Chapitre 1 - Concepts et notations de la théorie des ensembles

Le cours va commencer de façon bien abstraite, par une énumération peut-être un peu indigeste de non-définitions (il faut bien des mots non définis pour entamer les premières définitions...), de notations, de définitions. Qu'on se le dise, tout est essentiel pour la suite !

1 - Ensembles

Non-définition 1-1-1 : Le mot **ensemble** ne sera pas défini. Intuitivement, un ensemble est un paquet de choses (qui sont elles-mêmes des ensembles, mais glissons là dessus), non rangées, sans répétition possible.

Cette explication intuitive est particulièrement déficiente : la théorie des ensembles s'est définitivement constituée au début du (XXème) siècle lorsqu'on a pris conscience que certains paquets ne pouvaient décemment être appelés "ensembles". Mais il me faut savoir me taire pour pouvoir avancer.

Non-définition 1-1-2 : Le verbe **appartenir** ne sera pas défini. Intuitivement, on dit que a **appartient à** un ensemble A lorsqu'il fait partie des choses dont l'ensemble A est un paquet.

Définition 1-1-1 : Pour tous a et A, on dit que a est **élément** de A lorsque a appartient à A.

Notation 1-1-1 : On note $a \in A$ pour "a appartient à A", et $a \notin A$ pour "a n'appartient pas à A".

Non-définition 1-1-3 : L'expression **est égal à** ne sera pas définie. Intuitivement... vous savez bien ce que ça veut dire !

Notation 1-1-2 : On note $a = b$ pour "a est égal à b".

Définition 1-1-2 : On dit que deux objets a et b sont **distincts** ou **différents** lorsqu'ils ne sont pas égaux.

Notation 1-1-3 : On note $a \neq b$ pour "a est distinct de b".

Non-définition 1-1-4 : L'**ensemble vide** ne sera pas défini. Intuitivement, c'est un ensemble qui n'a aucun élément, par exemple l'ensemble des solutions réelles de l'équation $x^2 = -1$.

Notation 1-1-4 : $\emptyset$ désigne l'ensemble vide.

Au-delà de ces non-définitions, j'utiliserai un certain nombre de propriétés intuitives de ces diverses notions sans me risquer à les énoncer. Par exemple si je sais que trois réels x, y et z vérifient $x = y$ et $y = z$, j'en déduirai que $x = z$ sans m'expliquer davantage. Et d'autres manipulations, parfois un peu plus subtiles mais qui ne devraient pas poser de problème.

Notation 1-1-5 : Pour un certain nombre d'objets $a_1, a_2, \ldots, a_n$, on notera $\{a_1, a_2, \ldots, a_n\}$ l'ensemble dont les éléments sont exactement $a_1, a_2, \ldots, a_n$.

Ça a l'air simple, mais il y a déjà des pièges possibles parmi ces notions non définies, il faut donc se concentrer un peu.

Question : les notations $\{1, 3\}$ et $\{3, 1\}$ désignent-elles le même ensemble d'entiers ? Réponse : oui, bien sûr, le premier ensemble possède 1 et 3 pour éléments, le second possède 3 et 1. L'intuition qu'on peut avoir du mot "et" nous fait affirmer comme évident que ce sont les mêmes.

Question : la notation $\{2, 2, 2\}$ est-elle licite, et si oui que désigne-t-elle exactement ? Réponse : ben, oui, on ne voit pas ce qui l'interdirait ; c'est l'ensemble dont les éléments sont 2, 2 et 2. Vu ce qu'on comprend du mot "et" c'est une façon compliquée de parler de l'ensemble $\{2\}$, ensemble à un seul élément : l'entier 2.

La remarque paraît stupide, mais il arrive effectivement qu'on note des ensembles de cette façon apparemment tordue : par exemple, l'énoncé suivant est vrai :

Pour tous réels a (non nul), b et c tels que $b^2 - 4ac \geq 0$, l'ensemble des solutions réelles de l'équation (d'inconnue x) :

$$ax^2 + bx + c = 0$$

est l'ensemble :

$$\{\frac{-b + \sqrt{b^2 - 4ac}}{2a}, \frac{-b - \sqrt{b^2 - 4ac}}{2a}\}.$$

Or lorsqu'on écrit une vérité si notoire, dans le cas particulier où $b^2 - 4ac = 0$ on a répété deux fois le même élément !

Question : Combien d'éléments possède l'ensemble $\{\{\{3, 6\}\}\}$? Réponse : un seul bien sûr ! C'est par définition l'ensemble possédant l'unique élément $\{\{3, 6\}\}$.

Question : Les notations $\emptyset$, $\{\emptyset\}$, $\{\{\emptyset\}\}$ désignent-elles le même ensemble ? Réponse : non, certainement pas ! Le premier de ces trois ensembles —l'ensemble vide— n'a aucun élément, le second et le troisième en ont un seul et sont donc distincts de l'ensemble vide. Ils sont aussi distincts l'un de l'autre, parce que l'unique élément de $\{\emptyset\}$ est vide, alors que l'unique élément de $\{\{\emptyset\}\}$ ne l'est pas.

Reprenons le cours de nos notations.

Notation 1-1-6 : On note $\{x \mid p(x)\}$ l'ensemble formé des ensembles x qui vérifient la propriété $p(x)$.

Par exemple, $\{x \mid x \in \mathbf{R}$ et $ax^2 + bx + c = 0\}$ est l'ensemble des solutions réelles d'une équation du second degré.

Notation 1-1-7 : Pour un ensemble A, on note $\{x \in A \mid p(x)\}$ l'ensemble $\{x \mid x \in A$ et $p(x)\}$.

Par exemple, on notera plutôt $\{x \in \mathbf{R} \mid ax^2 + bx + c = 0\}$ l'ensemble de l'exemple précédent.

Définition 1-1-3 : On dit qu'un ensemble A est **inclus** dans un ensemble B (ou que A est une **partie** de B,ou que B **contient** A) lorsque la propriété suivante est réalisée : pour tout x, si x appartient à A, alors x appartient à B. (Pour le redire en termes moins formalistes : lorsque tous les éléments de A sont éléments de B).

Notation 1-1-8 : On note $A \subset B$ pour "A est inclus dans B".

Remarques : il est facile de se convaincre que pour tout ensemble A, l'inclusion $A \subset A$ est vraie ; il peut paraître un peu plus bizarre que l'inclusion $\emptyset \subset A$ le soit aussi, mais c'est bien vrai.

Notation 1-1-9 : On note parfois $A \subsetneq B$ pour "A est inclus dans B, mais distinct de B".

Définition 1-1-4 : On appelle **réunion** de deux ensembles A et B l'ensemble des éléments qui appartiennent à A **ou** appartiennent à B.

Notation 1-1-10 : On note $A \cup B$ cette réunion.

Définition 1-1-5 : On appelle **intersection** de deux ensembles A et B l'ensemble des éléments qui appartiennent à A **et** appartiennent à B.

Notation 1-1-11 : On note $A \cap B$ cette intersection.

Définition 1-1-6 : On appelle **différence** de deux ensembles A et B l'ensemble des éléments de A qui ne sont pas éléments de B.

Notation 1-1-12 : On note $A \setminus B$ cette différence.

Définition 1-1-7 : Quand B est inclus dans A, la différence $A \setminus B$ est appelée le **complémentaire** de B dans A.

Notation 1-1-13 : Le complémentaire est noté avec un symbole que je ne sais pas obtenir de mon traitement de textes.

Je n'énumèrerai pas ici les multiples relations très simples à vérifier entre réunions, intersections, etc... (un exemple : pour tous ensembles A, B et C, $(A \cup B) \cup C = A \cup (B \cup C)$).

2 - Ensemble des parties d'un ensemble

Si cette notion a droit a la faveur d'un numéro de section particulier —alors qu'elle se range parfaitement dans la suite de la litanie qui précède— c'est parce que je sais qu'elle est moins bien maîtrisée et qu'il s'agit simplement d'attirer votre attention sur la nécessité de la connaître, et, idéalement, de la comprendre.

Définition 1-2-8 : On appelle **ensemble des parties** d'un ensemble A l'ensemble dont les éléments sont les parties de A.

Notation 1-2-14 : L'ensemble des parties de A est noté $\mathcal{P}(A)$.

Exemple : Pour a et b deux objets, l'ensemble des parties de $\{a, b\}$ est $\{\emptyset, \{a\}, \{b\}, \{a, b\}\}$. Il possède donc "à première vue" quatre éléments. À seconde vue, il en possède quatre si a et b sont distincts, et deux si a et b sont égaux.

