation entre \mathbf{x} et \mathbf{y}

- si $\langle \mathbf{x}, \mathbf{y} \rangle > 0$, $|\theta(\mathbf{x}, \mathbf{y})| < \frac{\pi}{2}$
 $\rightarrow \mathbf{x}$ et \mathbf{y} pointent dans la même direction.
- si $\langle \mathbf{x}, \mathbf{y} \rangle < 0$, $|\theta(\mathbf{x}, \mathbf{y})| > \frac{\pi}{2}$
 $\rightarrow \mathbf{x}$ et \mathbf{y} pointent dans des directions opposées
- si $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, $\theta(\mathbf{x}, \mathbf{y}) = \frac{\pi}{2}$
 $\rightarrow \mathbf{x}$ et \mathbf{y} sont orthogonaux



$\mathbf{x}^\perp = \{\mathbf{y} \in \mathbb{R}^n, \langle \mathbf{x}, \mathbf{y} \rangle = 0\}$ définit l'espace orthogonal à \mathbf{x} et divise \mathbb{R}^n en deux demi-espaces :

- \rightarrow demi-espace **positif** (où pointe \mathbf{x})
- \rightarrow demi-espace **négatif** (opposé d'où pointe \mathbf{x})



D'une manière générale pour un ensemble $E \subset \mathbb{R}^n$, on appelle orthogonal de E l'ensemble $E^\perp = \{\mathbf{y} \in \mathbb{R}^n \text{ tq } \forall \mathbf{x} \in E, \langle \mathbf{x}, \mathbf{y} \rangle = 0\}$

Inégalité de Cauchy-Schwartz : $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, |\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$

Norme, distance et boule unité

Une *norme* est une fonction $\| \cdot \| : \mathbb{R}^n \rightarrow \mathbb{R}^+$ qui vérifie les propriétés de :

séparation :

homogénéité :

inégalité triangulaire :

Norme, distance et boule unité

Une *norme* est une fonction $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}^+$ qui vérifie les propriétés de :

séparation : $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}_n$

homogénéité : $\forall \mathbf{x} \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}, \|\lambda \mathbf{x}\| = |\lambda| \|\mathbf{x}\|$

inégalité triangulaire : $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

Norme, distance et boule unité

Une *norme* est une fonction $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}^+$ qui vérifie les propriétés de :

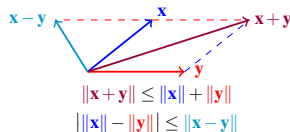
séparation : $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}_n$

homogénéité : $\forall \mathbf{x} \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}, \|\lambda \mathbf{x}\| = |\lambda| \|\mathbf{x}\|$

inégalité triangulaire : $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

→ permet de déduire l'inégalité triangulaire inversée :

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \left| \|\mathbf{x}\| - \|\mathbf{y}\| \right| \leq \|\mathbf{x} - \mathbf{y}\|$$



Norme, distance et boule unité

Une *norme* est une fonction $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}^+$ qui vérifie les propriétés de :

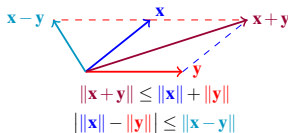
séparation : $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}_n$

homogénéité : $\forall \mathbf{x} \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}, \|\lambda \mathbf{x}\| = |\lambda| \|\mathbf{x}\|$

inégalité triangulaire : $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

→ permet de déduire l'inégalité triangulaire inversée :

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \left| \|\mathbf{x}\| - \|\mathbf{y}\| \right| \leq \|\mathbf{x} - \mathbf{y}\|$$



On appelle *distance* entre \mathbf{x} et \mathbf{y} la quantité $d_{\|\cdot\|}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$

→ $\|\mathbf{x}\| = d_{\|\cdot\|}(\mathbf{x}, \mathbf{0}_n)$: la norme de \mathbf{x} est sa distance par rapport à l'origine $\mathbf{0}_n$

→ La distance dépend de la norme. Deux normes différentes $\|\cdot\|_\alpha$ et $\|\cdot\|_\beta$ induisent deux distances différentes $d_{\|\cdot\|_\alpha}(\mathbf{x}, \mathbf{y}) \neq d_{\|\cdot\|_\beta}(\mathbf{x}, \mathbf{y})$ pour les deux mêmes vecteurs \mathbf{x} et \mathbf{y} .

Norme, distance et boule unité

Une *norme* est une fonction $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}^+$ qui vérifie les propriétés de :

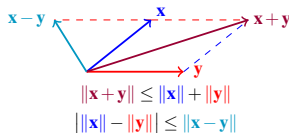
séparation : $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}_n$

homogénéité : $\forall \mathbf{x} \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}, \|\lambda \mathbf{x}\| = |\lambda| \|\mathbf{x}\|$

inégalité triangulaire : $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

→ permet de déduire l'inégalité triangulaire inversée :

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \left| \|\mathbf{x}\| - \|\mathbf{y}\| \right| \leq \|\mathbf{x} - \mathbf{y}\|$$



On appelle *distance* entre \mathbf{x} et \mathbf{y} la quantité $d_{\|\cdot\|}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$

→ $\|\mathbf{x}\| = d_{\|\cdot\|}(\mathbf{x}, \mathbf{0}_n)$: la norme de \mathbf{x} est sa distance par rapport à l'origine $\mathbf{0}_n$

→ La distance dépend de la norme. Deux normes différentes $\|\cdot\|_\alpha$ et $\|\cdot\|_\beta$ induisent deux distances différentes $d_{\|\cdot\|_\alpha}(\mathbf{x}, \mathbf{y}) \neq d_{\|\cdot\|_\beta}(\mathbf{x}, \mathbf{y})$ pour les deux mêmes vecteurs \mathbf{x} et \mathbf{y} .

On appelle *boule* centrée en \mathbf{x} et de rayon $r > 0$ pour la norme $\|\cdot\|$ l'ensemble

$\mathcal{B}_{\|\cdot\|}(\mathbf{x}, r) = \{\mathbf{y} \in \mathbb{R}^n, \|\mathbf{x} - \mathbf{y}\| \leq r\}$ → on parle de boule *fermée* si $\leq r$

→ on parle de boule *ouverte* si $< r$

L'ensemble $\mathcal{B}_{\|\cdot\|}(\mathbf{0}_n, 1) = \{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\| \leq 1\}$ est appelé *boule unité* pour la norme $\|\cdot\|$.

