

Bribes de géométrie

Résumé

Cette première série de séances pédagogiques se focalise sur les éléments de géométrie et, dans une certaine mesure, de topologie sous-jacents aux différents contextes qu'on aborde dans la suite.

Table des matières

1	Structure euclidienne usuelle sur \mathbb{R}^n	2
1.1	Produit scalaire de deux vecteurs	2
1.2	Dualité	3
2	Sous-espaces affines de \mathbb{R}^n	4
2.1	Définition	4
2.2	Se donner un sous-espace affine de \mathbb{R}^3	5
2.2.1	Paramétriquement	5
2.2.2	Implicitement	5
2.3	Cas extrémaux	6
3	Sous-espaces de \mathbb{R}^n	7
3.1	Se donner des parties de \mathbb{R}^n	7
3.2	Convexité dans \mathbb{R}^n	8
3.2.1	Parties convexes de \mathbb{R}^n	8
3.2.2	Fonctions convexes	9
4	Programmes linéaires dans le plan	10
A	Pivot de Gauss	13
A.1	Opérations élémentaires	13
A.2	Systèmes linéaires	14
A.3	Déterminant	14
A.4	Matrices équivalentes	14
B	Géométrie affine	14
B.1	Applications affines	14
B.1.1	Translations et homothéties	16
B.1.2	Projections affines	16
B.2	Transformations orthogonales	16
B.2.1	Aparté sur l'orthogonalité	16
B.2.2	Transformations vectorielles	17
B.2.3	Transformations affines	19
C	Les formes bilinéaires symétriques	19
C.1	Définition	19
C.2	Les formes quadratiques	20
C.3	Sous-ensembles des matrices symétriques	21

Convention 1. On fixe dans la suite un entier naturel n strictement positif. Le cas nul, même s'il garderait un sens dans les énoncés à suivre, nous est pas d'un grand intérêt. Afin de ne pas alourdir le texte, la i -ème coordonnée d'un vecteur $x \in \mathbb{R}^n$ sera systématiquement notée x_i . La transposée d'une matrice A sera notée A^T .

1 Structure euclidienne usuelle sur \mathbb{R}^n

Un espace euclidien est un espace vectoriel muni d'un produit scalaire, c'est-à-dire d'une forme bilinéaire symétrique définie positive. On vous renvoie à la section (C) pour un rafraîchissement de mémoire. L'espace vectoriel \mathbb{R}^n vient muni d'une structure euclidienne, dite usuelle, définie par le produit scalaire suivant : à tout couple de vecteurs (x, y) dans \mathbb{R}^n on associe le scalaire

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

Cette quantité est souvent exprimée de manière plus compacte par

$$\langle x, y \rangle = x^T y.$$

La longueur au sens intuitif du vecteur x , appelée norme euclidienne de x , s'exprime alors sous la forme

$$\|x\|_2 = \langle x, x \rangle^{1/2} = \sum_{i=1}^n x_i^2.$$

En ce sens la norme $\|\cdot\|_2$ est dite **norme associée au produit scalaire** $\langle \cdot, \cdot \rangle$. On rappelle que celle-ci permet d'écrire l'inégalité de CAUCHY-SCHWARZ sous la forme

$$|\langle x, y \rangle| \leq \|x\|_2 \|y\|_2$$

où x et y sont des vecteurs quelconques de \mathbb{R}^n .

1.1 Produit scalaire de deux vecteurs

Le signe du produit scalaire de deux vecteurs porte une signification géométrique dont la compréhension est essentielle à l'étude qualitative des problèmes d'optimisation, ainsi qu'aux méthodes itératives qui permettent d'en approcher une solution.

On se place en un premier temps dans le cas de dimension 2, du plan euclidien. Soient x et y deux vecteurs de \mathbb{R}^2 , on désigne par θ l'angle orienté (dans le sens direct) entre x et y et par ϕ (resp. ψ) celui entre x (resp. y) et la partie positive de l'axe des abscisses.

Question 1-1

1. Exprimer les coordonnées de x et y en fonction de leurs normes respectives et des angles ϕ et ψ .
2. En déduire une expression du produit scalaire de $\langle x, y \rangle$ en fonction de θ et des normes de x et y .

L'étude précédente s'étend relativement facilement au cas $n \geq 2$. Soit P un plan contenant les vecteurs x et y de \mathbb{R}^n , si ceux-ci sont indépendants P est bien entendu l'espace vectoriel engendré par x et y . La restriction du produit scalaire $\langle \cdot, \cdot \rangle$ à P prend la forme usuelle dans toute base orthonormée de P . Pour en trouver une dans P il suffit de normaliser un vecteur non nul f_1 de P puis de normaliser un vecteur f_2 de son orthogonal dans P . C'est la première étape de l'algorithme de GRAM-SCHMIDT. La base (f_1, f_2) nous ramène au cas de \mathbb{R}^2 précédent.

Le travail précédent permet d'établir en particulier le fait que l'orthogonalité de deux vecteurs, au sens intuitif que vous avez « *l'angle entre les deux vecteur est égal à $\pi/2[\pi]$* », correspond à l'annulation de leur produit scalaire. C'est cette définition, plus algébrique et n'impliquant pas de définir ce qu'est un angle, qui est utilisée en pratique.

Définition 1.1. Deux vecteurs x, y d'un espace euclidien $(E, \langle \cdot, \cdot \rangle)$ sont dits **orthogonaux** si leur produit scalaire est nul.

Cette notion d'orthogonalité de deux vecteurs va également permettre de décrire certains sous-espace d'un espace euclidien.

Définition 1.2. Soit $(E, \langle \cdot, \cdot \rangle)$ un espace euclidien. L'**orthogonal** d'un sous-espace vectoriel F de E est le sous-espace vectoriel défini par

$$F^\perp = \{x \in E \mid \forall y \in F, \langle x, y \rangle = 0\}.$$

Par abus on parlera d'orthogonal à un vecteur (ou une famille de vecteur) quand on sous-entend l'orthogonal au sous-espace vectoriel engendré. À noter que lorsque un vecteur est orthogonal à une famille de vecteurs, il est également orthogonal à toute combinaison linéaire d'éléments de celle-ci.

On revient à la signification géométrique du produit scalaire. On est désormais en mesure de voir qu'étant donné un vecteur non nul $\nu \in \mathbb{R}^2$, le lieu des vecteurs x sujets à la contrainte $\langle \nu, x \rangle = \nu^T x \geq 0$ est celui des vecteurs x tels que $\angle x\nu$ est aigu (ou orthogonal). C'est précisément le demi-espace de \mathbb{R}^2 délimité par $\{\nu\}^\perp$ et vers lequel pointe ν . La contrainte $\nu^T x \leq 0$ correspond de son côté au demi-espace délimité par $\{\nu\}^\perp$ et hors duquel pointe ν . C'est l'ensemble des vecteurs x qui décrivent un angle $\angle x\nu$ obtus.

Question 1-2 Décrire le lieu de \mathbb{R}^2 donné par la relation matricielle :

$$\begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Dans \mathbb{R}^n un demi-espace n'est plus délimité par une droite mais par un sous-espace vectoriel de dimension $n - 1$, c'est ce qu'on appelle un **hyperplan** de \mathbb{R}^n . Si ν est un vecteur non nul de \mathbb{R}^n , la condition de $\nu^T x \geq 0$ décrit les vecteurs x dans le demi-espace délimité par l'orthogonal de ν et vers lequel pointe ν . Inverser l'inégalité revient à étudier le demi-espace hors duquel pointe ν .