En détraquant subitement l'ordre logique du cours, et en faisant intervenir des entiers avant d'en avoir parlé, illustrons la notion d'"ensemble des parties" en comptant ses éléments ; plusieurs preuves en sont possibles, j'ai choisi d'écrire la preuve par récurrence, peu palpitante, parce qu'elle donne l'occasion d'écrire méthodiquement une preuve justement sans surprise.

Proposition 1-2-1 : Pour tout ensemble fini A, si n désigne le nombre d'éléments de A, le nombre d'éléments de $\mathcal{P}(A)$ est 2^n.

Démonstration : On va procéder à une démonstration par récurrence sur l'entier n.

- Cas particulier où n vaut 0.

Dans ce cas, l'ensemble A est vide. Son ensemble des parties est alors $\{\emptyset\}$, qui possède bien $1 = 2^0$ élément.

• Soit n un entier fixé ($n \geq 0$). Supposons la proposition vraie pour tous les ensembles à n éléments et prouvons la pour un ensemble A fixé possédant $n+1$ éléments.

Puisque $n+1$ vaut au moins 1, A n'est pas vide. Soit a un élément de A. Notons B l'ensemble $A \setminus \{a\}$ (en clair, l'ensemble formé des autres éléments de A). Ainsi B est un ensemble qui possède n éléments.

Les parties de A se subdivisent en deux catégories : celles dont a est un élément, et les autres. Commençons par examiner les autres, pour nous apercevoir que ce sont exactement les parties de B. Il y en a donc 2^n, par application de l'hypothèse de récurrence.

Comptons maintenant les parties de A dont a est un élément. Étant donnée une telle partie E, l'ensemble $E \setminus \{a\}$ est alors une partie de B ; et réciproquement chaque fois qu'on part d'une partie F de B, l'ensemble $F \cup \{a\}$ est une partie de A dont a est un élément. Il y a donc autant de parties de A dont a est un élément que de parties de B, donc encore 2^n.

Le nombre total de parties de A est donc $2^n + 2^n$, soit 2^{n+1}.

•

3 - Couples, produit cartésien

Non-définition 1-3-5 : Le mot **couple** ne sera pas défini. Intuitivement un couple est formé de deux objets, distincts ou égaux, et dans un ordre bien précis.
Notation 1-3-15 : Le couple formé des objets a et b est noté (a, b).
Définition 1-3-9 : Le **produit cartésien** de deux ensembles A et B est l'ensemble des couples (a, b) où a est un élément de A et b un élément de B.
Notation 1-3-16 : On note $A \times B$ le produit cartésien de A et B.
Exemple : Pour ceux qui n'auraient pas compris,

$$\{a, b\} \times \{c, d\} = \{(a, c), (a, d), (b, c), (b, d)\}.$$

Comme pour l'ensemble des parties, il est facile de compter combien d'éléments possède le produit cartésien de deux ensembles finis.
Proposition 1-3-2 : Pour tous ensembles finis A et B, si m désigne le nombre d'éléments de A et n le nombre d'éléments de B, le nombre d'éléments de $A \times B$ est mn.
Démonstration : Je ne la ferai pas ; on peut par exemple faire une récurrence sur n.

•

4 - Relations

Non-définition 1-4-6 : Le mot **relation** sur un ensemble E ne sera pas défini. Intuitivement, une relation est la description de liens entre certains éléments de l'ensemble.
Exemple : La relation "est inférieur ou égal" sur l'ensemble $\mathbf{R}$ des réels : pour deux réels x et y on peut avoir $x \leq y$ ou non.
Définition 1-4-10 : Le **graphe** d'une relation $\mathcal{R}$ sur un ensemble E est l'ensemble des couples (a, b) de $E \times E$ tels que $a \,\mathcal{R}\, b$.

Tiens, arrêtons nous un instant pour relire ensemble cette définition et voir comment elle peut être mal retenue par un étudiant peu scrupuleux. Il est facile (si on relit de temps en temps son cours tout de même !) de retenir que le graphe de $\mathcal{R}$ a un rapport avec $a \,\mathcal{R}\, b$. Mais combien en verra-t-on qui glisseront sur des mots anodins en apparence (et d'ailleurs anodins en réalité... si on ne les oublie pas !) Je tiens ici à souligner que le graphe est un **ensemble**. Point n'est besoin d'apprendre par cœur sans comprendre ; les divers objets qui sont définis dans ce cours se casent en effet dans un petit nombre de catégories : souvent des ensembles, assez souvent des applications, souvent des n-uplets (des applications particulières), souvent aussi des nombres (entiers, réels...), plus rarement des relations, etc... Il n'est pas difficile de sentir dans quelle boîte ranger les graphes : ce ne sont manifestement pas des triplets, ni des nombres complexes ! Le plus important est de ne pas oublier de les ranger quelque part. Savoir à quelle catégorie appartient un objet permet d'éviter les bourdes les plus monumentales : le symbole $\cap$ aura un sens entre deux ensembles, pas entre deux réels —et réciproquement pour le symbole $+$.
Exemple : Le graphe de la relation $\leq$ sur $\mathbf{R}$ est un demi-plan de $\mathbf{R}^2$.

Alignons maintenant quatre définitions rébarbatives, mais incontournables.

Définition 1-4-11 : Une relation $\mathcal{R}$ sur un ensemble E est dite **réflexive** lorsque pour tout élément a de E, $a \mathcal{R} a$.

Définition 1-4-12 : Une relation $\mathcal{R}$ sur un ensemble E est dite **symétrique** lorsque pour tous éléments a, b de E, si $a \mathcal{R} b$, alors $b \mathcal{R} a$.

Définition 1-4-13 : Une relation $\mathcal{R}$ sur un ensemble E est dite **transitive** lorsque pour tous éléments a, b, c de E, si $a \mathcal{R} b$ et $b \mathcal{R} c$, alors $a \mathcal{R} c$.

Et la plus difficile à bien mémoriser des quatre :

Définition 1-4-14 : Une relation $\mathcal{R}$ sur un ensemble E est dite **antisymétrique** lorsque pour tous éléments a, b de E, si $a \mathcal{R} b$ et $b \mathcal{R} a$, alors $a = b$.

Quelques commentaires sur cette dernière : c'est, comme son nom l'indique, en gros le contraire de la propriété de symétrie. La symétrie c'est exigé que quand deux éléments sont liés dans un sens, ils le sont aussi dans l'autre. L'antisymétrie, ce serait approximativement demander que si deux éléments sont liés dans un sens, ils ne le sont pas dans l'autre. Mais cette condition empêcherait un élément d'être lié à lui-même, ce qui ne serait pas désespérant mais ne serait pas conforme à l'usage. De ce fait, l'usage s'est fait de compliquer la définition afin de garder la permission pour un élément d'être lié à lui-même...

On comprendra peut-être un peu mieux la définition en écrivant la contraposée de l'implication qu'elle contient :

Autre formulation de la définition de l'antisymétrie : Une relation $\mathcal{R}$ sur un ensemble E est antisymétrique lorsque pour tous éléments a, b distincts de E, on ne peut avoir simultanément $a \mathcal{R} b$ et $b \mathcal{R} a$.

Comme nous sommes encore débutants, je fais encore l'effort d'expliciter une autre façon de présenter la même notion :

Autre formulation de la définition de l'antisymétrie : Une relation $\mathcal{R}$ sur un ensemble E est antisymétrique lorsque pour tous éléments a, b distincts de E, $a \not\mathcal{R} b$ ou $b \not\mathcal{R} a$.

Bien évidemment, ce genre de liste de formulations équivalentes n'est surtout pas à "savoir par cœur". Ce qui est par contre indispensable, c'est de se familariser avec les petites manipulations qui permettent de passer de l'une à l'autre, selon les besoins.

Question pour voir si on a bien tout compris : le graphe de la relation $\leq$ sur $\mathbf{R}$ est-il antisymétrique ? Réponse : c'était une question piège (grossier et crétin) ! Le mot "antisymétrique" s'applique à des relations, et on nous a bien dit de faire attention que le graphe est, lui, un ensemble. La réponse est non pour une raison tout à fait stupide. Si on répond "oui", on fait une erreur de distraction pas bien grave ; mais des à-peu-près analogues peuvent avoir des conséquences dramatiques s'ils ouvrent un raisonnement. Restons donc précis.

5 - Relations d'ordre

En pratique, les relations qui pourront nous intéresser cette année ne seront jamais bien compliquées ; le vocabulaire que nous avons dû ingurgiter à la section précédente n'a d'utilité que pour savoir reconnaître deux types très particuliers de relations : les relations d'ordre, puis, à la section prochaine, les relations d'équivalence.