Normes de Hölder et distances associées

On appelle norme de Hölder d'indice $p \geq 1$, ou norme \mathcal{L}_p , la quantité $\|\mathbf{x}\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p}$

Les cas particuliers fréquemment utilisés correspondent aux indices :

- $p = 1$: $\mathcal{L}_1(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$
- $p = 2$: $\mathcal{L}_2(\mathbf{x}) = \|\mathbf{x}\|_2 =$ norme euclidienne usuelle
- $p = \infty$: $\mathcal{L}_\infty(\mathbf{x}) = \|\mathbf{x}\|_\infty = \max_{i=1,\dots,n} |x_i|$

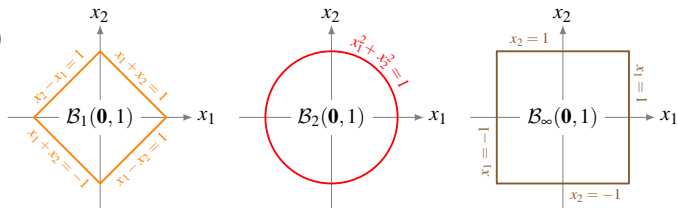
Normes de Hölder et distances associées

On appelle norme de Hölder d'indice $p \geq 1$, ou norme \mathcal{L}_p , la quantité $\|\mathbf{x}\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p}$

Les cas particuliers fréquemment utilisés correspondent aux indices :

- $p = 1$: $\mathcal{L}_1(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$
- $p = 2$: $\mathcal{L}_2(\mathbf{x}) = \|\mathbf{x}\|_2 =$ norme euclidienne usuelle
- $p = \infty$: $\mathcal{L}_\infty(\mathbf{x}) = \|\mathbf{x}\|_\infty = \max_{i=1,\dots,n} |x_i|$

La forme de $\mathcal{B}_p(\mathbf{0}, 1)$
dépend de l'indice p :



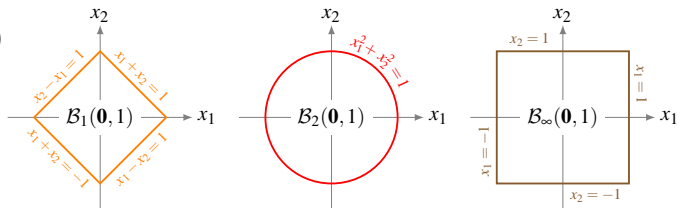
Normes de Hölder et distances associées

On appelle norme de Hölder d'indice $p \geq 1$, ou norme \mathcal{L}_p , la quantité $\|\mathbf{x}\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p}$

Les cas particuliers fréquemment utilisés correspondent aux indices :

- $p = 1$: $\mathcal{L}_1(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$
- $p = 2$: $\mathcal{L}_2(\mathbf{x}) = \|\mathbf{x}\|_2 =$ norme euclidienne usuelle
- $p = \infty$: $\mathcal{L}_\infty(\mathbf{x}) = \|\mathbf{x}\|_\infty = \max_{i=1,\dots,n} |x_i|$

La forme de $\mathcal{B}_p(\mathbf{0}, 1)$
dépend de l'indice p :

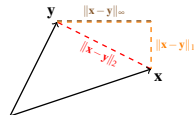


→ Distances associées aux normes 1, 2, et ∞ :

$\|\mathbf{x} - \mathbf{y}\|_1 = \sum_i |x_i - y_i|$: distance de Manhattan

$\|\mathbf{x} - \mathbf{y}\|_2$: distance euclidienne usuelle

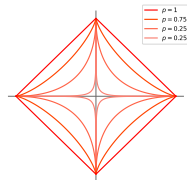
$\|\mathbf{x} - \mathbf{y}\|_\infty = \max_{i=1,\dots,n} |x_i - y_i|$: distance de Chebyshev



Et si $p < 1$?

Pour $0 < p < 1$

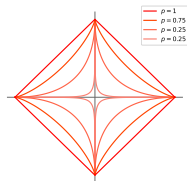
Si l'indice p dans la norme de Hölder est $0 < p < 1$, l'inégalité triangulaire n'est plus respectée : $\|\mathbf{x}\|_p$ n'est pas une norme, mais une *quasi-norme*.



Et si $p < 1$?

Pour $0 < p < 1$

Si l'indice p dans la norme de Hölder est $0 < p < 1$, l'inégalité triangulaire n'est plus respectée : $\|\mathbf{x}\|_p$ n'est pas une norme, mais une *quasi-norme*.

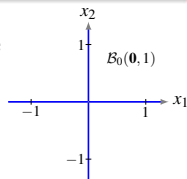


Pour $p = 0$

Il est d'usage de définir la "norme" \mathcal{L}_0 d'un vecteur \mathbf{x} comme le nombre de coordonnées non nulles de \mathbf{x} : $\|\mathbf{x}\|_0 = \sum_i |x_i|^0$ (avec $0^0 = 0$)

→ Propriété d'homogénéité non respectée.

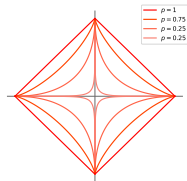
→ Distance associée : $\|\mathbf{x} - \mathbf{y}\|_0 \equiv$ distance de Hamming.



Et si $p < 1$?

Pour $0 < p < 1$

Si l'indice p dans la norme de Hölder est $0 < p < 1$, l'inégalité triangulaire n'est plus respectée : $\|\mathbf{x}\|_p$ n'est pas une norme, mais une *quasi-norme*.

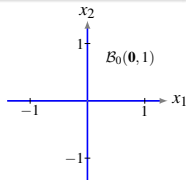


Pour $p = 0$

Il est d'usage de définir la "norme" \mathcal{L}_0 d'un vecteur \mathbf{x} comme le nombre de coordonnées non nulles de \mathbf{x} : $\|\mathbf{x}\|_0 = \sum_i |x_i|^0$ (avec $0^0 = 0$)

→ Propriété d'homogénéité non respectée.

→ Distance associée : $\|\mathbf{x} - \mathbf{y}\|_0 \equiv$ distance de Hamming.



$\mathbf{y} \in \mathbb{R}^m$ $\mathbf{A} \in \mathbb{R}^{m \times n}$ $\mathbf{x} \in \mathbb{R}^n$

\mathbf{y} = observation
 \mathbf{A} = dictionnaire
 \mathbf{x} = inconnue

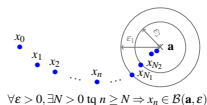
$\min \|\mathbf{x}\|_0 \text{ tq } \mathbf{y} = \mathbf{Ax}$
 $\min \|\mathbf{y} - \mathbf{Ax}\|_2^2 \text{ tq } \|\mathbf{x}\|_0 \leq N$

La norme \mathcal{L}_0 a un grand nombre d'applications en traitement du signal (reconstruction parcimonieuse) et en machine learning (régularisation de modèles).