Question 1-3 Représenter le lieu de \mathbb{R}^3 décrit par la relations $x_1 + x_2 + x_3 \geq 0$.

1.2 Dualité

Si E est un espace vectoriel on note E^\vee l'espace vectoriel $\mathcal{L}(E, \mathbb{R})$ des formes linéaires sur E . C'est ce qu'on appelle l'espace **dual** de E , il est de même dimension que E . Quand E est euclidien, c'est-à-dire muni d'un produit scalaire $\langle \cdot, \cdot \rangle$ et de dimension finie, on a un isomorphisme d'espace vectoriel donné par l'application linéaire naturelle

$$\begin{aligned} \Psi : E &\longrightarrow E^\vee \\ x &\longmapsto [\Phi_x : y \longmapsto \langle x, y \rangle] \end{aligned}$$

Pour s'en convaincre il suffit d'étudier le noyau de Ψ : si celui-ci est trivial Ψ est injective et donc bijective car E et E^\vee ont même dimension, finie. Un vecteur x est dans le noyau de Ψ si

$$\forall y \in \mathbb{R}^2, \quad \langle x, y \rangle = 0.$$

En particulier, on doit avoir $\langle x, x \rangle = 0$ donc $x = 0$.

Le résultat précédent signifie qu'étudier, par exemple, l'image d'une sous-partie A de \mathbb{R}^n par toutes les formes linéaires de son dual revient à regarder l'image de A par les applications linéaires $\Phi_x = \langle x, \cdot \rangle$

pour x qui varie dans \mathbb{R}^n . Cette seconde approche est plus explicite, elle est donc plus évidente à implémenter (avec toutes les précautions d'usage quant aux questions de discrétisation). Elle est également loin d'être dénuée d'intérêt applicatif, on pourrait par exemple s'intéresser à la trace que laisse un objet 3D par projections sur des droites vectorielles de \mathbb{R}^3 .

2 Sous-espaces affines de \mathbb{R}^n

Afin de rendre compte des situations géométriques auxquelles on fera face, il va nous falloir élargir le panel des sous-ensembles de \mathbb{R}^n que nous sommes capables de manipuler. La première étape est d'être en mesure de travailler avec des translatés d'espaces vectoriels : les sous-espaces affines de \mathbb{R}^n . On se limitera à une approche utilitariste qui ne s'attarde pas sur la définition d'espace affine en général (on vous renvoie vers l'annexe pour plus de formalisme). De ce point de vue, de nombreux abus de terminologie seront fait.

2.1 Définition

Définition 2.1. Un sous-espace affine de \mathbb{R}^n est un translaté de sous-espace vectoriel de \mathbb{R}^n . Plus précisément, une partie A de \mathbb{R}^n est un sous-espace affine de \mathbb{R}^n s'il existe un point x dans \mathbb{R}^n et un sous-espace vectoriel F de \mathbb{R}^n tels que

$$A = x + F$$

ou la notation $x + F$ désigne l'ensemble $\{x + y \mid y \in F\}$.

Quand cette écriture a lieu, l'espace vectoriel F est unique, il est appelé la direction de A . Le point x ne l'est cependant pas, voir pour un exemple (2-4).

La dimension d'un sous-espace affine de \mathbb{R}^n est la dimension de sa direction. En particulier, un hyperplan affine est le translaté d'un hyperplan vectoriel de \mathbb{R}^n . Dans la pratique on omet les adjectifs, le fait d'être vectoriel ou non est clair du contexte.

Exemple 2.1. Tout sous-espace vectoriel de \mathbb{R}^n est un sous-espace affine de \mathbb{R}^n . En effet, on a toujours $F = 0 + F$.

Exemple 2.2. Tout singleton $\{x\}$ de \mathbb{R}^n définit un sous-espace affine de \mathbb{R}^n . Pour tout $x \in \mathbb{R}^n$ on peut toujours écrire $\{x\} = x + \{0\}$.

Exemple 2.3. La partie $A = \{(1, t) \in \mathbb{R}^2 \mid t \in \mathbb{R}\}$ est un sous-espace affine de \mathbb{R}^2 . On constate effectivement que

$$A = (1, 0) + \{(0, t) \mid t \in \mathbb{R}\}$$

le second terme étant évidemment un sous-espace vectoriel de \mathbb{R}^2 . On peut par exemple remarquer que c'est le noyau de la projection $\pi_1 : \mathbb{R}^2 \rightarrow \mathbb{R}$ sur la première coordonnée (ou encore l'espace vectoriel engendré par $(0, 1)$). Donc

$$A = (1, 0) + \text{Ker}(\pi_1)$$

Question 2-4

1. Donner une écriture différente de l'espace affine A dans l'exemple (2.3).
2. Décrire toutes les écritures possibles de l'espace affine A ; que pouvez vous en déduire en général?

Question 2-5 Décrire les sous-espaces affines de \mathbb{R} , \mathbb{R}^2 et \mathbb{R}^3 .

2.2 Se donner un sous-espace affine de \mathbb{R}^3

Un sous-espace vectoriel de \mathbb{R}^n est ou bien décrit par les équations que vérifient ses éléments ou alors explicitement à l'aide d'un certains nombres de paramètres. Dans le premier cas on dit que le sous-espace vectoriel est donné implicitement, dans le second qu'il est donné paramétriquement. Cela correspond à se donner un sous-espace vectoriel ou bien comme le noyau ou alors comme l'image d'une application linéaire. On peut encore dire qu'il est donné par contrainte ou par le biais d'une base par laquelle l'espace vectoriel est généré. La situation des sous-espaces affines est similaire.

2.2.1 Paramétriquement

Soit P le sous-espace vectoriel de \mathbb{R}^3 donné paramétriquement par

$$P = \{(t, 3t + u, -u) \mid (t, u) \in \mathbb{R}^2\}$$

C'est l'espace vectoriel engendré par la famille libre de \mathbb{R}^3 , $\{(1, 3, 0), (0, 1, -1)\}$. C'est encore l'image de \mathbb{R}^2 par l'application linéaire $(t, u) \mapsto (t, 3t + u, -u)$. Tout translaté de P est un sous-espace affine de \mathbb{R}^3 , par exemple

$$A = (1, 2, -1) + P = \{(1 + t, 3t + u + 2, -u - 1) \mid (t, u) \in \mathbb{R}^2\}$$

est un sous-espace affine de \mathbb{R}^3 donné paramétriquement. On voit que la forme des coordonnées de A est affine au sens que vous entendiez plus jeunes. Une application affine est une application linéaire une fois qu'on se débarrasse des constantes, plus formellement

Définition 2.2. Une application $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ est affine s'il existe $x_0 \in \mathbb{R}^p$ tel que $x \mapsto f(x) - x_0$ soit linéaire.

Dans notre cas $x_0 = (1, 2, -1)$. Afin d'éviter un détour qui nous écarterait de nos objectifs premiers nous garderons l'approche intuitive de la notion d'applications affines : une applications dont l'expression des coordonnées est affines dans les variables de départ. Pour plus détail on vous renvoie à (B.1).

Question 2-6 Écrire paramétriquement :

1. la droite de \mathbb{R}^2 de vecteur directeur $(1, -1)$ et passant par $(2, 3)$;
2. le plan de \mathbb{R}^3 donné par l'équation $x_1 + x_2 + x_3 = 2$.