Définition 1-5-15 : Une relation est dite **relation d'ordre** lorsqu'elle est simultanément réflexive, transitive et antisymétrique.

Intuitivement, une relation d'ordre est une relation qui peut raisonnablement être appelée "est plus grand que" (ou bien sûr "est plus petit que").

Exemples : La relation "$\leq$" sur $\mathbf{R}$ est une relation d'ordre. Pour A ensemble fixé, la relation "$\subset$" sur $\mathcal{P}(A)$ est une relation d'ordre (plus compliquée à maîtriser, dans la mesure où deux parties de A ne sont pas forcément comparables l'une à l'autre).

6 - Relations d'équivalence et partitions

Le morceau est plus sérieux que pour les relations d'ordre, car on ne va pas se contenter de donner une définition, mais on va aussi voir le lien avec un autre concept. Pour expliquer intuitivement ce qui va suivre, une relation d'équivalence est une relation qui peut raisonnablement s'appeler "est du même groupe que" et une partition est une répartition en groupes.

Définition 1-6-16 : Une relation est dite **relation d'équivalence** lorsqu'elle est simultanément réflexive, symétrique et transitive.
Exemples : L'égalité sur n'importe quel ensemble E fixé. La relation “a même parité” sur l'ensemble $\mathbf{N}$ des entiers naturels. La relation “est parallèle à” sur l'ensemble des droites d'un plan (affine).

Avalons encore trois définitions de plus en plus indigestes (mais ce n'est pas gratuit, les concepts serviront plus loin, notamment en arithmétique...)
Définition 1-6-17 : Soit $\sim$ une relation d'équivalence sur un ensemble E, et a un élément de E. On appelle **classe d'équivalence** de a l'ensemble

$$\{x \in E \mid a \sim x\}.$$

Avec des mots, la classe d'équivalence de a est l'ensemble formé des éléments de la même catégorie que a.
Notation 1-6-17 : On notera $\dot{a}$ la classe d'équivalence de a.

Sans commentaires —il y en aura plus loin— un objet plus étrange :
Définition 1-6-18 : Soit $\sim$ une relation d'équivalence sur un ensemble E. On appelle **ensemble-quotient** de E par la relation $\sim$ l'ensemble

$$\{\dot{a} \mid a \in E\}.$$

Attention tout de même ! Comme $\dot{a}$ est une partie (et non un élément) de E, l'ensemble-quotient est un ensemble de parties de E. Ce n'est pas une partie de E mais une partie de $\mathcal{P}(E)$. Ce n'est pas si compliqué, mais il ne faut pas s'y perdre.
Notation 1-6-18 : L'ensemble-quotient de E par $\sim$ est noté $E/\sim$.
Définition 1-6-19 : Une **partition** d'un ensemble E est un ensemble $\mathcal{Q}$ de parties de E vérifiant les trois propriétés suivantes :
(i) L'ensemble vide n'est pas un élément de $\mathcal{Q}$.
(ii) Deux éléments distincts de $\mathcal{Q}$ sont disjoints.
(iii) Tout élément de E appartient à un élément de $\mathcal{Q}$.

C'est dur à avaler parce qu'on rentre inévitablement dans le monde des ensembles dont les éléments sont eux-mêmes des ensembles. Les éléments de $\mathcal{Q}$, qui sont des parties de E, doivent être intuités comme des groupes d'éléments de E vérifiant une condition commune. Exemple de partition : en notant $I \subset \mathbf{N}$ l'ensemble des entiers impairs et $P \subset \mathbf{N}$ l'ensemble des entiers pairs, $\{I, P\}$ est une partition de $\mathbf{N}$. Tentons maintenant de commenter les conditions... La condition (i) est sans intérêt, juste là pour que les énoncés marchent bien. La condition (ii) nous assure qu'on n'a inscrit aucun élément de E dans deux groupes à la fois. La condition (iii) signifie qu'on n'a oublié d'inscrire personne : tout élément de E est dans un groupe.

On remarquera qu'on peut regrouper les deux conditions significatives et donner une
Autre formulation de la définition d'une partition : Une **partition** d'un ensemble E est un ensemble $\mathcal{Q}$ de parties de E vérifiant les deux propriétés suivantes :
(i) L'ensemble vide n'est pas un élément de $\mathcal{Q}$.
(ii) Tout élément de E appartient à un et un seul élément de $\mathcal{Q}$.

Bien évidemment là encore il n'est pas question d'apprendre par cœur ce genre de reformulation. Il faut se convaincre —et ici ce n'est peut-être pas facile— qu'elle est bien équivalente à la précédente.

Et maintenant la synthèse finale, qui expliquera ce qui est un ensemble-quotient à ceux qui ont compris ce qu'est une partition, et expliquera ce qu'est une partition à ceux qui ont compris ce qu'est un ensemble-quotient.
Proposition 1-6-3 : Soit $\sim$ une relation d'équivalence sur un ensemble E. L'ensemble-quotient $E/\sim$ est une partition de E.
Complément : Toute partition de A peut s'obtenir ainsi comme quotient par une (unique) relation d'équivalence de E.
Démonstration : (la preuve du complément étant “laissée au lecteur”)

Vérifions successivement les trois propriétés définissant une partition.
Vérification de (i) : Soit A un élément de $E/\sim$. Par définition de $E/\sim$, on peut prendre un a dans E tel que $A = \dot{a}$. Comme $\sim$ est réflexive, $a \sim a$, donc $a \in \dot{a} = A$. Ainsi A n'est pas réduit à l'ensemble vide.

Vérification de (ii) : Soit A et B deux éléments de $E/\sim$. On peut trouver des éléments a et b de E tels que $A = \dot{a}$ et $B = \dot{b}$. On doit montrer que si A et B sont distincts, ils sont alors disjoints, et on va y procéder par contraposition, c'est-à-dire en montrant que si A et B ne sont pas disjoints, ils sont égaux.
Supposons donc A et B non disjoints.

L'objectif est de prouver que $A = B$, on va montrer successivement les inclusions $A \subset B$ et $B \subset A$.
Par l'hypothèse qu'on vient de faire, on peut prendre un c qui appartienne simultanément à A et à B.
Montrons tout d'abord que $A \subset B$.

Pour ce faire, prenons un x quelconque dans A et prouvons que $x \in B$.
Comme $x \in A = \dot{a}$, par définition d'une classe d'équivalence, on obtient $a \sim x$. Comme $c \in A = \dot{a}$, on obtient de même $a \sim c$ —puis, grâce à la symétrie de $\sim$, on obtient $c \sim a$. Comme $c \in B = \dot{b}$, on obtient enfin $b \sim c$. En mettant bout à bout les trois informations ainsi obtenues ($b \sim c$, $c \sim a$ et $a \sim x$) et en jouant deux fois sur la transitivité de $\sim$, on obtient alors que $b \sim x$, c'est-à-dire que $x \in B$.

Ceci prouve bien que $A \subset B$.
Passons à l'inclusion dans l'autre sens.

L'astuce est ici classique : elle consiste à remarquer que nos hypothèses (à savoir que A et B sont des classes d'équivalence, et qu'elles ne sont pas disjointes) sont totalement symétriques en A et B. Dès lors, en échangeant A et B dans le morceau précédent de la preuve, on obtient bien l'inclusion

$B \subset A$.
La double inclusion étant désormais prouvée, on a ainsi prouvé que $A = B$.

On a ainsi prouvé que si $A \cap B \neq \emptyset$, alors $A = B$. La propriété (ii) est prouvée. Ouf, c'était le plus gros morceau !

Vérification de (iii) : Soit a un élément de E. Comme $\sim$ est réflexive, $a \in \dot{a}$, et de ce fait on a bien trouvé un élément de $A/\sim$ dont a est lui-même élément. C'est fini !

•

7 - Applications

Non-définition 1-7-7 : Le mot **application** ne sera pas défini. Intuitivement, une application f est un moyen de faire correspondre à chaque élément x d'un ensemble E (son **ensemble de départ**) un élément noté $f(x)$ (et appelé l'**image** de x) d'un ensemble F (son **ensemble d'arrivée**).
Remarque : Bien que le concept ne soit pas défini, il est courant de demander dans un exercice ou un problème de "montrer que" telle ou telle formule "définit bien une application de E vers F". Ce qui est conventionnellement attendu lorsqu'on pose une telle question, c'est qu'il soit prouvé que la formule associe bien sans ambiguïté une et une seule image $f(x)$, qui se trouve bien dans l'ensemble F, à chaque élément x de E. Évidemment, on ne le demande que quand ça pose une difficulté, plus ou moins cachée, l'essentiel du travail étant alors de repérer où elle se cache !
Définition 1-7-20 : On appelle graphe d'une application f d'un ensemble E vers un ensemble F l'ensemble

$$\{(x, f(x)) \mid x \in E\}.$$

C'est une partie de l'ensemble-produit $E \times F$.
Notation 1-7-19 : L'ensemble de toutes les applications d'un ensemble E vers un ensemble F est noté F^E. (On ne s'en servira guère).