→ La minimisation est NP-difficile 😞

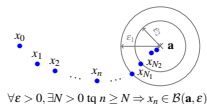
Normes équivalentes

Les normes induisent des topologies différentes en fonction de la forme de leur boule unité \rightarrow certains comportements locaux/asymptotiques (limite d'une suite de \mathbb{R}^n , continuité d'une fonction en un point de \mathbb{R}^n) peuvent-ils être impactés par le choix de la norme ? 🤔



Normes équivalentes

Les normes induisent des topologies différentes en fonction de la forme de leur boule unité \rightarrow certains comportements locaux/asymptotiques (limite d'une suite de \mathbb{R}^n , continuité d'une fonction en un point de \mathbb{R}^n) peuvent-ils être impactés par le choix de la norme ? 🤖



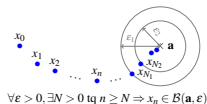
Normes équivalentes

Deux normes $\|\cdot\|_\alpha$ et $\|\cdot\|_\beta$ sont *équivalentes* s'il existe des constantes $A > 0$ et $B > 0$ telles que $\forall \mathbf{x} \in \mathbb{R}^n, A\|\mathbf{x}\|_\alpha \leq \|\mathbf{x}\|_\beta \leq B\|\mathbf{x}\|_\alpha$

Deux normes équivalentes induisent la même topologie

Normes équivalentes

Les normes induisent des topologies différentes en fonction de la forme de leur boule unité \rightarrow certains comportements locaux/asymptotiques (limite d'une suite de \mathbb{R}^n , continuité d'une fonction en un point de \mathbb{R}^n) peuvent-ils être impactés par le choix de la norme ? 🤖

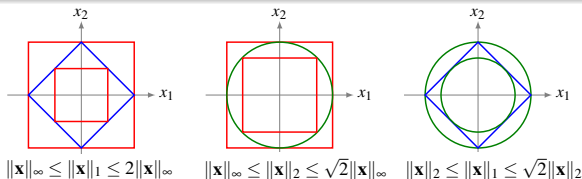


Normes équivalentes

Deux normes $\|\cdot\|_\alpha$ et $\|\cdot\|_\beta$ sont *équivalentes* s'il existe des constantes $A > 0$ et $B > 0$ telles que $\forall \mathbf{x} \in \mathbb{R}^n, A\|\mathbf{x}\|_\alpha \leq \|\mathbf{x}\|_\beta \leq B\|\mathbf{x}\|_\alpha$

Deux normes équivalentes induisent la même topologie

Théorème : Toutes les normes sur \mathbb{R}^n sont équivalentes

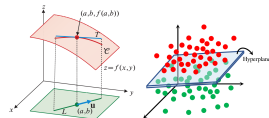


\Rightarrow Pas besoin de se soucier de l'identité de la norme dans \mathbb{R}^n , si ça converge pour l'une d'entre elles, ça converge pour toutes les autres !

Description des sous espaces de \mathbb{R}^n

La description des sous-espaces $\mathbb{R}^p \subset \mathbb{R}^n$ intervient fréquemment lors de la formulation et la résolution d'un problème d'optimisation :

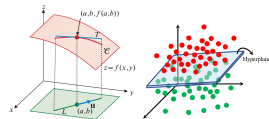
- droite : dérivée directionnelle, direction de descente, etc
- hyperplan : description d'une séparatrice, etc
- le reste : description explicite des potentielles contraintes



Description des sous espaces de \mathbb{R}^n

La description des sous-espaces $\mathbb{R}^p \subset \mathbb{R}^n$ intervient fréquemment lors de la formulation et la résolution d'un problème d'optimisation :

- droite : dérivée directionnelle, direction de descente, etc
- hyperplan : description d'une séparatrice, etc
- le reste : description explicite des potentielles contraintes



On se restreint ici uniquement à la description des droites et des hyperplans de \mathbb{R}^n

Hyperplan (rappel)

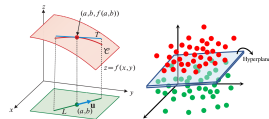
Un hyperplan \mathcal{H} de \mathbb{R}^n est un sous-espace de dimension $(n - 1)$

- dans \mathbb{R}^2 , les hyperplans sont des droites
- dans \mathbb{R}^3 , les hyperplans sont des plans
- L'espace orthogonal \mathcal{H}^\perp d'un hyperplan \mathcal{H} est une droite (et vice versa)

Description des sous espaces de \mathbb{R}^n

La description des sous-espaces $\mathbb{R}^p \subset \mathbb{R}^n$ intervient fréquemment lors de la formulation et la résolution d'un problème d'optimisation :

- droite : dérivée directionnelle, direction de descente, etc
- hyperplan : description d'une séparatrice, etc
- le reste : description explicite des potentielles contraintes



On se restreint ici uniquement à la description des droites et des hyperplans de \mathbb{R}^n

Hyperplan (rappel)

Un hyperplan \mathcal{H} de \mathbb{R}^n est un sous-espace de dimension $(n - 1)$

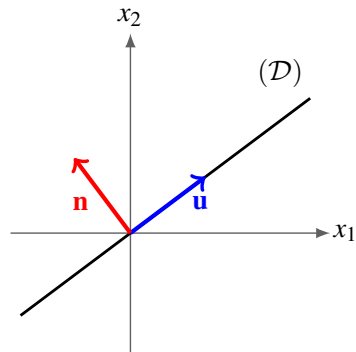
- dans \mathbb{R}^2 , les hyperplans sont des droites
- dans \mathbb{R}^3 , les hyperplans sont des plans
- L'espace orthogonal \mathcal{H}^\perp d'un hyperplan \mathcal{H} est une droite (et vice versa)

Deux représentations possibles : représentation *paramétrique* et représentation *implicite*

Sous espaces de \mathbb{R}^2

Droite linéaire

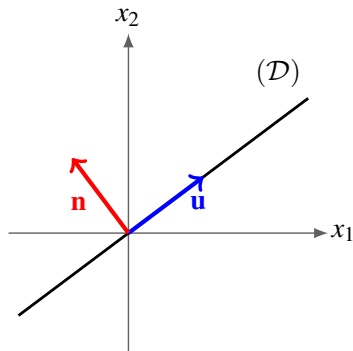
Une droite linéaire (\mathcal{D}) peut être décrite de deux façons : par son vecteur directeur **u** ou par son vecteur normal **n**.



Sous espaces de \mathbb{R}^2

Droite linéaire

Une droite linéaire (\mathcal{D}) peut être décrite de deux façons : par son vecteur directeur \mathbf{u} ou par son vecteur normal \mathbf{n} .