2.2.2 Implicitement

Un sous-espace vectoriel F de \mathbb{R}^n est donné implicitement s'il est décrit comme le noyau d'une application linéaire. Soit M une matrice de taille (p, n) et de coefficients $(m_{i,j})$. Un vecteur $x \in \mathbb{R}^n$ est dans le noyau de l'application linéaire associée à M si

$$Mx = 0$$

relation compacte pour designer le système linéaire

$$\begin{cases} m_{1,1}x_1 + \cdots + m_{1,n}x_n = 0 \\ m_{2,1}x_1 + \cdots + m_{2,n}x_n = 0 \\ \vdots \\ m_{p,1}x_1 + \cdots + m_{p,n}x_n = 0. \end{cases}$$

Un sous-espace affine de \mathbb{R}^n est de manière similaire donné par un système linéaire avec un second membre non nul, qui prend la forme

$$Mx = R \iff \begin{cases} m_{1,1}x_1 + \cdots + m_{1,n}x_n = r_1 \\ m_{2,1}x_1 + \cdots + m_{2,n}x_n = r_2 \\ \vdots \\ m_{p,1}x_1 + \cdots + m_{p,n}x_n = r_p. \end{cases}$$

où R désigne un vecteur de \mathbb{R}^n . Pour s'en convaincre on se donne une solution particulière x_0 de $Mx = R$. L'équation précédente est donc équivalente à

$$Mx = Rx_0 \iff M(x - x_0) = 0.$$

L'ensemble des solutions de $Mx = R$ est donc ou bien vide ou alors $x_0 + \text{Ker}(M)$ pour une solution particulière x_0 . Une fois donné une base de $\text{Ker}(M)$ on obtient une écriture paramétrique de F .

Remarque 1. Dans la pratique, on n'a pas à trouver d'abord une solution particulière de notre système afin d'en trouver toutes les solutions. Le pivot de Gauss s'en charge, vous êtes invités à vous rafraîchir la mémoire avec l'annexe (A).

Question 2-7 Dessiner le lieu de \mathbb{R}^2 décrit par les contraintes

$$\begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \leq \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

- Décrire chacun des composants du lieu géométrique précédent paramétriquement.
- Que change le fait de rajouter la contrainte $x - 3y \leq 6$?
- Quel lieu correspond à la situation où l'on change le sens de toutes les inégalités?

Question 2-8 On désigne par A la partie de \mathbb{R}^2 donnée par ^a

$$A = \left\{ (x, y) \in \mathbb{R}^2 \mid \begin{array}{l} x + 2y \leq 3 \\ x - y \geq 2 \end{array} \right\}.$$

1. Représenter A graphiquement en indiquant les éléments qui permettent de décomposer votre représentation.
2. Quel est le lieu qu'on obtient si on inverse les inégalités.
3. Donner l'équation d'un demi-espace dont l'intersection avec A est un lieu non-vide borné de \mathbb{R}^2 .

^a. Question 1-1 Exam 2019.

2.3 Cas extrémaux

On a toujours le choix entre exprimer un sous-espace affine paramétriquement ou implicitement. Il est cependant plus simple d'exprimer une droite paramétriquement (on a besoin que d'un paramètre pour le faire) et un hyperplan implicitement (une seule équation y suffit).

- On pense une droite comme un point de départ auquel on ajoute les multiples scalaires d'un vecteur. On peut par exemple écrire

$$D = \{(0, 1) + \lambda(2, 3) \mid \lambda \in \mathbb{R}\}$$

Dans la pratique on verra souvent une écriture équivalente où l'on pense une droite comme l'unique droite passant par deux points donnés, dans le cas de D on peut écrire

$$D = \{(1 - t)(0, 1) + t(2, 4) \mid t \in \mathbb{R}\}$$

- Un hyperplan affine H est pour sa part donné par une équation de la forme $\nu^T x = r$. Si x_0 est une solution particulière de l'équation précédente celle-ci est équivalente à

$$\nu^T (x - x_0) = 0$$

Donc $H = x_0 + \{\nu\}^\perp$. C'est le translaté de l'orthogonal de ν par x_0 .

Question 2-9 Retrouver une écriture implicite de la droite D précédente. Comment s'y prendre pour transformer une écriture paramétrique en une écriture implicite ? ^a

a. Vous êtes invités à étudier plus en profondeur le passage représentation implicite \rightarrow représentation paramétriques et inversement

3 Sous-espaces de \mathbb{R}^n

On ne verra dans le cadre de ce cours qu'un contexte restreint de la géométrie des parties de \mathbb{R}^n . On se limite par exemple aux parties de \mathbb{R}^n qu'on se donne à l'aide de fonctions de \mathbb{R}^n vers \mathbb{R} . De plus, dans le cas des descriptions paramétriques on se limite aux graphes.

3.1 Se donner des parties de \mathbb{R}^n

Tout comme dans le cas des sous-espaces affines de \mathbb{R}^n , on peut se donner une partie de \mathbb{R}^n de manière *implicite* ou *paramétrique*.

Une partie de \mathbb{R}^n est décrite *implicitement* lorsqu'elle est donnée comme zéros de fonctions. Dans notre cadre, une telle partie A est de la forme

$$A = \{x \in \mathbb{R}^n \mid f(x) = 0\}$$

où f est une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Notez que cette écriture regroupe les parties décrites par les équations de la forme $f(x) = \star$ pour une constante $\star \in \mathbb{R}$. Une telle partie est zéros de la fonction $g = f - \star$.

Définition 3.1. Étant donné une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ on appelle *courbe de niveau r* de f la partie de \mathbb{R}^n définie par

$$C_r = \{x \in \mathbb{R}^n \mid f(x) = r\}.$$

Dans la suite on s'intéresse également au lieu de sous-niveau r d'une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$; c'est la partie de \mathbb{R}^n

$$C_{\leq r} = \{x \in \mathbb{R}^n \mid f(x) \leq r\}.$$

Question 3-10 Dessiner les courbes de niveaux 0, 1 et 2 des fonctions $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ et $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ données par les expressions $f(x, y) = x^2 + y^2$ et $g(x, y) = x^2 + 4y^2$.

Question 3-11 Comment décrire la surface délimitée par les deux branches d'une hyperbole ?

On dira qu'une partie de \mathbb{R}^{n+1} est décrite *paramétriquement* si elle prend la forme du graphe d'une fonction

$$\Gamma(f) = \{(t, f(t)) \mid t \in \mathbb{R}^n\}$$

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Une partie de \mathbb{R}^{n+1} décrite paramétriquement au sens précédent s'écrit facilement sous forme implicite : c'est les zéros de la fonction

$$\begin{aligned} \mathbb{R}^n \times \mathbb{R} &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto f(x) - y \end{aligned}$$

L'inverse n'est cependant pas vrai en général.

Question 3-12 Trouver un exemple d'une partie de \mathbb{R}^2 qu'on peut décrire implicitement mais pas paramétriquement. Que faudrait-il modifier pour qu'il soit possible de parler d'écriture paramétrique dans votre exemple ?

Remarque 2. Même avec une notion plus fine d'écriture paramétrique le passage implicite - paramétrique n'est pas toujours possible. Je vous invite à vous intéresser au théorème des fonctions implicites pour plus de détails.