Il n'est pas inutile de savoir que lorsque E et F sont finis avec respectivement m et n éléments, F^E possède alors n^m éléments. Pas question d'en donner une "démonstration" formelle, puisqu'il nous manque la définition d'"application", mais on peut l'expliquer assez bien : on a n choix dans F pour l'image d'un premier élément x_1 de E, puis encore n choix pour l'image d'un deuxième élément x_2, ces choix se faisant indépendamment : on a donc $n \times n = n^2$ choix pour les images de ces deux éléments. Puis on a n choix pour l'image de x_3, donc n^3 choix pour l'image des trois premiers éléments, et ainsi de suite.

Cette propriété explique le choix de la notation F^E.

Les deux définitions qui vont suivre sont, d'expérience, trop mal connues des étudiants. Elles sont accompagnées d'une notation qui peut être source de confusion, et doivent donc être maîtrisées sous risque de morsures graves.

Définition 1-7-21 : Soit f une application d'un ensemble E vers un ensemble F. Pour toute partie A de E, on appelle **image directe** de A l'ensemble

$$\{f(x) \mid x \in A\}.$$

Notation 1-7-20 : L'image directe de A est notée $f(A)$ (certains ouvrages, prudents, y apportent des variantes comme $f{<}A{>}$). Évidemment cette notation est dangereuse, car elle ressemble trop à la notation $f(x)$ (image d'un élément de E) et l'étudiant veillera bien à ne pas mélanger ces deux concepts !

Exemples : Pour f l'application de $\mathbf{R}$ vers $\mathbf{R}$ définie par $f(x) = x^2$, on a $f(\{1,2\}) = \{1,4\}$, $f(\mathbf{R}^+) = \mathbf{R}^+$, $f(\mathbf{R}) = \mathbf{R}^+$, $f(\emptyset) = \emptyset$...

Définition 1-7-22 : Soit f une application d'un ensemble E vers un ensemble F. Pour toute partie B de F, on appelle **image réciproque** de B l'ensemble

$$\{x \in E \mid f(x) \in B\}.$$

Notation 1-7-21 : L'image réciproque de B est notée $f^{-1}(B)$. Là encore, cette notation est dangereuse, car un autre sens de f^{-1} va apparaître plus bas. **Merci de ne pas confondre**.

Passons à une problématique qui modélise en jargon la vieille problématique de la résolution d'équations ; ainsi la recherche d'antécédents d'un élément m de l'ensemble d'arrivée n'est autre que la résolution de l'équation $f(x) = m$.

Définition 1-7-23 : Soit f une application d'un ensemble E vers un ensemble F. Pour y élément de F et x élément de E, on dit que x est un **antécédent** de y lorsque $f(x) = y$.

Définition 1-7-24 : Soit f une application d'un ensemble E vers un ensemble F. On dit que f est une **injection** lorsque tout élément de F possède **au plus un** antécédent par f.

Définition 1-7-25 : Soit f une application d'un ensemble E vers un ensemble F. On dit que f est une **surjection** lorsque tout élément de F possède **au moins un** antécédent par f.

Définition 1-7-26 : Soit f une application d'un ensemble E vers un ensemble F. On dit que f est une **bijection** lorsque tout élément de F possède **exactement un** antécédent par f.

J'espère qu'il saute aux yeux de tout le monde sans que j'aie à l'énoncer qu'une application est une bijection si et seulement si c'est simultanément une injection et une surjection.

Il n'est en revanche peut-être pas inutile de mettre en relief une :

Autre formulation de la définition d'une injection Soit f une application d'un ensemble E vers un ensemble F. L'application f est une injection si et seulement si

$$\text{pour tous } x_1, x_2 \text{ distincts dans } E,\ f(x_1) \neq f(x_2).$$

Par simple contraposition, on obtient une variante pas désagréable, car elle invite à manipuler des égalités :

Autre formulation de la définition d'une injection Soit f une application d'un ensemble E vers un ensemble F. L'application f est une injection si et seulement si

$$\text{pour tous } x_1, x_2 \text{ dans } E,\ (f(x_1) = f(x_2)) \Rightarrow (x_1 = x_2)\,.$$

Les applications peuvent être appliquées l'une après l'autre, ce qui conduit à poser la

Définition 1-7-27 : Soit E, F et G trois ensembles, soit f une application de E vers F et g une application de F vers G. On appelle **application composée** de g et f l'application h de E vers G définie par : $h(x) = g[f(x)]$ pour tout élément x de E.

Notation 1-7-22 : La composée de g et f est notée $g \circ f$.

Enfin un concept idiot, mais fort souvent utilisé : l'application qui ne déplace rien.

Définition 1-7-28 : Soit E un ensemble, l'application f de E vers E définie par $f(x) = x$ pour tout élément x de E est appelée l'**application identique** de E.

Notation 1-7-23 : L'application identique de E est notée Id_E.

8 - Réciproque d'une bijection

Danger ! La notion étudiée dans cette section est notée f^{-1} et peut donc être confondue avec la notion d'image réciproque d'un ensemble définie à la section précédente. Vous aurez été prévenus.

L'idée est des plus simples : une bijection met en correspondance deux ensembles point par point, donc peut se retourner (il suffit de changer les flèches de sens) ; le nouvel objet est alors une nouvelle bijection.

À idée simple, théorème pas très compliqué mais qu'il ne faut pas méconnaître.

Théorème 1-8-1 : Soit f une application d'un ensemble E vers un ensemble F. f est une bijection si et seulement s'il existe une application g de F vers E telle que

$$g \circ f = Id_E \qquad \text{et} \qquad f \circ g = Id_F.$$

Démonstration : Le théorème est énoncé comme un "si et seulement si", on va le prouver très méthodiquement et classiquement en prouvant un sens, puis l'autre.
Prouvons d'abord l'implication "$\Rightarrow$".
Supposons que f est une bijection.

On me demande de montrer l'existence de g, la façon la plus simple de procéder est encore de le trouver. Ici, ce n'est pas difficile si on comprend le sens des concepts manipulés ; tout élément y de F possède un et un seul antécédent par f : choisissons donc d'appeler $g(y)$ cet unique antécédent. Il faut vérifier que le g ainsi construit vérifie les deux identités réclamées.

Vérifions d'abord que $g \circ f = Id_E$.

Soit x un élément de E. L'antécédent de $f(x)$ par f est évidemment x, ce qui s'écrit : $g[f(x)] = x$ en revenant à la définition de g.

Cette identité étant vraie pour tout x, on a bien prouvé l'égalité entre applications : $g \circ f = Id_E$.

Prouvons l'autre identité, en vérifiant que $f \circ g = Id_F$.

Soit y un élément de F. Puisque $g(y)$ est par définition un antécédent de y, ceci signifie que $f[g(y)] = y$.

Cette identité étant vraie pour tout y, on a bien prouvé l'égalité entre applications : $f \circ g = Id_F$.

On a donc bien réussi à construire le g que l'on souhaitait.

Et ceci prouve l'implication "$\Rightarrow$".
Passons maintenant à la preuve de l'implication "$\Leftarrow$".
Supposons que g existe, vérifiant les identités $g \circ f = Id_E$ et $f \circ g = Id_F$, et prouvons que f est une bijection.

Montrons d'abord que f est une surjection

Soit y un élément de F ; comme $f[g(y)] = y$, on sait trouver au moins un antécédent de y par f, à savoir $g(y)$.

Ceci étant vrai pour tout y, f est donc une surjection.

Prouvons maintenant que f est une injection.

Soit x_1 et x_2 deux éléments de E tels que $f(x_1) = f(x_2)$, on va montrer que $x_1 = x_2$. En appliquant g aux deux termes de l'identité $f(x_1) = f(x_2)$, on obtient $g[f(x_1)] = g[f(x_2)]$, et comme $g \circ f = Id_E$, ceci se réduit à $x_1 = x_2$.

On a donc prouvé que f est injective.

Puisque f est injective et surjective, elle est bijective.

On a donc prouvé l'implication "$\Leftarrow$".