Représentation paramétrique

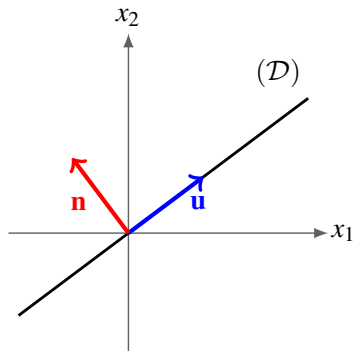
Elle fait intervenir le vecteur directeur \mathbf{u}

- (\mathcal{D}) est l'ensemble des vecteurs \mathbf{x} colinéaires à \mathbf{u}
- $(\mathcal{D}) = \{\mathbf{x} \in \mathbb{R}^2, \mathbf{x} = \lambda \mathbf{u}, \lambda \in \mathbb{R}\}$
- (\mathcal{D}) est l'image de l'application linéaire $\lambda \mapsto \lambda \mathbf{u}$
- besoin de 1 paramètre ($\lambda \in \mathbb{R}$) pour décrire (\mathcal{D})

Sous espaces de \mathbb{R}^2

Droite linéaire

Une droite linéaire (\mathcal{D}) peut être décrite de deux façons : par son vecteur directeur \mathbf{u} ou par son vecteur normal \mathbf{n} .



Représentation paramétrique

Elle fait intervenir le vecteur directeur \mathbf{u}

- (\mathcal{D}) est l'ensemble des vecteurs \mathbf{x} colinéaires à \mathbf{u}
- $(\mathcal{D}) = \{\mathbf{x} \in \mathbb{R}^2, \mathbf{x} = \lambda \mathbf{u}, \lambda \in \mathbb{R}\}$
- (\mathcal{D}) est l'image de l'application linéaire $\lambda \mapsto \lambda \mathbf{u}$
- besoin de 1 paramètre ($\lambda \in \mathbb{R}$) pour décrire (\mathcal{D})

Représentation implicite

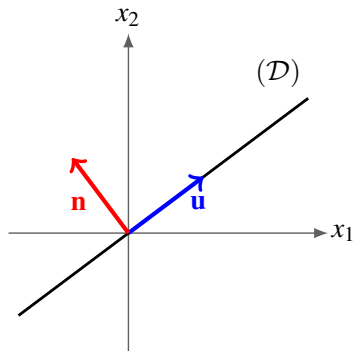
Elle fait intervenir le vecteur normal \mathbf{n}

- (\mathcal{D}) est l'ensemble des vecteurs \mathbf{x} orthogonaux à \mathbf{n}
- $(\mathcal{D}) = \mathbf{n}^\perp = \{\mathbf{x} \in \mathbb{R}^2, \langle \mathbf{n}, \mathbf{x} \rangle = \mathbf{n}^T \mathbf{x} = 0\}$
- (\mathcal{D}) est le noyau de l'application linéaire $\mathbf{x} \mapsto \mathbf{n}^T \mathbf{x}$
- besoin de 1 vecteur ($\mathbf{n} \in \mathbb{R}^2$) pour décrire (\mathcal{D})

Sous espaces de \mathbb{R}^2

Droite linéaire

Une droite linéaire (\mathcal{D}) peut être décrite de deux façons : par son vecteur directeur \mathbf{u} ou par son vecteur normal \mathbf{n} .



Équation cartésienne $ax + by = 0$

\Leftrightarrow

représentation implicite de vecteur
normal $\mathbf{n} = (a, b)^T$

Représentation paramétrique

Elle fait intervenir le vecteur directeur \mathbf{u}

- (\mathcal{D}) est l'ensemble des vecteurs \mathbf{x} colinéaires à \mathbf{u}
- (\mathcal{D}) = $\{\mathbf{x} \in \mathbb{R}^2, \mathbf{x} = \lambda \mathbf{u}, \lambda \in \mathbb{R}\}$
- (\mathcal{D}) est l'image de l'application linéaire $\lambda \mapsto \lambda \mathbf{u}$
- besoin de 1 paramètre ($\lambda \in \mathbb{R}$) pour décrire (\mathcal{D})

Représentation implicite

Elle fait intervenir le vecteur normal \mathbf{n}

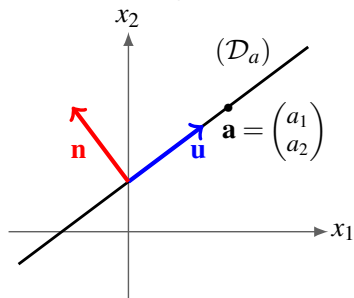
- (\mathcal{D}) est l'ensemble des vecteurs \mathbf{x} orthogonaux à \mathbf{n}
- (\mathcal{D}) = $\mathbf{n}^\perp = \{\mathbf{x} \in \mathbb{R}^2, \langle \mathbf{n}, \mathbf{x} \rangle = \mathbf{n}^T \mathbf{x} = 0\}$
- (\mathcal{D}) est le noyau de l'application linéaire $\mathbf{x} \mapsto \mathbf{n}^T \mathbf{x}$
- besoin de 1 vecteur ($\mathbf{n} \in \mathbb{R}^2$) pour décrire (\mathcal{D})

Sous espaces de \mathbb{R}^2

Droite affine

Un sous-espace $F \subset \mathbb{R}^n$ est dit *affine* s'il existe un sous-espace linéaire $E \subset \mathbb{R}^n$ et un point $\mathbf{a} \in \mathbb{R}^n$ tels que $F = \mathbf{a} + E = \{\mathbf{a} + \mathbf{x}, \mathbf{x} \in E\}$

Une droite affine (\mathcal{D}_a) passant par \mathbf{a} , de vecteur directeur \mathbf{u} et de vecteur normal \mathbf{n} peut être décrite de deux façons :

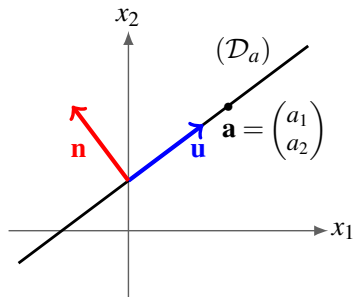


Sous espaces de \mathbb{R}^2

Droite affine

Un sous-espace $F \subset \mathbb{R}^n$ est dit *affine* s'il existe un sous-espace linéaire $E \subset \mathbb{R}^n$ et un point $\mathbf{a} \in \mathbb{R}^n$ tels que $F = \mathbf{a} + E = \{\mathbf{a} + \mathbf{x}, \mathbf{x} \in E\}$

Une droite affine (\mathcal{D}_a) passant par \mathbf{a} , de vecteur directeur \mathbf{u} et de vecteur normal \mathbf{n} peut être décrite de deux façons :