Les graphes de fonctions apparaissent régulièrement dans la suite ; parmi les parties définies à partir de ceux-ci on retrouve l'épigraphe d'une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$

$$\text{epi}(f) = \{(x, t) \mid t \geq f(x)\}.$$

Question 3-13 Représenter l'épigraphe de la fonction \sin sur $[0, \frac{\pi}{2}]$ puis sur $[\frac{\pi}{2}, \pi]$.

Question 3-14 Représenter graphiquement l'intersection de l'épigraphe de la fonction $x \mapsto -\sqrt{x}$ sur \mathbb{R}_+ et de la partie

$$\{(x, y) \mid y \leq \sqrt{x}\}.$$

Question 3-15 On considère les fonctions suivantes notées f , g et h respectivement données par les expressions

$$f(x, y) = x + 4y, \quad g(x, y) = xy, \quad h(x, y) = \frac{x^2}{2} + \frac{y^2}{4}.$$

Représenter les courbes de niveaux $\mathcal{C}_1(f)$, $\mathcal{C}_2(g)$ et $\mathcal{C}_4(h)$.

3.2 Convexité dans \mathbb{R}^n

Les techniques d'optimisation se déroulent au mieux dans des contextes convexes. C'est une notion qu'il est important de garder en mémoire ; des comportements pathologiques voient le jour hors de ce cadre.

3.2.1 Parties convexes de \mathbb{R}^n

Définition 3.2. Une partie $A \subset \mathbb{R}^n$ est dite **convexe** si l'on peut relier 2 points quelconques de A par un segment intégralement inclus dans A . Plus formellement : A est convexe si

$$\forall x, y \in A, \forall t \in [0, 1] \quad (1 - t)x + ty \in A.$$

Exemple 3.1. Tout intervalle de \mathbb{R} est convexe.

Exemple 3.2. Les sous-espaces affines de \mathbb{R}^n sont convexes, il en va de même des demi-espaces de \mathbb{R}^n .

Dans la pratique reconnaître des parties convexes de \mathbb{R}^n revient à les construire à partir de parties convexes à l'aide d'opérations préservant la convexité.

- L'intersection de parties convexes est convexe ; le lieu admissible d'un programme linéaire est donc convexe.
- L'image d'une partie convexe par une fonction affine est convexe ; les changements de variables affines dans les contraintes d'un programme linéaire ne changent donc pas le caractère convexe de celui-ci.

Question 3-16 Donner deux arguments qui permettent de justifier le fait que la partie décrite question (3-14) est convexe.

Question 3-17 Est-ce que l'union de parties convexes est convexe ?

Question 3-18 Qu'est-ce que l'enveloppe convexe d'une partie de \mathbb{R}^n ?

Les parties convexes de \mathbb{R}^n jouissent de propriétés géométriques qui permettent de les isoler (séparer), lorsqu'elles ne s'intersectent pas¹. Ces problématiques sont en dehors de la portée de ce cours, elles nous donnent cependant une information géométrique utile pour généraliser l'étude des programmes linéaires au cas (convexe) non-linéaire.

Définition 3.3. Un point x d'une partie $A \subset \mathbb{R}^n$ est **point du bord** de A si pour tout disque D centré en x et de rayon non nul $D \cap A \neq \emptyset$.

L'ensemble des point du bord d'une partie A est appelée **bord** de A , elle est notée ∂A . L'union $\bar{A} = A \cup \partial A$ est appelée **adhérence** de A .

Question 3-19 Dessiner la définition précédente.

Définition 3.4. Soit A une partie de \mathbb{R}^n et $x \in \partial A$. Un **hyperplan d'appui** à A en x est un hyperplan H

- passant par x
- défini par un vecteur normal ν tel que $\forall y \in A, \langle y - x, \nu \rangle \leq 0$.

Un hyperplan d'appui à A cantonne donc la partie A à l'un des demi-espaces qu'il définit.

Question 3-20 Dessiner un exemple de partie

- n'ayant pas d'hyperplan d'appui en un point donné de son bord ;
- n'ayant aucun hyperplan d'appui ;
- ayant plus d'un hyperplan d'appui en un même point.

Proposition 3.1. Une partie convexe de \mathbb{R}^n admet un hyperplan d'appui en tout point de son bord².

Cette propriété est une conséquence du théorème de séparations des convexes fermés par des hyperplans.

Question 3-21 On reprend les notations de la question (3-15)

1. Donner un hyperplan d'appui à $\mathcal{C}_{\leq 4}(h)$.
2. Justifier le fait que $\mathcal{C}_{\leq 2}(g)$ n'a pas d'hyperplan d'appui.

3.2.2 Fonctions convexes

Définition 3.5. Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est **convexe** si

- son domaine de définition est une partie convexe de \mathbb{R}^n ;
- $\forall x, y$ dans le domaine de définition de f ,

$$\forall t \in [0, 1], f(ty + (1 - t)x) \leq tf(y) + (1 - t)f(x).$$

¹. Il faut un peu plus de rigueur, mais l'idée est là ...
². Celui-ci pouvant être vide.

Question 3-22 Dessiner la définition précédente et en donner une interprétation géométrique.

Définition 3.6. Une fonction f est dite **concave** si son opposé $-f$ est convexe.

Question 3-23 Donner des exemples de fonctions convexes.

Question 3-24 Prouver, en n'utilisant que la définition, que la fonction $x \mapsto x^2$ définie sur \mathbb{R} est convexe.

Question 3-25 Justifier le fait

1. que la somme pondérée, de poids positifs de fonctions convexes est convexe ;
2. qu'une fonction sous-linéaire est convexe ? en donner des exemples ;
3. que le max de fonctions convexes est convexe ;
4. que la composition d'une fonction $f \circ g$ d'une fonction f convexe avec une fonction affine g est convexe.

Question 3-26 Généraliser l'étude du max au contexte qui suit : soit $g : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$ une fonction qui pour tout $x \in A \subset \mathbb{R}^n$ a une fonction partielle $y \mapsto g(x, y)$ qui est convexe, montrer que $f(y) = \sup_{x \in A} g(x, y)$ est également convexe.

Question 3-27 Montrer que les sous-niveaux d'une fonction convexe sont convexes. La réciproque est-elle vraie ?

4 Programmes linéaires dans le plan

Un programme linéaire est un problème d'optimisation qui prend la forme

$$\begin{array}{ll} \min & f_0(x) \\ \text{sujet à} & \\ & f_i(x) \leq 0, 1 \leq i \leq p. \end{array}$$

où la fonction objectif f_0 ainsi que les f_i sont à valeurs réelles et de domaine \mathbb{R}^n . Les programmes linéaires forment la plus simple famille de problèmes d'optimisation contraints. Ils modélisent cela dit la majeure partie des problèmes de recherche opérationnelle classiques, ils permettent de résoudre les problèmes de plus courts chemins dans un graphes, de flots maximaux ou encore de problèmes de transports. Leur aspect unificateurs permet d'évaluer la difficulté de résoudre ces problèmes les uns par rapport aux autres.

Le travail effectué jusqu'à présent pose déjà le cadre pour résoudre autrement que par l'algorithme du simplexe ces types de problèmes. On se limite au cas de la dimension 2 afin de représenter les objets qu'on manipule.