•

Un exemple d'utilisation de ce théorème sera donné un peu plus loin (preuve de l'existence d'une bijection entre **N** et **Z**).

Complément : Lorsque f est une bijection, l'application g donnée dans l'énoncé du théorème ci-dessus est unique.

Démonstration : Soit deux applications g et h vérifiant à elles deux les quatre identités :

$$g \circ f = Id_E \quad , \; f \circ g = Id_F \qquad h \circ f = Id_E \qquad \text{et} \qquad f \circ h = Id_F.$$

On a alors $g = g \circ (f \circ h) = (g \circ f) \circ h = h$.

•

Cette existence et cette unicité permettent d'énoncer la :

Définition 1-8-29 : Pour f bijection d'un ensemble E vers un ensemble F, on appelle **bijection réciproque** de f l'unique application g telle que $g \circ f = Id_E$ et $f \circ g = Id_F$.
Notation 1-8-24 : La bijection réciproque de f est notée f^{-1}.

Au risque d'être un peu lourd, j'insiste encore une fois sur le point suivant : la bijection réciproque de f n'existe que lorsque f est elle-même une bijection!

9 - Restrictions

Définition 1-9-30 : Soit f une application d'un ensemble E vers un ensemble F. Pour E_1 partie de E, on appelle **restriction** de f à E_1 l'application g de E_1 vers F définie par : $g(x) = f(x)$ pour tout x de E_1.
Notation 1-9-25 : La restriction de f à E_1 est notée $f_{|E_1}$.

C'est une notion simple ; techniquement on utilisera un concept un peu plus lourd à énoncer (mais en pratique sans même s'en rendre compte !)

Variante : Soit f une application d'un ensemble E vers un ensemble F, soit E_1 une partie de E et F_1 une partie de F telles que $f(E_1) \subset F_1$. On définit sans nom ni notation bien clairement fixée une notion de "restriction" de f de E_1 vers F_1 encore définie par $f(x) = g(x)$ pour x dans E_1.

Reste à comprendre l'intérêt de définitions aussi creuses en apparence. Une des utilités de cette technique est de permettre de retaper avec assez peu de travaux une application qui n'est pas injective ou pas surjective —ou ni l'un ni l'autre— et d'en faire une nouvelle application ayant de bien meilleures propriétés.

Exemples : Soit f l'application de $\mathbf{R}^*$ vers $\mathbf{R}$ définie par $f(x) = \ln|x|$ pour tout x réel non nul. Cette application est surjective, mais pas injective. Si nous considérons plutôt la restriction $f_{|\mathbf{R}^{+*}}$, il s'agit alors d'une bijection.

De même soit g l'application de $\mathbf{R}$ vers $\mathbf{R}$ définie par $g(x) = e^x$ pour tout x réel. Cette application, elle, est injective mais pas surjective. Si nous faisons appel à la deuxième notion de "restriction" pour la restreindre de $\mathbf{R}$ vers $\mathbf{R}^{+*}$, nous tombons encore sur une bijection (réciproque de la précédente).

Toujours plus fort, soit h l'application de $\mathbf{R}$ vers $\mathbf{R}$ définie par $h(x) = x^2$ pour tout x réel. Cette application n'est pas injective (les réels strictement positifs ont deux antécédents), ni surjective (les réels strictement négatifs n'en ont aucun). Mais on remarque que $h(\mathbf{R}) = \mathbf{R}^+$ et donc *a fortiori* $h(\mathbf{R}^+) \subset \mathbf{R}^+$. Il est donc possible de restreindre h en une application de $\mathbf{R}^+$ vers $\mathbf{R}^+$. Et on obtient alors une bijection (il faut encore le prouver rigoureusement, on le fera peut-être au second semestre !). La fonction racine carrée peut alors être définie comme l'inverse de cette bijection.

De même la fonction sin n'a rien d'une bijection vue comme fonction de $\mathbf{R}$ vers $\mathbf{R}$, mais en devient une si on la restreint en une application de $[-\frac{\pi}{2}, \frac{\pi}{2}]$ vers $[-1, 1]$. Ceci permet encore de définir sa réciproque, une nouvelle fonction nommée arc sinus.

Chapitre 2 - Juste quelques mots sur les entiers naturels

Je supposerai —avec raison— que les lecteurs de ces notes savent compter, et je ne tenterai donc pas de donner des définitions de choses bien connues -je renonce même à énumérer les “non-définitions” implicites tout le long, comme celles de “nombre entier naturel” (il n’est peut-être tout de même pas inutile de rappeler ici que les entiers “naturels” sont nos bons entiers du comptage, c’est-à-dire les positifs, y compris 0), d’addition, et quelques autres.

De même, je ne dirai rien de ce qu’est un “ensemble fini” ou de ce qu’est son “nombre d’éléments”.

Ces mises au point fort négatives étant faites, recommençons à apprendre des choses.

1 - Récurrences

Vous savez sans doute tous faire une récurrence correctement (enfin espérons-le), en revanche tout le monde ne connaît peut-être pas la méthode parfois appelée “de récurrence forte”. Une petite mise au point ne sera donc de ce fait peut-être pas inutile.

Il serait possible de donner des énoncés précis et ensemblistes (qu’on se garderait de démontrer) décrivant ce qu’est une récurrence, il sera sans doute plus clair de donner des “principes” d’aspect un peu inhabituel —des énoncés qui parlent d’énoncés— et surtout un exemple pour celui qui est nouveau.

Principe de récurrence (“faible”) : Soit (H_n) un énoncé dépendant d’un paramètre entier n. Si les deux énoncés suivants :

(1) (H_0)

(2) Pour tout $n \geq 0$, $\big((H_n) \Rightarrow (H_{n+1})\big)$

sont vrais, alors la conclusion :

Pour tout $n \geq 0$, (H_n)

est également vraie.

Et voilà maintenant l’énoncé plus technique encore en apparence expliquant ce qu’est une “récurrence forte”. Il est recommandé de ne le regarder qu’en diagonale, d’examiner de près l’exemple, et, éventuellement, de s’y pencher de nouveau après avoir acquis soi-même un peu de pratique.

Principe de récurrence forte : Soit (H_n) un énoncé dépendant d’un paramètre entier n. Si les deux énoncés suivants :

(1) (H_0)

(2) Pour tout $n \geq 0$, $\Big(\big(\text{pour tout } k \leq n,\ (H_k)\big) \Rightarrow (H_{n+1})\Big)$

sont vrais, alors la conclusion :

Pour tout $n \geq 0$, (H_n)

est également vraie.

Comme promis, un exemple d’utilisation de ce principe amélioré, que j’espère plus lisible que l’énoncé du principe par lui-même.

Proposition 2-1-4 : Tout entier naturel supérieur ou égal à 2 peut s’écrire comme un produit de nombres premiers.

Démonstration : Démontrons par récurrence “forte” sur l’entier $n \geq 2$ la propriété suivante :

(H_n) : n peut s’écrire comme produit de nombres premiers.

- Vérifions tout d’abord (H_2)

  L’écriture $2 = 2$ nous montre que 2 est produit de nombres premiers (d’un seul, en l’occurence !)

- Soit n un entier avec $n \geq 2$, supposons que l’hypothèse (H_k) est vraie pour tout entier $k \geq 2$ inférieur ou égal à n, et montrons (H_{n+1}).

  On distinguera deux cas :

  $*$ Le cas où $n+1$ est premier ; on le traite de la même façon qu’on a traité le cas de 2 : l’écriture $n+1 = n+1$ répond à la question.

  $*$ Le cas où $n+1$ n’est pas premier. Dans ce cas, on peut écrire $n+1 = kl$, où k et l sont deux entiers différents de $n+1$; comme $k = n+1/l$, k est inférieur (au sens large) à $n+1$; puisque $k \neq n+1$ on a même $k < n+1$ et donc $k \leq n$; puisque $l \neq n+1$, on a aussi $k = n+1/l \neq 1$ et donc $2 \leq k \leq n$. L'entier k est donc bien dans le domaine de valeurs qui garantit la validité de (H_k) et on peut donc écrire $k = p_1 p_2 \cdots p_a$ pour un entier $a \geq 1$ et des nombres premiers $p_1, p_2, \ldots, p_a$. En échangeant les rôles de k et l on peut de la même façon écrire $l = q_1 q_2 \cdots q_b$ pour un entier $b \geq 1$ et des nombres premiers $q_1, q_2, \ldots, q_b$. Il n'y a plus qu'à juxtaposer l'information accumulée : en écrivant $n+1 = kl = p_1 p_2 \cdots p_a q_1 q_2 \cdots q_b$, on parvient à écrire $n+1$ comme produit de nombres premiers ce qui démontre bien (H_{n+1}).