Représentation paramétrique

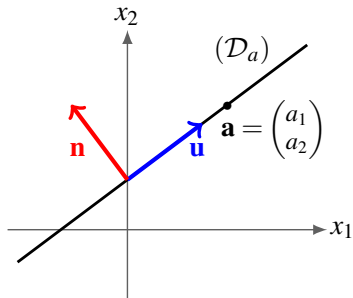
- Elle fait intervenir le point \mathbf{a} et le vecteur directeur \mathbf{u}
- $\rightarrow (\mathcal{D}_a)$ est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et colinéaires à \mathbf{u}
- $\rightarrow (\mathcal{D}_a) = \mathbf{a} + (\mathcal{D}) = \mathbf{a} + \{\lambda \mathbf{u}, \lambda \in \mathbb{R}\} = \{\mathbf{a} + \lambda \mathbf{u}, \lambda \in \mathbb{R}\}$
- $\rightarrow (\mathcal{D}_a)$ est l'image de l'application affine $\lambda \mapsto \mathbf{a} + \lambda \mathbf{u}$
- \rightarrow besoin de 1 paramètre ($\lambda \in \mathbb{R}$) pour décrire (\mathcal{D}_a)

Sous espaces de \mathbb{R}^2

Droite affine

Un sous-espace $F \subset \mathbb{R}^n$ est dit *affine* s'il existe un sous-espace linéaire $E \subset \mathbb{R}^n$ et un point $\mathbf{a} \in \mathbb{R}^n$ tels que $F = \mathbf{a} + E = \{\mathbf{a} + \mathbf{x}, \mathbf{x} \in E\}$

Une droite affine (\mathcal{D}_a) passant par \mathbf{a} , de vecteur directeur \mathbf{u} et de vecteur normal \mathbf{n} peut être décrite de deux façons :



Représentation paramétrique

Elle fait intervenir le point \mathbf{a} et le vecteur directeur \mathbf{u}
→ (\mathcal{D}_a) est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et colinéaires à \mathbf{u}
→ $(\mathcal{D}_a) = \mathbf{a} + (\mathcal{D}) = \mathbf{a} + \{\lambda \mathbf{u}, \lambda \in \mathbb{R}\} = \{\mathbf{a} + \lambda \mathbf{u}, \lambda \in \mathbb{R}\}$
→ (\mathcal{D}_a) est l'image de l'application affine $\lambda \mapsto \mathbf{a} + \lambda \mathbf{u}$
→ besoin de 1 paramètre ($\lambda \in \mathbb{R}$) pour décrire (\mathcal{D}_a)

Représentation implicite

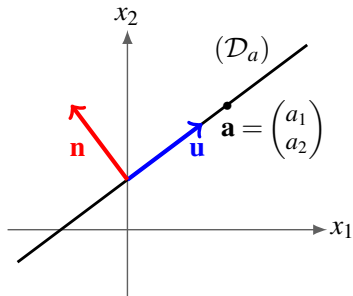
Elle fait intervenir le point \mathbf{a} et vecteur normal \mathbf{n}
→ (\mathcal{D}_a) est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et orthogonaux à \mathbf{n}
→ $(\mathcal{D}_a) = \{\mathbf{x} \in \mathbb{R}^2, \langle \mathbf{n}, \mathbf{x} - \mathbf{a} \rangle = \mathbf{n}^T (\mathbf{x} - \mathbf{a}) = 0\}$
→ (\mathcal{D}_a) est le noyau de l'application affine $\mathbf{x} \mapsto \mathbf{n}^T \mathbf{x} - \mathbf{n}^T \mathbf{a}$
→ besoin de 1 vecteur ($\mathbf{n} \in \mathbb{R}^2$) pour décrire (\mathcal{D}_a)

Sous espaces de \mathbb{R}^2

Droite affine

Un sous-espace $F \subset \mathbb{R}^n$ est dit *affine* s'il existe un sous-espace linéaire $E \subset \mathbb{R}^n$ et un point $\mathbf{a} \in \mathbb{R}^n$ tels que $F = \mathbf{a} + E = \{\mathbf{a} + \mathbf{x}, \mathbf{x} \in E\}$

Une droite affine (\mathcal{D}_a) passant par \mathbf{a} , de vecteur directeur \mathbf{u} et de vecteur normal \mathbf{n} peut être décrite de deux façons :



Équation cartésienne $ax + by + c = 0$

⇔

représentation implicite de vecteur normal $\mathbf{n} = (a, b)^T$ et passant par $\mathbf{a} = (0, -\frac{c}{b})^T$

Représentation paramétrique

Elle fait intervenir le point \mathbf{a} et le vecteur directeur \mathbf{u}
→ (\mathcal{D}_a) est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et colinéaires à \mathbf{u}
→ $(\mathcal{D}_a) = \mathbf{a} + (\mathcal{D}) = \mathbf{a} + \{\lambda \mathbf{u}, \lambda \in \mathbb{R}\} = \{\mathbf{a} + \lambda \mathbf{u}, \lambda \in \mathbb{R}\}$
→ (\mathcal{D}_a) est l'image de l'application affine $\lambda \mapsto \mathbf{a} + \lambda \mathbf{u}$
→ besoin de 1 paramètre ($\lambda \in \mathbb{R}$) pour décrire (\mathcal{D}_a)

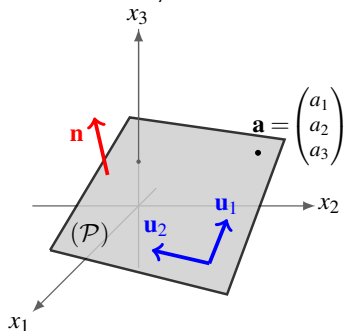
Représentation implicite

Elle fait intervenir le point \mathbf{a} et vecteur normal \mathbf{n}
→ (\mathcal{D}_a) est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et orthogonaux à \mathbf{n}
→ $(\mathcal{D}_a) = \{\mathbf{x} \in \mathbb{R}^2, \langle \mathbf{n}, \mathbf{x} - \mathbf{a} \rangle = \mathbf{n}^T (\mathbf{x} - \mathbf{a}) = 0\}$
→ (\mathcal{D}_a) est le noyau de l'application affine $\mathbf{x} \mapsto \mathbf{n}^T \mathbf{x} - \mathbf{n}^T \mathbf{a}$
→ besoin de 1 vecteur ($\mathbf{n} \in \mathbb{R}^2$) pour décrire (\mathcal{D}_a)

Sous espaces de \mathbb{R}^3

Plan affine

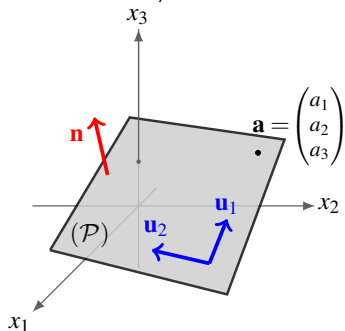
Un plan (\mathcal{P}) passant par \mathbf{a} , de vecteurs générateurs $(\mathbf{u}_1, \mathbf{u}_2)$ et de vecteur normal \mathbf{n} peut être décrit de deux façons :