Question 4-28 On revient sur les lieux géométriques décrits question (1-2). On note \mathcal{A}_u le lieu

décrit par les contraintes

$$\mathcal{A}_u : \begin{array}{rcl} -x + 2y & \leq & -1 \\ x + y & \leq & 1 \end{array}$$

et \mathcal{A}_b celui décrit par les contraintes précédentes auxquelles on ajoute

$$x - 3y \leq 6.$$

1. Étudier le programme linéaire de lieu admissible \mathcal{A}_u et minimisant y . Que se passe-t-il si l'on remplace y par $-y$?
2. A-t-on toujours une valeur minimal pour un programme linéaire ayant pour lieu admissible \mathcal{A}_b ?
3. Étudier le programme linéaire de lieu admissible \mathcal{A}_b et minimisant $x+y$. Effectuer cette même étude pour la fonction objectif $-x - y$.
4. Décrire une procédure géométrique qui permet de résoudre par coups les programmes linéaires en dimension 2.

Question 4-29 Donner un exemple d'un programme linéaire :

1. non borné ;
2. de lieu admissible non-borné mais de solution fini ;
3. ayant une infinité de solution ;
4. ayant une unique solution.

Question 4-30 On considère le programme linéaire $(P1)$ suivant :

$$\begin{array}{ll} \text{minimiser} & f_0(x_1, x_2) = 3x_1 + 2x_2 \\ \text{sujet à} & \end{array}$$

$$\begin{array}{l} x_1 - x_2 \leq 0 \\ 4x_1 - x_2 \geq 1 \\ -x_1 - x_2 \geq -5 \end{array}$$

1. Représenter le lieu admissible de $(P1)$ dans le plan euclidien.
2.
 - a. Tracer la courbe de niveau 6 de la fonction objectif de $(P1)$. Elle sera notée C_6 .
 - b. Indiquer les demi-espaces positif et négatif défini par C_6 . Dans quelle direction doit-on translater C_6 afin de minimiser f_0 .
3. Tracer la courbe de niveau qui réalise le minimum de $(P1)$ et calculer l'unique point optimal de $(P1)$. Quelle est la valeur optimale de $(P1)$?
4. Résumer votre démarche en rappelant la condition d'optimalité d'une solution de $(P1)$.

Question 4-31 On considère le programme linéaire (**P2**) donné par

$$\begin{array}{ll} \text{minimiser} & f_0(x_1, x_2) = x_1 + 2x_2 \\ \text{sujet à} & \\ & x_2 \leq 1 \\ & x_1 \leq 1 \\ & -x_2 \leq 1 \\ & -x_1 \leq 1 \end{array}$$

Résoudre (**P2**) en suivant la démarche précédente.

Question 4-32 On considère le programme linéaire (**P3**) suivant :

$$\begin{array}{ll} \text{minimiser} & x + 2y \\ \text{sujet à} & \\ & x + y \leq 1 \\ & -x + 2y \leq 2 \\ & x - 3y \leq 3 \end{array}$$

1. Représenter le lieu admissible de (**P3**) dans \mathbb{R}^2 .
2.
 - a. Tracer la courbe de niveau 0 de la fonction objectif de (**P3**). Elle sera notée \mathcal{C}_0 .
 - b. Indiquer les demi-espaces positif et négatif définis par \mathcal{C}_0 .
 - c. Indiquer dans quelle direction on doit translater \mathcal{C}_0 afin de minimiser la fonction objectif.
3. Tracer la courbe de niveau qui réalise le minimum de (**PL3**) et calculer l'unique point optimal de (**P3**). Quelle est la valeur optimale de (**P3**) ?

La démarche qui a été introduite précédemment permet de résoudre une famille un peu plus large de problèmes d'optimisation dans \mathbb{R}^2 .

Question 4-33 On considère la fonction différentiable $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ donnée par

$$f(x, y) = 3x^2 + y^2.$$

1. Représenter les courbes de niveaux 2 et 4 de f dans le plan euclidien.
2. À quel lieu correspond la condition $f(x, y) \leq 4$?
3. On s'intéresse au problème d'optimisation (**P4**)

$$\begin{array}{ll} \text{minimiser} & 2x + y \\ \text{sujet à} & \\ & 3x^2 + y^2 \leq 4 \end{array}$$

Représenter la courbe de niveau de la fonction objectif qui correspond à la valeur optimale de (**P4**).

4. Comment trouver le point optimale correspondant à (**P4**) ? Faire le calcul.

Question 4-34 On considère le problème d'optimisation (**P5**) suivant

$$\begin{array}{ll} \text{minimiser} & x + y \\ \text{sujet à} & \\ & x + 2y \leq 3 \\ & x \in B \end{array}$$

où B est la partie de \mathbb{R}^2 qu'on a décrit section (3-14).

1. Dessiner le lieu admissible du problème de (**P5**).
2. Représenter la courbe de niveau de la fonction objective qui réalise le minimum de (**P5**).
3. Calculer le point optimal ainsi que la valeur optimale de (**P5**).

La résolution de problèmes d'optimisation à la main comme on vient de le faire se heurte à deux obstacles :

- les dimensions des espaces concernés, qui représentent souvent des *features* d'un dataset, dépassent de loin les dimensions représentables graphiquement ;
- les fonctions concernées, que cela soit au niveau des contraintes ou des fonctions objectifs ne sont pas souvent linéaires.

Cela étant dit la démarche qu'on vient de mettre en évidence ici est le point de départ de l'essentiel des démarches itératives pour résoudre les problèmes d'optimisation. Pour être en mesure d'aborder cette généralisation il nous faudra, pour commencer, étudier les conditions sous lesquelles on est en mesure d'approcher localement et par des fonctions affines les courbes de niveau d'une fonction objectif. Cette étude recoupe le besoin de généraliser les notions de dérivabilité et dérivée d'une fonction numérique de domaine réel au cas multivarié.

A Pivot de Gauss

On se donne une matrice $M \in \mathcal{M}_{m,n}(\mathbb{R})$. Une opération élémentaire sur M est une des deux opérations suivantes :

- pour $i \in \{1, \dots, m\}$ et $j \in \{1, \dots, n\}$ interchanger les lignes L_i et L_j

$$L_i \leftrightarrow L_j$$

- pour $i \in \{1, \dots, m\}$, $j \in \{1, \dots, n\}$, $i \neq j$, $\lambda_i \in \mathbb{R}^*$ et $\lambda_j \in \mathbb{R}$, remplacer la ligne L_i par $\lambda_i L_i + \lambda_j L_j$:

$$L_i \leftarrow \lambda_i L_i + \lambda_j L_j.$$

Le pivot de Gauss est un algorithme de transformation d'une matrice M en matrice triangulaire supérieure et éventuellement en matrice identité, dans le but d'inverser une matrice, de calculer un déterminant ou de résoudre un système linéaire. Il est d'utilisation constante en calcul numérique.³

A.1 Opérations élémentaires

On traduit par la suite les opérations élémentaires précédentes à l'aide du produit matriciel.

Question 1-35 Trouver des matrices $P(i, j)$ et $U(i, j, \lambda, \nu)$ dont la multiplications avec M réalise les opérations élémentaires précédentes.

3. L'expression de l'inverse d'une matrice inversible à l'aide des cofacteurs est à proscrire ! Elle est théoriquement intéressante et particulièrement esthétique mais numériquement inefficace.

Question 1-36 Trouver les matrices $P_t(i, j)$ et $U_t(i, j, \lambda, \nu)$ qui réalisent les opérations précédentes sur les colonnes.

A.2 Systèmes linéaires

Question 1-37 Dédurre de la section A.1 que les opérations élémentaires du pivot de Gauss transforment tout système linéaire dont M est une matrice (si appliqué des deux côtés de l'égalité) en un système linéaire ayant les mêmes solutions.