On a donc bien prouvé (H_{n+1}) dans les deux cas.

On a donc bien prouvé la deuxième condition requise pour la méthode de "récurrence forte". La propriété (H_n) est donc vraie pour tout entier $n \geq 2$.

•

2 - Deux faits qu'on sait déjà, mais qu'on peut toutefois apprendre

Les "faits" en question concernent la relation d'ordre sur $\mathbf{N}$. Pour pouvoir les énoncer, deux définitions préalables :

Définition 2-2-31 : Soit A une partie de $\mathbf{N}$ et N un élément de A. On dit que N est le **plus grand élément** de A lorsque pour tout k de A, $k \leq N$.

Définition 2-2-32 : Soit A une partie de $\mathbf{N}$ et n un élément de A. On dit que n est le **plus petit élément** de A lorsque pour tout k de A, $n \leq k$.

Ce sont vraiment des notions dont la définition ne fait que répéter le nom, en plus formalisé.

Les deux "faits" suivants sont de bon sens, mais (surtout pour le second) les avoir en tête permet de trouver la bonne idée pour débuter une démonstration.

Fait : toute partie finie non vide de $\mathbf{N}$ admet un plus grand élément.

Fait : toute partie non vide de $\mathbf{N}$ admet un plus petit élément.

(Est-il la peine de souligner que les parties infinies de $\mathbf{N}$ n'ont, elles, pas de plus grand élément. Il n'y a pas d'entier "∞" !)

Ces deux faits sont utiles pour effectuer des démonstrations par l'absurde ; ainsi si on vous demande de prouver qu'une partie de $\mathbf{N}$ est infinie, ce peut être un bon réflexe de commencer par la supposer finie, de considérer son plus grand élément puis travailler jusqu'à trouver cela absurde —souvent en construisant un élément encore plus grand. Symétriquement, si on vous demande de prouver qu'une équation n'a aucune solution entière, ce peut être astucieux de supposer l'ensemble des solutions non vide, de considérer la plus petite solution puis travailler jusqu'à trouver cela absurde —et là souvent en construisant une solution encore plus petite.

Bref, vous le saviez déjà, mais c'est encore mieux en sachant que vous le savez.

3 - Dénombrabilité

L'objectif est de parvenir à distinguer des ensembles infinis moins gros que les autres - ceux "de la taille" de $\mathbf{N}$. Intuitivement, un ensemble dénombrable est un ensemble dont les points peuvent être numérotés : on appelle x_0 le premier, x_1 le second, x_3 le troisième, et ainsi de suite... jusqu'à les épuiser tous (en faisant une infinité d'efforts tout de même). Très informellement, de tels ensembles devraient garder un aspect de nuages de points - avec un peu d'habitude du concept il peut paraître intuitif que $\mathbf{R}$ ne saurait être un ensemble dénombrable : avec son aspect géométrique de droites, il a manifestement trop d'éléments.

Le concept ne sera utilisé nulle part ailleurs dans le cours de cette année, mais fera d'occasionnelles (mais importantes) apparitions en probabilités en deuxième année et une première couche ne peut donc pas faire de mal.

Définition 2-3-33 : On dit qu'un ensemble E est **dénombrable** lorsqu'il existe une bijection entre $\mathbf{N}$ et E.

Exemples : L'ensemble $\mathbf{N}$ est lui-même dénombrable. L'ensemble $\mathbf{N}^*$ est dénombrable : considérer l'application f de $\mathbf{N}$ vers $\mathbf{N}^*$ définie par $f(n) = n+1$ pour tout n de $\mathbf{N}$. On va démontrer ci-dessous que $\mathbf{Z}$ est dénombrable (par une preuve rédigée de façon volontairement lourde, histoire d'illustrer tant qu'il est encore frais le théorème concernant la caractérisation des bijections par l'existence d'une réciproque). L'ensemble $\mathbf{N}^2$ est dénombrable -ça s'explique bien avec un dessin, c'est beaucoup plus énigmatique si on donne seulement une formule, plaisir auquel je ne parviens à résister ; ainsi $g : \mathbf{N}^2 \to \mathbf{N}$ définie pour tout (s,t) de $\mathbf{N}^2$ par $g(s,t) = \dfrac{(s+t)(s+t+1)}{2} + t$ est elle une bijection entre $\mathbf{N}^2$ et $\mathbf{N}$. Il me reste à prouver, comme promis la :

Proposition 2-3-5 : $\mathbf{Z}$ est dénombrable.

Démonstration : Définissons deux applications $f : \mathbf{N} \to \mathbf{Z}$ et $g : \mathbf{Z} \to \mathbf{N}$ par les formules respectives suivantes :

$$\text{pour tout } n \in \mathbf{N}, \quad \begin{cases} f(n) = -n/2 & \text{si } n \text{ est pair} \\ f(n) = \dfrac{n+1}{2} & \text{si } n \text{ est impair} \end{cases}$$

et

$$\text{pour tout } r \in \mathbf{Z}, \quad \begin{cases} g(r) = -2r & \text{si } r \leq 0 \\ g(r) = 2r-1 & \text{si } r > 0. \end{cases}$$

Il convient tout d'abord de vérifier que f et g sont “bien des applications” au sens suivant : il n'est pas tout à fait clair que les formules qui les définissent fournissent un résultat situé dans l'ensemble d'arrivée demandé.

* Vérifions que f définit bien une application. Si n est pair, $n/2$ (qui est *a priori* seulement une fraction) est bien lui-même un entier, si n est impair, $n+1$ est pair et donc $\dfrac{n+1}{2}$ est lui aussi entier. La formule proposée pour $f(n)$ définit donc bien un élément de $\mathbf{Z}$.

* Vérifions que g définit bien une application. Si r est négatif, $-2r$ est positif, donc bien dans $\mathbf{N}$; si r est strictement positif, $2r$ vaut au moins 2 et donc $2r-1$ est aussi dans $\mathbf{N}$ (et est même strictement positif).

Vérifions maintenant que $g \circ f$ est bien l'application identique. Prenons un n dans $\mathbf{N}$.

* Si n est pair, $f(n) = -n/2$ est négatif ; on calcule donc $g[f(n)] = g[-n/2]$ par la première formule pour g et on trouve :

$$g[f(n)] = -2\,(-n/2) = n.$$

* Si n est impair, $f(n) = \dfrac{n+1}{2}$ est strictement positif ; on calcule donc $g[f(n)] = g[\dfrac{n+1}{2}]$ par la deuxième formule pour g et on trouve :

$$g[f(n)] = 2\left(\frac{n+1}{2}\right) - 1 = n.$$

La conjonction des deux cas prouve bien que $g \circ f = Id_{\mathbf{N}}$.

Vérifions maintenant que $f \circ g$ est bien l'application identique. Prenons un r dans $\mathbf{Z}$.

* Si r est négatif, $g(r) = -2r$ est pair ; on calcule donc $f[g(r)] = f[-2r]$ par la première formule pour f et on trouve :

$$f[g(r)] = -[-2r/2] = r.$$

* Si r est négatif, $g(r) = 2r-1$ est impair ; on calcule donc $f[g(r)] = f[2r-1]$ par la deuxième formule pour f et on trouve :

$$f[g(r)] = \frac{(2r-1)+1}{2} = r.$$

La conjonction des deux cas prouve bien que $f \circ g = Id_{\mathbf{Z}}$.

Tout ceci prouve que f est une bijection, et donc que $\mathbf{Z}$ est dénombrable. •

Pour en finir avec les ensembles dénombrables, deux propriétés d'aspect plus ou moins évident selon la précision de votre intuition de la question, et que je n'essaierai pas de prouver :

Proposition 2-3-6 : Toute partie d'un ensemble dénombrable est finie ou dénombrable. •

Proposition 2-3-7 : La réunion de deux ensembles dénombrable est dénombrable.

•

Passons à la non-dénombrabilité. Comme je l'ai déjà écrit plus haut, $\mathbf{R}$ n'est pas dénombrable, car trop gros (la preuve n'est pas infaisable, mais est tout de même relativement difficile par rapport à ce que nous faisons cette année, disons raisonnablement intéressante pour la licence). On peut montrer de façon très brève que le très gros ensemble des parties de $\mathbf{N}$ n'est pas dénombrable. Avertissement : la preuve qui suit est brève mais obscure, elle est surtout là pour donner un exemple de preuve ingénieuse ; ne vous affolez pas si elle vous affole, et passez au plus vite à la suite.