Sous espaces de \mathbb{R}^3

Plan affine

Un plan (\mathcal{P}) passant par \mathbf{a} , de vecteurs générateurs $(\mathbf{u}_1, \mathbf{u}_2)$ et de vecteur normal \mathbf{n} peut être décrit de deux façons :



Représentation paramétrique

Plan contenant le point \mathbf{a} et engendré par $(\mathbf{u}_1, \mathbf{u}_2)$, famille libre de \mathbb{R}^3

→ (\mathcal{P}) est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et combinaison linéaire de $(\mathbf{u}_1, \mathbf{u}_2)$

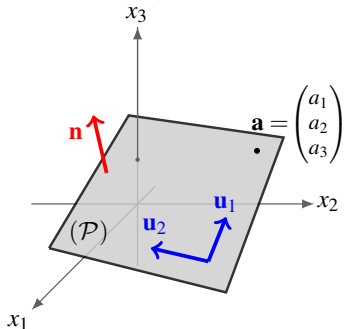
$$\begin{aligned} \rightarrow (\mathcal{P}) &= \mathbf{a} + \{\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2, (\lambda_1, \lambda_2) \in \mathbb{R}^2\} \\ &= \{\mathbf{a} + \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2, (\lambda_1, \lambda_2) \in \mathbb{R}^2\} \end{aligned}$$

→ besoin de 2 paramètres $(\lambda_1, \lambda_2) \in \mathbb{R}^2$ pour décrire (\mathcal{P})

Sous espaces de \mathbb{R}^3

Plan affine

Un plan (\mathcal{P}) passant par \mathbf{a} , de vecteurs générateurs ($\mathbf{u}_1, \mathbf{u}_2$) et de vecteur normal \mathbf{n} peut être décrit de deux façons :



Représentation paramétrique

Plan contenant le point \mathbf{a} et engendré par ($\mathbf{u}_1, \mathbf{u}_2$), famille libre de \mathbb{R}^3

→ (\mathcal{P}) est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et combinaison linéaire de ($\mathbf{u}_1, \mathbf{u}_2$)

$$\begin{aligned} \rightarrow (\mathcal{P}) &= \mathbf{a} + \{\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2, (\lambda_1, \lambda_2) \in \mathbb{R}^2\} \\ &= \{\mathbf{a} + \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2, (\lambda_1, \lambda_2) \in \mathbb{R}^2\} \end{aligned}$$

→ besoin de 2 paramètres $(\lambda_1, \lambda_2) \in \mathbb{R}^2$ pour décrire (\mathcal{P})

Représentation implicite

Plan contenant le point \mathbf{a} et de vecteur normal \mathbf{n}

→ (\mathcal{P}) est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et orthogonaux à \mathbf{n}

$$\rightarrow (\mathcal{P}) = \{\mathbf{x} \in \mathbb{R}^3, \langle \mathbf{n}, \mathbf{x} - \mathbf{a} \rangle = \mathbf{n}^T (\mathbf{x} - \mathbf{a}) = 0\}$$

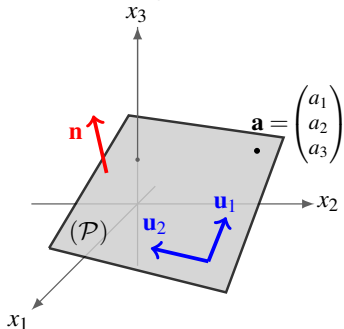
→ (\mathcal{P}) a la **même** écriture implicite que (\mathcal{D}_a) dans \mathbb{R}^2

→ besoin de 1 vecteur (\mathbf{n}) pour décrire (\mathcal{P})

Sous espaces de \mathbb{R}^3

Plan affine

Un plan (\mathcal{P}) passant par \mathbf{a} , de vecteurs générateurs ($\mathbf{u}_1, \mathbf{u}_2$) et de vecteur normal \mathbf{n} peut être décrit de deux façons :



Équation cartésienne
 $ax + by + cz + d = 0$

\Leftrightarrow

représentation implicite de vecteur
normal $\mathbf{n} = (a, b, c)^T$ et passant par
 $\mathbf{a} = \left(0, 0, -\frac{d}{c}\right)^T$

Représentation paramétrique

Plan contenant le point \mathbf{a} et engendré par ($\mathbf{u}_1, \mathbf{u}_2$), famille libre de \mathbb{R}^3

$\rightarrow (\mathcal{P})$ est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et combinaison linéaire de ($\mathbf{u}_1, \mathbf{u}_2$)

$$\rightarrow (\mathcal{P}) = \mathbf{a} + \{\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2, (\lambda_1, \lambda_2) \in \mathbb{R}^2\}$$
$$= \{\mathbf{a} + \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2, (\lambda_1, \lambda_2) \in \mathbb{R}^2\}$$

\rightarrow besoin de 2 paramètres $(\lambda_1, \lambda_2) \in \mathbb{R}^2$ pour décrire (\mathcal{P})

Représentation implicite

Plan contenant le point \mathbf{a} et de vecteur normal \mathbf{n}

$\rightarrow (\mathcal{P})$ est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et orthogonaux à \mathbf{n}

$$\rightarrow (\mathcal{P}) = \{\mathbf{x} \in \mathbb{R}^3, \langle \mathbf{n}, \mathbf{x} - \mathbf{a} \rangle = \mathbf{n}^T (\mathbf{x} - \mathbf{a}) = 0\}$$

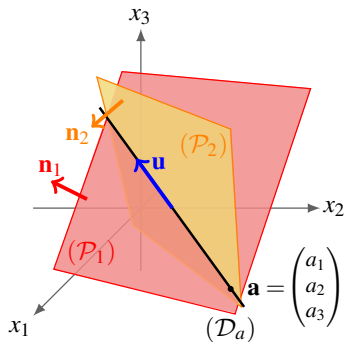
$\rightarrow (\mathcal{P})$ a la même écriture implicite que (\mathcal{D}_a) dans \mathbb{R}^2

\rightarrow besoin de 1 vecteur (\mathbf{n}) pour décrire (\mathcal{P})

Sous espaces de \mathbb{R}^3

Droite affine

Une droite affine (\mathcal{D}_a) passant par \mathbf{a} et de vecteur directeur \mathbf{u} peut être décrite de deux façons :



Sous espaces de \mathbb{R}^3

Droite affine

Une droite affine (\mathcal{D}_a) passant par \mathbf{a} et de vecteur directeur \mathbf{u} peut être décrite de deux façons :

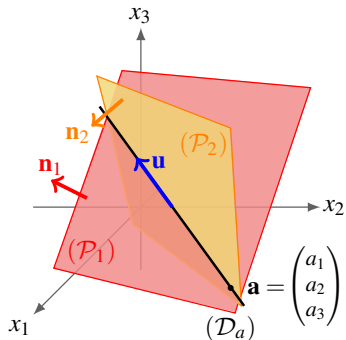
Représentation paramétrique

→ (\mathcal{D}_a) est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et colinéaires à \mathbf{u}