Question 1-38 Comment procéder pour décrire toute les solutions d'un système linéaire ?

Question 1-39 Discuter de l'efficacité de l'implémentation de l'algorithme

A.3 Déterminant

Question 1-40 Retrouver les relations standards sur le déterminant d'une matrice qui subit une opération élémentaire à l'aide de la formule

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

Question 1-41 Comment procéder pour calculer le déterminant d'une matrice à l'aide du pivot de Gauss ?

A.4 Matrices équivalentes

Deux matrices A et B sont dites équivalentes s'il existe deux matrices inversibles P et Q telle que

$$A = PBQ.$$

Question 1-42 Justifier le fait que toute matrice est équivalente à une matrice ayant un bloc identité et des zéros partout ailleurs.

B Géométrie affine

Cette feuille est centrée autour des applications affines de \mathbb{R}^n , plus particulièrement autours des projections et transformations de l'espace affine euclidien. L'objectif est de vous apporter les éléments de langage et les résultats de structure qui vous permettent de les expliciter et les utiliser dans les contextes de ML. Pour rappel la majeure partie des algorithmes de ML sont de nature géométrique.

B.1 Applications affines

Cette section est un kit de survie en milieu hostile. Il s'agit de s'armer des différentes notions de transformations et applications affines dont vous aurez l'usage lors de vos cours de ML. Notre démarche dans la suite est particulièrement pragmatique⁴; on se limite au cas des applications affines de \mathbb{R}^n dans \mathbb{R}^m .

4. Elle pourrait heurter certaines âmes sensibles et quelques matheux qui ont gardés un peu de sens esthétique.

Définition B.1. Une application $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ est dite affine s'il existe un point $\mathbf{a} \in \mathbb{R}^m$ tel que $f - \mathbf{a}$ soit une application linéaire.

Cette définition formelle nécessite une réalisation plus concrète afin qu'on puisse l'utiliser. Pour cela on va introduire une généralisation de la notion de bases dans le cas de \mathbb{R}^n qui nous permettra de se placer en un point quelconque pour y travailler.

Définition B.2. Un repère de l'espace affine \mathbb{R}^n est la donnée d'un point \mathbf{o} appelé origine du repère et d'une base $\mathbf{v} = (v_1, \dots, v_n)$ de \mathbb{R}^n . On note (\mathbf{o}, \mathbf{v}) une telle donnée.

Le repère canonique de \mathbb{R}^n est le repère $(\mathbf{0}, e_1, \dots, e_n)$, que vous aviez eu l'habitude de manipuler dans le secondaire et en physique. Du moins dans les cas de dimensions 1 ou 2.

Définition B.3. Soient (\mathbf{o}, \mathbf{v}) et (\mathbf{l}, \mathbf{w}) deux repères respectivement de \mathbb{R}^n et \mathbb{R}^m . L'écriture de f dans les repères (\mathbf{o}, \mathbf{v}) et (\mathbf{l}, \mathbf{w}) est la donnée d'une matrice M telle que pour tout $x \in \mathbb{R}^n$,

$$f(x) = M(x - \mathbf{o})_{\mathbf{v}} + f(\mathbf{o}). \quad (1)$$

où $\bullet_{\mathbf{v}}$ désigne l'écriture du vecteur \bullet dans la base \mathbf{v} . Le membre de droite de l'équation 1 est sous-entendu être décrit dans le repère (\mathbf{l}, \mathbf{w}) . Cette équation est parfois appelée équation de f dans les repères (\mathbf{o}, \mathbf{v}) et (\mathbf{l}, \mathbf{w}) .

Exemple B.1. Une application affine peut être donnée par une expression dans le repère canonique

$$x \mapsto Mx + b$$

où x et b sont des vecteurs respectivement dans \mathbb{R}^n et \mathbb{R}^m et $M \in \mathcal{M}_{m,n}(\mathbb{R})$. Par exemple

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$

Ce qui nous donne une fonction de \mathbb{R}^2 dans \mathbb{R}^2 affine en x et y au sens qu'on entend intuitivement.

Dans le but de trouver des expressions agréables ou standardisées d'une application affine on est amené à écrire ces applications affines dans des repères adaptés. Pour pouvoir exprimer un tel contexte il nous faut être en mesure de formaliser la notion de changement de repères.

Définition B.4. Le changement de repères de \mathbb{R}^n de (\mathbf{o}, \mathbf{v}) vers (\mathbf{l}, \mathbf{w}) est l'application affine donnée dans les repères précédents par

$$x \mapsto P_{\mathbf{v}}^{\mathbf{w}}(x - \mathbf{o})_{\mathbf{v}} + \mathbf{l}$$

où $P_{\mathbf{v}}^{\mathbf{w}}$ est la matrice de passage de la base \mathbf{v} vers la base \mathbf{w} ; c'est-à-dire la matrice des vecteurs de \mathbf{v} écrits dans la base \mathbf{w} .

Voici enfin comment formaliser le passage de l'écriture d'une application dans un repère vers un autre.

Proposition B.1. Soient (\mathbf{o}, \mathbf{v}) et $(\mathbf{o}', \mathbf{v}')$ deux bases de \mathbb{R}^n , (\mathbf{l}, \mathbf{w}) et $(\mathbf{l}', \mathbf{w}')$ deux bases de \mathbb{R}^m . On note M la matrice d'une application affine f dans les repères (\mathbf{o}, \mathbf{v}) et (\mathbf{l}, \mathbf{w}) , et M' celle de f dans les repères $(\mathbf{o}', \mathbf{v}')$ et $(\mathbf{l}', \mathbf{w}')$. Si $P_{\mathbf{v}}^{\mathbf{v}'}$ désigne la matrice de passage de \mathbf{v} vers \mathbf{v}' et $P_{\mathbf{w}'}^{\mathbf{w}}$ la matrice correspondante dans le cas de \mathbf{w} et \mathbf{w}' alors pour tout $x \in \mathbb{R}^n$

$$M'(x - \mathbf{o}')_{\mathbf{v}'} + f(\mathbf{o}') = f(x) = P_{\mathbf{w}'}^{\mathbf{w}} M (P_{\mathbf{v}}^{\mathbf{v}'})^{-1} (x - \mathbf{o})_{\mathbf{v}} + f(\mathbf{o})$$

Question 2-43 Écrire la fonction du premier exemple dans le repère de \mathbb{R}^2 donné au départ et à l'arrivée par $((0, 0), \{(1, 1), (1, -1)\})$.

Remarque 3. Dans les faits, on ne fera pas souvent ce type de changement de repères. À partir des propriétés des applications à l'étude on sera à même de trouver un bon repère dans lesquels les décrire. Il reste que de tels changements de repères doivent être compris et faits quand nécessaire.

Exemple B.2. Les applications affines constantes sont celles dont l'expression dans tous les repères a une matrice qui est nulle.

Exemple B.3. Avec la définition qu'on a donné toute application linéaire va décrire une application affine.

B.1.1 Translations et homothéties

On s'attarde un instant sur les plus simples des applications affines non triviales.

Exemple B.4. La translation de vecteur b est l'application affine donnée dans le repère canonique de \mathbb{R}^n par $x \mapsto x + b$.

Exemple B.5. L'*homothétie* de rapport λ et centre \mathbf{o} est l'application affine de \mathbb{R}^n dans lui-même dont l'écriture dans le repère (\mathbf{o}, \mathbf{v}) prend la forme

$$x \mapsto \lambda(x - \mathbf{o})_{\mathbf{v}} + \mathbf{o}$$

L'homothétie de rapport 1 est simplement l'identité.