Proposition 2-3-8 : L'ensemble des parties de $\mathbf{N}$ n'est pas dénombrable.

Démonstration : Supposons qu'il existe une bijection f de $\mathbf{N}$ vers $\mathcal{P}(\mathbf{N})$. Considérons alors l'ensemble B défini par :

$$B = \{n \in \mathbf{N} \mid n \notin f(n)\}.$$

Puisque f est une bijection, il existe un b dans $\mathbf{N}$ tel que $f(b) = B$. Maintenant, si b est élément de B, par définition de B, c'est que $b \notin f(b)$ et comme $f(b) = B$ on a donc $b \notin B$ —c'est absurde. Réciproquement, si b n'est pas élément de B, par le même cheminement, $b \in f(b)$ puis $b \in B$ et là encore c'est absurde.

L'existence de f conduit donc à une absurdité : f ne peut exister.

•

4 - Deux définitions que je ne sais où caser, pourquoi pas là ?

Définition 2-4-34 : Soit E un ensemble. Une **suite** d'éléments de E est une application de $\mathbf{N}$ vers E.

Comme vous le savez certainement déjà, l'usage est d'utiliser une notation pour les suites différente de celle utilisée pour les applications "ordinaires". Une suite sera notée $(x_n)_{n\in\mathbf{N}}$ (où plus brièvement (x_n)) au lieu du simple "f" des applications ; sa valeur en l'entier n sera notée x_n (au lieu du $f(n)$ des applications).

Bien évidemment, des variantes sont possibles : suites indexées par $\mathbf{N}^*$, $\mathbf{Z}$, etc...

Définition 2-4-35 : Soit E un ensemble et $k \geq 0$. Un **k-uplet** d'éléments de E est une application de $\{1, 2, \ldots, k\}$ vers E.

De façon analogue à ce qui se passe avec les suites, le k-uplet sera noté $(x_1, x_2, \cdots, x_k)$ et sa valeur en l'entier n (sa "n-ème coordonnée" sera notée x_n. Des variantes dans l'indexation —notamment des k-uplets indexés par $\{0, 1, \ldots, k-1\}$— sont bien sûr possibles.

Les lecteurs observateurs croiront peut-être ma définition insensée pour $k = 0$ mais ils ont tort car je sous-entends que l'ensemble $\{1, 2, \ldots, k\}$ désigne alors le vide, et il est tout à fait cohérent d'accepter alors l'existence d'un 0-uplet que je noterai () ne contenant aucun élément. Je n'aime pas épiloguer sur ce genre de cas dégénéré, mais le 0-uplet apparaîtra épisodiquement en algèbre linéaire (c'est la base de $\{0\}$) alors mettons par avance les choses au point, en insistant sur le peu d'importance de cette remarque.

Enfin un mot supplémentaire pour éviter de trop prononcer le peu euphonique mot "k-uplet".

Définition 2-4-36 : Soit E un ensemble. On appelle **système** d'éléments de E tout objet qui est un k-uplet d'éléments de E pour un $k \geq 0$.

Chapitre 3 - Rudiments d'algèbre linéaire : l'espace $\mathbf{R}^n$

0 - Quelques conventions de notations

Pour chaque $n \geq 0$, la notation $\mathbf{R}^n$ désigne l'ensemble des n-uplets de réels. On identifiera abusivement (pour la lisibilité !) $\mathbf{R}^1$ à $\mathbf{R}$ (ainsi on notera par exemple 127 l'élément (127) de $\mathbf{R}^1$).

Notation 3-0-26 : On notera 0 l'élément $(0, \ldots, 0)$ de $\mathbf{R}^n$ (avec la même notation pour toute valeur de n).

Avec cette notation simplificatrice, on a : $\mathbf{R}^0 = \{0\}$ (sans la notation simplificatrice, $\mathbf{R}^0 = \{()\}$, ce qui est tout à fait correct mais peut laisser perplexe).

Enfin on manipulera tout le long du chapitre des systèmes d'éléments de $\mathbf{R}^n$, c'est-à-dire des listes de listes ; par exemple $((1,2,3),(4,5,6))$ est un 2-uplet d'éléments de $\mathbf{R}^3$ tandis que $((1,2),(3,4),(5,6))$ est un 3-uplet d'éléments de $\mathbf{R}^2$ et $((1,2,3,4,5,6))$ est un 1-uplet comportant un élément unique de $\mathbf{R}^6$.

Enfin on notera sans toujours le préciser explicitement $\underline{e}$ pour un k-uplet introduit sous forme $(e_1, \ldots, e_k)$.

1 - Addition et multiplication externe sur $\mathbf{R}^n$

Définition 3-1-37 : Soit $n \geq 0$ fixé, et soit $e = (x_1, \ldots, x_n)$ et $f = (y_1, \ldots, y_n)$. La **somme** de e et f est l'élément $(x_1 + y_1, \ldots, x_n + y_n)$ de $\mathbf{R}^n$.

Notation 3-1-27 : La somme de e et f est notée $e + f$.

Définition 3-1-38 : Soit $n \geq 0$ fixé, soit $e = (x_1, \ldots, x_n)$ et soit $\lambda \in \mathbf{R}$. Le **produit** de e par λ est l'élément $(\lambda x_1, \ldots, \lambda x_n)$ de $\mathbf{R}^n$.

Notation 3-1-28 : Le produit de e par λ est noté λe.

Proposition 3-1-9 : Pour chaque $n \geq 0$, l'ensemble $\mathbf{R}^n$ est un groupe commutatif pour l'addition.

Démonstration : Il n'y a qu'à vérifier méthodiquement l'associativité et la commutativité (évidentes !) l'existence d'un neutre (c'est 0) et celle des symétriques (le symétrique de $(x_1, \ldots, x_n)$ étant $(-x_1, \ldots, -x_n)$). •

Convention de vocabulaire : dans le contexte provisoire où nous nous trouvons, le mot **vecteur** sera utiliser pour désigner les éléments de $\mathbf{R}^n$, le mot **scalaire** pour les réels. Ainsi cela a un sens de multiplier entre eux deux scalaires, ou un scalaire par un vecteur, mais bien évidemment seuls les étourdis essaient de multiplier entre eux deux vecteurs.

Conventions de notation : quand $e_1, \ldots, e_k$ sont k vecteurs d'un même $\mathbf{R}^n$, la notation $e_1 + \cdots + e_k$ signifie très précisément $((\cdots((e_1 + e_2) + e_3) + \cdots + e_{k-1}) + e_k)$. Évidemment on peut déplacer les parenthèses à volonté (puisqu'il y a associativité). Cette définition sera complétée par la convention que cette somme vaut 0 lorsque $k = 0$ (en bonne rigueur, toute cette "convention de notation" devrait être appelée "définition" et être énoncée sous forme d'une définition faisant l'objet d'une récurrence sur k, mais ce serait à mon avis peu lisible).

2 - Combinaisons linéaires ; ensembles engendrés

Définition 3-2-39 : Soit $n \geq 0$ fixé, soit $(e_1, \ldots, e_k)$ un système d'éléments de $\mathbf{R}^n$, et soit f un élément de $\mathbf{R}^n$. On dit que f est une **combinaison linéaire** de $(e_1, \ldots, e_k)$ lorqu'il existe des scalaires $\alpha_1, \ldots, \alpha_k \in \mathbf{R}$ tels que $f = \alpha_1 e_1 + \cdots + \alpha_k e_k$.

Remarque : (pour les puristes) Cette définition devrait être écrite "lorsqu'il existe un k-uplet de scalaires $(\alpha_1, \ldots, \alpha_k)$" pour avoir aussi un sens lorsque $k = 0$; mais l'ajout de parenthèses fait à mon goût perdre de la lisibilité aussi je m'en dispense, et m'en dispenserai encore plusieurs fois.

Définition 3-2-40 : Soit $n \geq 0$ fixé et soit $\underline{e} = (e_1, \ldots, e_k)$ un système d'éléments de $\mathbf{R}^n$. On appelle **ensemble engendré** par $\underline{e}$ la partie de $\mathbf{R}^n$ ensemble des combinaisons linéaires de $\underline{e}$.

Notation 3-2-29 : Au moins trois notations sont usuelles pour l'ensemble engendré par $(e_1, \ldots, e_k)$. On peut le noter $\mathrm{Vect}\,(e_1, \ldots, e_k)$, ou $<e_1, \ldots, e_k>$, ou $\mathbf{R}e_1 + \cdots + \mathbf{R}e_k$(c'est la notation que j'utiliserai).