→ $(\mathcal{D}_a) = \mathbf{a} + \{\lambda \mathbf{u}, \lambda \in \mathbb{R}\} = \{\mathbf{a} + \lambda \mathbf{u}, \lambda \in \mathbb{R}\}$

→ (\mathcal{D}_a) a la **même** écriture paramétrique que dans \mathbb{R}^2

→ besoin de 1 paramètre ($\lambda \in \mathbb{R}$) pour décrire (\mathcal{D}_a)



Sous espaces de \mathbb{R}^3

Droite affine

Une droite affine (\mathcal{D}_a) passant par \mathbf{a} et de vecteur directeur \mathbf{u} peut être décrite de deux façons :

Représentation paramétrique

→ (\mathcal{D}_a) est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et colinéaires à \mathbf{u}

→ $(\mathcal{D}_a) = \mathbf{a} + \{\lambda \mathbf{u}, \lambda \in \mathbb{R}\} = \{\mathbf{a} + \lambda \mathbf{u}, \lambda \in \mathbb{R}\}$

→ (\mathcal{D}_a) a la **même** écriture paramétrique que dans \mathbb{R}^2

→ besoin de 1 paramètre ($\lambda \in \mathbb{R}$) pour décrire (\mathcal{D}_a)

Représentation implicite

→ (\mathcal{D}_a) est l'intersection de deux (hyper)plans (\mathcal{P}_1) et (\mathcal{P}_2) contenant le point \mathbf{a} et de vecteurs normaux \mathbf{n}_1 et \mathbf{n}_2

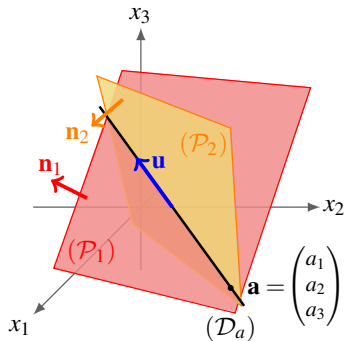
→ $(\mathcal{P}_1) : \mathbf{n}_1^T(\mathbf{x} - \mathbf{a}) = 0$ et $(\mathcal{P}_2) : \mathbf{n}_2^T(\mathbf{x} - \mathbf{a}) = 0$

→ $\mathbf{x} \in \mathbb{R}^3 \in (\mathcal{D}_a) \Leftrightarrow \begin{cases} \mathbf{n}_1^T(\mathbf{x} - \mathbf{a}) = 0 \\ \mathbf{n}_2^T(\mathbf{x} - \mathbf{a}) = 0 \end{cases} \Leftrightarrow \mathbf{N}(\mathbf{x} - \mathbf{a}) = 0$

avec $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2]^T \in \mathbb{R}^{2 \times 3}$

→ (\mathcal{D}_a) est le noyau de l'application affine $\mathbf{x} \mapsto \mathbf{N}(\mathbf{x} - \mathbf{a})$

→ besoin de 2 vecteurs ($\mathbf{n}_1, \mathbf{n}_2 \in \mathbb{R}^3$) pour décrire (\mathcal{D}_a)



Sous espaces de \mathbb{R}^n

Cas général

Pour un sous-espace $E \subset \mathbb{R}^n$ de dimension $p < n$ et contenant un point $\mathbf{a} \in \mathbb{R}^n$

Sous espaces de \mathbb{R}^n

Cas général

Pour un sous-espace $E \subset \mathbb{R}^n$ de dimension $p < n$ et contenant un point $\mathbf{a} \in \mathbb{R}^n$

Représentation paramétrique

→ E est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et combinaison linéaire de p vecteurs générateurs $(\mathbf{u}_1, \dots, \mathbf{u}_p)$ de E : $E = \mathbf{a} + \text{Vect}(\mathbf{u}_1, \dots, \mathbf{u}_p) = \mathbf{a} + \left\{ \sum_{i=1}^p \lambda_i \mathbf{u}_i, (\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p \right\}$

→ E est l'image de l'application affine $\boldsymbol{\lambda} \mapsto \mathbf{a} + \mathbf{U}\boldsymbol{\lambda}$ avec $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p] \in \mathbb{R}^{n \times p}$ et $\boldsymbol{\lambda} \in \mathbb{R}^p$

→ Besoin de $p = \dim(E)$ paramètres pour la représentation paramétrique:

👍 la description d'une droite affine nécessite un seul paramètre λ : $\mathcal{D}_a = \mathbf{a} + \{\lambda \mathbf{u}, \lambda \in \mathbb{R}\}$

👎 la description d'un hyperplan nécessite $(n - 1)$ paramètres $(\lambda_1, \dots, \lambda_{n-1})$

Sous espaces de \mathbb{R}^n

Cas général

Pour un sous-espace $E \subset \mathbb{R}^n$ de dimension $p < n$ et contenant un point $\mathbf{a} \in \mathbb{R}^n$

Représentation paramétrique

→ E est l'ensemble des vecteurs \mathbf{x} d'origine \mathbf{a} et combinaison linéaire de p vecteurs générateurs $(\mathbf{u}_1, \dots, \mathbf{u}_p)$ de E : $E = \mathbf{a} + \text{Vect}(\mathbf{u}_1, \dots, \mathbf{u}_p) = \mathbf{a} + \left\{ \sum_{i=1}^p \lambda_i \mathbf{u}_i, (\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p \right\}$

→ E est l'image de l'application affine $\boldsymbol{\lambda} \mapsto \mathbf{a} + \mathbf{U}\boldsymbol{\lambda}$ avec $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p] \in \mathbb{R}^{n \times p}$ et $\boldsymbol{\lambda} \in \mathbb{R}^p$

→ Besoin de $p = \dim(E)$ paramètres pour la représentation paramétrique:

👍 la description d'une droite affine nécessite un seul paramètre λ : $\mathcal{D}_a = \mathbf{a} + \{\lambda \mathbf{u}, \lambda \in \mathbb{R}\}$

👎 la description d'un hyperplan nécessite $(n - 1)$ paramètres $(\lambda_1, \dots, \lambda_{n-1})$

Représentation implicite

→ E est l'intersection de $(n - p)$ hyperplans $\mathcal{H}_1, \dots, \mathcal{H}_{n-p}$ contenant \mathbf{a} et de vecteurs normaux $(\mathbf{n}_1, \dots, \mathbf{n}_{n-p})$: $E = \mathcal{H}_1 \cap \dots \cap \mathcal{H}_{n-p} : \mathbf{x} \in E \Leftrightarrow (\mathbf{x} \in \mathcal{H}_1) \cap \dots \cap (\mathbf{x} \in \mathcal{H}_{n-p})$