Question 2-44 Donner une particularité que partagent les homothéties et les translations.

Question 2-45 Que donne la composition de deux translations ? Celle de deux homothéties ?

Question 2-46 Dessiner l'image du carré de sommets $(\pm 1, \pm 1)$ dans \mathbb{R}^2 par les homothéties suivantes :

- de rapport 2 et de centre l'origine
- de rapport 2 et de centre $(1, 0)$.

B.1.2 Projections affines

Les projections affines sont la généralisation des projections linéaires. Pour rappel une projection $p : E \rightarrow E$ de l'espace vectoriel E sur lui-même est une application linéaire qui satisfait la relation $p^2 = p$. Cette relation garantit le fait que $G = \text{Ker}(p)$ et $F = \text{Ker}(p - \text{id})$ sont des espaces vectoriels supplémentaires dans E . Ainsi, pour tout $x = x_F + x_G$, la projection p est définie par $p(x) = x_F$. L'application p est appelée projection sur F parallèlement à G .

Définition B.5. Une application affine f de \mathbb{R}^n dans \mathbb{R}^n est une projection affine si sa matrice dans un repère est la matrice d'une projection linéaire.

Une projection affine, dans un repère (\mathbf{o}, \mathbf{v}) de \mathbb{R}^n , s'écrit donc sous la forme

$$x \mapsto M(x - \mathbf{o})_{\mathbf{v}} + \mathbf{o}$$

On voit dans cette écriture que \mathbf{o} est fixé par la projection. On vient d'écrire la projection de $\mathbf{o} + \text{Ker}(M - I_n)$ parallèlement à $\mathbf{o} + \text{Ker}(M)$.

Question 2-47 Étudier des exemples de projections affines dans \mathbb{R}^2 . Dessiner systématiquement les droites affines qui caractérisent cette projection.

B.2 Transformations orthogonales

B.2.1 Aparté sur l'orthogonalité

On se donne un produit scalaire $\langle \cdot, \cdot \rangle$ sur \mathbb{R}^n ⁵. On rappelle que c'est une forme bilinéaire symétrique définie positive. Elle vient avec une norme associée dite norme euclidienne définie par

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

5. Pensez au produit scalaire usuel si ça vous facilite la vie.

Définition B.6. Deux vecteurs x, y in \mathbb{R}^n sont dit orthogonaux pour $\langle \cdot, \cdot \rangle$ si $\langle x, y \rangle = 0$.

Nous avons vu que cette notion correspond dans le cas du produit scalaire usuel au fait que l'angle entre x et y est $\pi/2$ [π].

Définition B.7. Une base \mathbf{v} est dite *orthogonale* si les vecteurs qui la constituent sont orthogonaux deux à deux. Elle est *orthonormale* si ses vecteurs sont de norme 1.

L'exemple le plus simple d'une base orthonormée est celui de la base canonique de \mathbb{R}^n pour le produit scalaire usuel.

Question 2-48 Donner d'autres bases orthonormales de \mathbb{R}^3 .

Question 2-49 Comment construire une base orthonormée de \mathbb{R}^3 à partir de la base

$$\{(1, 0, 1), (0, 1, 1), (1, 1, 0)\}?$$

Définition B.8. Étant donné une partie K de \mathbb{R}^n on appelle orthogonal de K pour $\langle \cdot, \cdot \rangle$ le sous-espace vectoriel

$$K^\perp = \{x \in \mathbb{R}^n \mid \forall y \in K, \langle x, y \rangle = 0\}.$$

Ainsi un hyperplan H de \mathbb{R}^n quand il est donné implicitement est décrit comme l'orthogonal d'un vecteur de \mathbb{R}^n qu'on appelle un vecteur *normal*.

Question 2-50 Décrire dans \mathbb{R}^3 l'orthogonal à la partie $\{(1, 1, 1), (1, 0, 1)\}$.

B.2.2 Transformations vectorielles

La notion de produit scalaire permet à la fois de formaliser la notion d'angle ainsi que celle de distance⁶. Les transformations affines de \mathbb{R}^n , c'est-à-dire les applications affines bijectives, qui préservent les angles et les distances sont d'un grand intérêt géométrique; elles ne touchent pas aux dispositions relatives de configurations de parties dans \mathbb{R}^n et donc préservent les propriétés géométriques de celles-ci. Elles préservent par exemple la perpendicularité ou le parallélisme. Pour l'instant on se limite au cas vectoriel avant d'exposer la situation affine, plus générale.

Avant de poursuivre notre travail on prend un instant pour étudier de plus prêt une écriture matricielle de $\langle \cdot, \cdot \rangle$. Étant donné une base \mathbf{v} de \mathbb{R}^n , on appelle matrice de $\langle \cdot, \cdot \rangle$ dans la base \mathbf{v} la matrice

$$\mathbf{A} = (\langle v_i, v_j \rangle)_{i,j}.$$

Question 2-51 Montrer que, dans la base \mathbf{v} , on a pour tout $x, y \in \mathbb{R}^n$, $\langle x, y \rangle = x^T \mathbf{A} y$.

Si \mathbf{v} est une base orthonormale dans ce cas $\mathbf{A} = \mathbf{I}$ et on a, pour tout $x, y \in \mathbb{R}^n$ écrits dans cette base

$$\langle x, y \rangle = x^T y.$$

Ce qui nous ramène à l'écriture usuelle du produit scalaire.

Hypothèse B.2. Quitte à se ramener à une base orthonormée appropriée⁷ on peut supposer que notre produit scalaire a la forme usuelle.

⁶. On étudiera plus en détail la seconde question un peu plus tard.

⁷. Ça existe toujours. Voir algorithme de Gram-Schmidt.

Soit M une matrice inversible dans une base orthonormée de \mathbb{R}^n . Dire que M préserve le produit scalaire $\langle \cdot, \cdot \rangle$ signifie que pour tout $x, y \in \mathbb{R}^n$ on a

$$(Mx)^T(My) = x^T y.$$

D'où l'on obtient, pour tout $x, y \in \mathbb{R}^n$

$$x^T (M^T M - I_n) y = 0.$$

En prenant x et y ayant des coefficients nuls sauf un on peut montrer que les coefficients de $M^T M - I_n$ sont tous nuls. Donc

$$M^T M = I_n.$$

Proposition B.3. Une isomorphisme de \mathbb{R}^n préserve un produit scalaire si et seulement sa matrice dans une base orthonormée satisfait

$$M^T M = I_n.$$

Une matrice qui satisfait la propriété précédente est dite **orthogonale**.

Question 2-52 Quel peut être le déterminant d'une matrice orthogonale ?

Question 2-53 Traduire la proposition précédente en fonction des vecteurs colonnes de la matrice M .

On étudie un premier exemple de matrices orthogonales. On se fixe une base orthonormée de \mathbb{R}^n dans laquelle on travaille. Soit $\theta \in \mathbb{R}$ on note $R(\theta)$ la matrice

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Question 2-54 Montrer que $R(\theta)$ est une matrice orthogonale pour tout θ . Quel est son déterminant ?

Question 2-55 Dessiner l'image du carré de points extrémaux $(\pm 1, \pm 1)$ par $R(\theta)$ pour $\theta = 0, \pi/4$.

Un exemple de matrice orthogonale qui n'est pas de déterminant 1 est celui-ci

$$s = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Question 2-56 Interpréter géométriquement cette matrice.