→ E est le noyau de l'application affine $\mathbf{x} \mapsto \mathbf{N}(\mathbf{x} - \mathbf{a})$ avec $\mathbf{N} = [\mathbf{n}_1, \dots, \mathbf{n}_{n-p}]^T \in \mathbb{R}^{(n-p) \times n}$

→ Besoin de $(n - p) = \dim(E^\perp)$ vecteurs normaux pour la représentation implicite:

👍 la description d'un hyperplan nécessite un seul vecteur normal \mathbf{n} : $\mathcal{H} : \mathbf{n}^T(\mathbf{x} - \mathbf{a}) = 0$

(en machine learning, un hyperplan se note classiquement $\mathbf{w}^T \mathbf{x} + b = 0$)

👎 la description d'une droite affine nécessite $(n - 1)$ vecteurs normaux $(\mathbf{n}_1, \dots, \mathbf{n}_{n-1})$

Un petit tour du côté des matrices

Valeurs propres et vecteurs propres

Vecteurs et valeurs propres, diagonalisation et spectre (rappels)

Pour une matrice $\mathbf{A} \in \mathbb{R}^{n \times n}$, on appelle *vecteur propre* \mathbf{u} associé à la *valeur propre* λ le couple (\mathbf{u}, λ) tel que $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$.

On dit que \mathbf{A} est *diagonalisable* s'il existe une matrice $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ et une matrice inversible $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n] \in \mathbb{R}^{n \times n}$ tels que $\mathbf{\Lambda} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$.

La $i^{\text{ème}}$ colonne \mathbf{u}_i de \mathbf{U} est un vecteur propre associé à la $i^{\text{ème}}$ valeur propre λ_i de la diagonale de $\mathbf{\Lambda}$.

$(\lambda_1, \dots, \lambda_n)$ s'appelle le *spectre* de \mathbf{A} , et se note $Sp(\mathbf{A})$

Un petit tour du côté des matrices

Valeurs propres et vecteurs propres

Vecteurs et valeurs propres, diagonalisation et spectre (rappels)

Pour une matrice $\mathbf{A} \in \mathbb{R}^{n \times n}$, on appelle *vecteur propre* \mathbf{u} associé à la *valeur propre* λ le couple (\mathbf{u}, λ) tel que $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$.

On dit que \mathbf{A} est *diagonalisable* s'il existe une matrice $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ et une matrice inversible $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_n] \in \mathbb{R}^{n \times n}$ tels que $\mathbf{\Lambda} = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$.

La $i^{\text{ème}}$ colonne \mathbf{u}_i de \mathbf{U} est un vecteur propre associé à la $i^{\text{ème}}$ valeur propre λ_i de la diagonale de $\mathbf{\Lambda}$.

$(\lambda_1, \dots, \lambda_n)$ s'appelle le *spectre* de \mathbf{A} , et se note $Sp(\mathbf{A})$

Propriétés

→ Toute matrice symétrique ($\mathbf{A} = \mathbf{A}^T$) est (ortho)diagonalisable (\mathbf{U} unitaire : $\mathbf{U}^{-1} = \mathbf{U}^T$)

→ $\begin{cases} \text{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i \\ \det(\mathbf{A}) = \prod_{i=1}^n \lambda_i \end{cases} \quad \rightarrow \text{en particulier lorsque } n = 2 : \begin{cases} \text{tr}(\mathbf{A}) = \lambda_1 + \lambda_2 \\ \det(\mathbf{A}) = \lambda_1 \lambda_2 \end{cases}$

→ \mathbf{A} inversible ($\det(\mathbf{A}) \neq 0$) $\Leftrightarrow 0 \notin Sp(\mathbf{A})$

Un petit tour du côté des matrices

Le cas particulier des matrices symétriques

Positivité/négativité d'une matrice symétrique

On dit qu'une matrice $\mathbf{A} \in \mathbb{R}^{n \times n}$ symétrique est :

- (semi-définie) positive ssi $\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0 \Leftrightarrow Sp(\mathbf{A}) \subset \mathbb{R}^+$
Notation usuelle : $\mathbf{A} \succeq 0$
- définie positive ssi $\forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}_n\}, \mathbf{x}^T \mathbf{A} \mathbf{x} > 0 \Leftrightarrow Sp(\mathbf{A}) \subset \mathbb{R}_*^+$
 $\rightarrow \mathbf{A}$ est positive et inversible
Notation usuelle : $\mathbf{A} \succ 0$
- (semi-définie) négative ssi $\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x}^T \mathbf{A} \mathbf{x} \leq 0 \Leftrightarrow Sp(\mathbf{A}) \subset \mathbb{R}^-$
Notation usuelle : $\mathbf{A} \preceq 0$
- définie négative ssi $\forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}_n\}, \mathbf{x}^T \mathbf{A} \mathbf{x} < 0 \Leftrightarrow Sp(\mathbf{A}) \subset \mathbb{R}_*^-$
 $\rightarrow \mathbf{A}$ est négative et inversible
Notation usuelle : $\mathbf{A} \prec 0$

Un petit tour du côté des matrices

Le cas particulier des matrices symétriques

Positivité/négativité d'une matrice symétrique

On dit qu'une matrice $\mathbf{A} \in \mathbb{R}^{n \times n}$ symétrique est :

- (semi-définie) positive ssi $\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0 \Leftrightarrow Sp(\mathbf{A}) \subset \mathbb{R}^+$
Notation usuelle : $\mathbf{A} \succeq 0$
- définie positive ssi $\forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}_n\}, \mathbf{x}^T \mathbf{A} \mathbf{x} > 0 \Leftrightarrow Sp(\mathbf{A}) \subset \mathbb{R}_*^+$
 $\rightarrow \mathbf{A}$ est positive et inversible
Notation usuelle : $\mathbf{A} \succ 0$
- (semi-définie) négative ssi $\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x}^T \mathbf{A} \mathbf{x} \leq 0 \Leftrightarrow Sp(\mathbf{A}) \subset \mathbb{R}^-$
Notation usuelle : $\mathbf{A} \preceq 0$
- définie négative ssi $\forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}_n\}, \mathbf{x}^T \mathbf{A} \mathbf{x} < 0 \Leftrightarrow Sp(\mathbf{A}) \subset \mathbb{R}_*^-$
 $\rightarrow \mathbf{A}$ est négative et inversible
Notation usuelle : $\mathbf{A} \prec 0$

Lien entre matrice symétrique et produit scalaire

Soit \mathbf{A} une matrice symétrique définie positive, alors l'application $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x}^T \mathbf{A} \mathbf{y}$ est un produit scalaire.