Question 2-57 Dessiner l'image du carré de points extrémaux $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$ par $R(\theta)$ pour $\theta = 0, \pi/4$.

Question 2-58 Déterminer toutes les matrices orthogonales de \mathbb{R}^2 muni de son produit scalaire usuel. Vous pourrez les étudier suivant leurs spectres.

Question 2-59 Donner des exemples de matrices orthogonales en toute dimension. Essayez d'être le plus général possible.

Question 2-60 Étendre la classification des matrices orthogonales au cas de dimension 3.

B.2.3 Transformations affines

TODO

C Les formes bilinéaires symétriques

Les formes bilinéaires symétriques apparaissent sous différents aspects en mathématiques parmi ceux-ci, en réalité dès que l'on travaille avec une matrice symétrique. On peut par exemple évoquer le produit scalaire sur \mathbb{R}^n qui nous permet de décrire un objet qui donne à la fois une information sur la norme d'un vecteur et l'angle que deux vecteurs forment. On peut encore évoquer le cas de la matrice de covariance, d'usage fréquent et symétrique. Ou enfin les polynôme de degré 2 ayant un nombre quelconque de variables.

C.1 Définition

Définition C.1. Soit E un \mathbb{R} -espace vectoriel. Une **forme bilinéaire** est une application $\phi : E \times E \rightarrow \mathbb{R}$ pour laquelle les applications partielles en tout point $x \in E$

- $t \mapsto \phi(x, t)$
- $t \mapsto \phi(t, x)$

sont linéaires.

Exemple C.1. La multiplication matricielle $\times : \mathcal{M}_n(\mathbb{R}) \times \mathcal{M}_n(\mathbb{R}) \rightarrow \mathbb{R}$ est une application bilinéaire sur l'espace vectoriel des matrices carrées de taille n .

Définition C.2. Une forme bilinéaire ϕ sur un \mathbb{R} -espace vectoriel E est dite

- **symétrique** si $\forall x, y \in E, \phi(x, y) = \phi(y, x)$;
- **antysymétrique** si $\forall x, y \in E, \phi(x, y) = -\phi(y, x)$;
- **positive** si $\forall x, \phi(x, x) \geq 0$;
- **négative** si $\forall x, \phi(x, x) \leq 0$;
- **définie** si $\forall x, \phi(x, x) = 0 \Rightarrow x = 0$.

Question 3-61 Justifier le fait que le produit scalaire usuel sur \mathbb{R}^n est une forme bilinéaire symétrique définie positive.

Dans la pratique, et quand on travaille sur \mathbb{R}^n , une forme bilinéaire symétrique est donnée par une matrice symétrique P de taille (n, n) . L'application bilinéaire Φ associée s'écrit

$$\forall x, y \in \mathbb{R}^n, \quad \Phi(x, y) = x^T P y.$$

L'écriture à droite de l'égalité donne effectivement une forme bilinéaire symétrique. Si l'on s'est placés dans la base b_1, \dots, b_n pour écrire les vecteurs x et y , la matrice P est la matrice de coefficients $P_{ij} = b_i^T P b_j$.

C.2 Les formes quadratiques

Les formes bilinéaires symétriques donnent naissance aux formes quadratiques. Celles-ci donnent des programmes d'optimisations qu'on résout assez bien par les méthodes introduites dans ce cours. On passe d'un contexte linéaire dans le cas des programmes linéaires à celui de degré 2.

Définition C.3. Soit E un espace vectoriel sur \mathbb{R} . On dit que $Q : E \rightarrow \mathbb{R}$ est une forme quadratique sur E s'il existe une forme bilinéaire symétrique $\phi : E \times E \rightarrow \mathbb{R}$ telle que pour tout $x \in E$, $Q(x) = \phi(x, x)$.

Cette définition permet déjà de voir qu'étant donné un produit scalaire $\langle \cdot, \cdot \rangle$ sur E , l'application $x \mapsto \langle x, x \rangle$ est une forme quadratique ; c'est la norme au carré (la norme associée au produit scalaire). Une forme quadratique Q sur \mathbb{R}^n est donc la donnée d'une matrice symétrique P telle que

$$\forall x \in \mathbb{R}^n, \quad Q(x) = x^T P x.$$

Exemple C.2. La matrice carrée $P = \begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix}$ donne lieu à la forme quadratique Q l'expression pour $(x, y) \in \mathbb{R}^2$ est

$$Q((x, y)) = x^2 + 4xy.$$

On remarque que l'expression obtenue est un polynôme de degré total⁸ 2. Ce fait est général : étant donné une forme quadratique non nulle Q sur \mathbb{R}^n , l'expression $Q(x)$ est un polynôme de degré total 2 en les coordonnées du vecteur $x \in \mathbb{R}^n$.

Remarque 4. Par abus, dans le cas de \mathbb{R}^n , ces adjectifs sont accolés à la matrice symétrique P telle que $Q(x) = x^T P x$.

Question 3-62 Donner, quand possible, un exemple de forme quadratique sur \mathbb{R}^2 qui soit (n'oubliez pas de justifier vos réponses) :

- positive, non définie ;
- définie positive ;
- définie non positive ni négative.

Dessiner un croquis illustrant le graphe de la fonction quadratique donnée dans chacun des cas^a.

^a. Vous êtes autorisés (encouragés ?) à utiliser `matplotlib`.

On admet par la suite le résultat important suivant :

Théorème C.1. *Toute matrice symétrique $P \in M_n(\mathbb{R})$ est diagonalisable sur \mathbb{R} dans une base ortho-normée.*

Ce résultat, important, est valable sous des conditions beaucoup plus larges sur le corps de base ; on le montre en partie à l'aide d'un algorithme dit de Gauss.

Question 3-63 À l'aide du déterminant d'une matrice carrée, donner une condition sur une forme quadratique Q sur \mathbb{R}^n afin que celle-ci soit :

- définie ;
- non positive et non négative.

Pouvez-vous en dire plus dans le cas des formes quadratiques sur \mathbb{R}^2 ? Sait-on, dans ce cas, quand une forme quadratique est positive ? Comment différencier le cas positif du cas négatif ? Formaliser le cas des formes quadratiques sur \mathbb{R}^2 .

8. C'est le degré du polynôme obtenu en remplaçant toutes les variables par une unique variable t .

Question 3-64 Suffit-il qu'une matrice symétrique ait des coefficients positifs pour que la forme quadratique associée le soit ?

Question 3-65 Donner des conditions suffisantes sur la positivité de la matrice associée à une forme quadratique pour que celle-ci soit

- convexe ou concave ;
- ni l'un ni l'autre.

Ces conditions sont-elles nécessaires ?

C.3 Sous-ensembles des matrices symétriques

On note S_n l'ensemble des matrices symétriques de $M_n(\mathbb{R})$. L'ensemble des matrices symétriques positives (resp. définies positives) est noté S_n^+ (resp. S_n^{++}). Les notations dans le cas négatif sont S_n^- et S_n^{--} .

Question 3-66 On s'intéresse à la structure de S_n .

1. Montrer que S_n est un sous-espace vectoriel de $M_n(\mathbb{R})$. Peut-on en dire plus ? Calculer la dimension de S_n .
2. Montrer que S_n^+ est une partie convexe de S_n . Peut-on en dire autant de S_n^{++} ?
3. Est-ce que le résultat précédent est valable dans le cas de S_n^- ?