3 - Sous-espaces vectoriels de $\mathbf{R}^n$

Définition 3-3-41 : Soit $n \geq 0$ fixé. On dit qu'un sous-ensemble E de $\mathbf{R}^n$ est un **sous-espace vectoriel** de $\mathbf{R}^n$ lorsque les trois conditions suivantes sont vérifiées :
(i) E n'est pas vide.
(ii) Pour tous u, v de E, la somme $u + v$ est aussi dans E.
(iii) Pour tout u de E et tout scalaire λ, le produit λu est aussi dans E.

Voici une propriété utile, très peu profonde mais qui permet parfois d'économiser un peu de papier lors de vérifications faciles.

Proposition 3-3-10 : Soit $n \geq 0$ fixé. Un sous-ensemble E de $\mathbf{R}^n$ est un sous-espace vectoriel de $\mathbf{R}^n$ si et seulement si les deux conditions suivantes sont vérifiées :
(*1*) E n'est pas vide.
(*2*) Pour tous u, v de E et tout scalaire λ, le vecteur $u + \lambda v$ est aussi dans E.

Démonstration : Il s'agit d'une équivalence, vérifions la double implication...
Supposons que E est un sous-espace vectoriel de $\mathbf{R}^n$, c'est-à-dire qu'il vérifie (i),(ii) et (iii). Il est alors clair que (*1*) —qui coïncide avec (i) !— est vérifiée.

Montrons que E vérifie (*2*). Soit u, v deux éléments de E et λ un scalaire. En appliquant (iii) à u et λ, on constate que λu est aussi dans E, puis en appliquant (ii) à u et λv que la somme $u + \lambda v$ aussi. C'est déjà fini !

Supposons maintenant que E vérifie (*1*) et (*2*). Vérifier (i) est bien sûr sans problème.

Montrons que E vérifie (ii). Soit u, v deux éléments de E. En appliquant (*2*) à u, v et 1, on obtient $u + 1v \in E$, c'est-à-dire $u + v \in E$.

Montrons préalablement que $0 \in E$. En effet E n'étant pas vide, on peut prendre un élément w dans E, puis appliquer l'hypothèse (*2*) à w, w et -1 pour conclure que $w + (-1)w = 0 \in E$.

Montrons que E vérifie (iii). Soit u un élément de E et λ un scalaire. Maintenant qu'on sait que $0 \in E$, on peut appliquer (*2*) à 0, u et λ pour obtenir $0 + \lambda u \in E$, c'est-à-dire $\lambda u \in E$.

•

La proposition qui suit nous donne d'un coup tout plein d'exemples de sous-espaces (on verra même dès le prochain chapitre que tous les sous-espaces de $\mathbf{R}^n$ sont de cette forme).

Proposition 3-3-11 : Soit $n \geq 0$ fixé et soit $(e_1, \ldots, e_k)$ un système de vecteurs de $\mathbf{R}^n$. L'ensemble de leurs combinaisons linéaires, soit l'ensemble $\mathbf{R}e_1 + \cdots + \mathbf{R}e_k$ est un sous-espace vectoriel de $\mathbf{R}^n$.

Démonstration : C'est une simple vérification sans astuces.
Vérification de (*1*) : on peut écrire $0 = 0e_1 + \cdots + 0e_k$. Le vecteur nul est donc combinaison linéaire de $(e_1, \ldots, e_k)$, et E n'est donc pas vide.
Vérification de (*2*) : soit u et v deux éléments de E. On peut donc trouver des scalaires $\alpha_1, \ldots, \alpha_k$ et $\beta_1, \ldots, \beta_k$ tels que $u = \alpha_1 e_1 + \cdots + \alpha_k e_k$ et $v = \alpha_1 e_1 + \cdots + \alpha_k e_k$. Soit maintenant en outre λ un scalaire. On a alors :

$$\begin{aligned} u + \lambda v &= (\alpha_1 e_1 + \cdots + \alpha_k e_k) + \lambda\,(\beta_1 e_1 + \cdots + \beta_k e_k) \\ &= (\alpha_1 + \lambda\beta_1)e_1 + \cdots + (\alpha_k + \lambda\beta_k)e_k. \end{aligned}$$

Donc $u + \lambda v \in E$.

•

Cette proposition justifie donc de considérer comme obsolète l'expression “sous-ensemble engendré” à peine une page après son introduction ; on dira désormais “sous-espace engendré”.

4 - Systèmes générateurs, systèmes libres, bases

Définition 3-4-42 : Soit $n \geq 0$ fixé, $(g_1, \ldots, g_k)$ un système de vecteurs de $\mathbf{R}^n$ et E un sous-espace de $\mathbf{R}^n$. On dit que $\underline{g}$ est **générateur** de E (ou qu'il **engendre** E) lorsque $E = \mathbf{R}g_1 + \cdots + \mathbf{R}g_k$.

On remarquera (est-ce la peine de le dire ?) que comme chacun des e_i $(1 \leq i \leq k)$ est dans $\mathbf{R}e_1 + \cdots + \mathbf{R}e_k$, tous les vecteurs d'un système générateur de E sont forcément des vecteurs de E.

Définition 3-4-43 : Soit $n \geq 0$ fixé et $(f_1, \ldots, f_k)$ un système de vecteurs de $\mathbf{R}^n$. On dit que $\underline{f}$ est **libre** lorsque :

pour tous $\lambda_1, \ldots, \lambda_k \in \mathbf{R}$, si $\lambda_1 f_1 + \cdots + \lambda_k f_k = 0$, alors $\lambda_1 = \cdots = \lambda_k = 0$.

Définition 3-4-44 : On dit qu'un système de vecteurs de $\mathbf{R}^n$est **lié** lorsqu'il n'est pas libre.

Si la définition précise d'un système libre est indubitablement à connaître sur le bout du doigt pour pouvoir manipuler correctement cette notion, tenter de la réexprimer avec des mots peut permettre de mieux comprendre ce qu'elle raconte (je n'en suis pas si sûr !).

On doit être capable d'écrire automatiquement la propriété négation de celle définissant la liberté, ce qui n'est pas si simple (négation d'une implication...). Je le fais pour vous ci-dessous, mais vous devez savoir reconstituer ce qui suit plutôt que de l'apprendre sottement par cœur :

un système $(e_1, \ldots, e_k)$ est lié s'il existe des scalaires $\alpha_1, \ldots, \alpha_k$ tels que $\alpha_1 e_1 + \cdots + \alpha_k e_k = 0$ et $\lambda_1 \neq 0$ ou ... ou $\lambda_k \neq 0$.

Revoici la même formulation, avec un peu moins de symboles et un peu plus de mots :

un système $(e_1, \ldots, e_k)$ est lié s'il existe des scalaires $\alpha_1, \ldots, \alpha_k$ non tous nuls tels que $\alpha_1 e_1 + \cdots + \alpha_k e_k = 0$.

En la lisant, voyez-vous bien la nuance entre "non tous nuls" (que j'ai, avec raison, écrit) et "tous non nuls" ?

Une reformulation de la liberté, plus vague :

un système est libre si la seule façon d'obtenir 0 comme combinaison linéaire de ce système est la façon évidente.

Définition 3-4-45 : Soit $n \geq 0$ fixé, E un sous-espace vectoriel de $\mathbf{R}^n$ et $(e_1, \ldots, e_k)$ un système de vecteurs. On dit que $\underline{e}$ est une **base** de E lorsque $\underline{e}$ est libre et engendre E.

5 - Propriétés élémentaires des systèmes générateurs, des systèmes libres

Proposition 3-5-12 : Un système où un vecteur est répété est lié.

Démonstration : Soit $(e_1, \ldots, e_k)$ un tel système, où on suppose que pour deux indices i et j qui vérifient $1 \leq i < j \leq k$, $e_i = e_j$.

Prenons alors $\lambda_1 = \cdots = \lambda_{i-1} = 0$, $\lambda_i = 1$, $\lambda_{i+1} = \cdots = \lambda_{j-1} = 0$, $\lambda_j = -1$, $\lambda_{j+1} = \cdots = \lambda_k = 0$.

On a alors :

$$\begin{aligned}
\lambda_1 e_1 + \cdots + \lambda_k e_k &= 0e_1 + \cdots + 0e_{i-1} + 1e_i + 0e_{i+1} + \cdots + 0e_{j-1} + (-1)e_j + 0e_{j+1} + \cdots + 0e_k \\
&= e_i - e_j \\
&= 0.
\end{aligned}$$

Pourtant les coefficients ne sont pas tous nuls (λ_i, notamment, ne l'est pas).

•

Proposition 3-5-13 : Si on enlève un vecteur à un système libre, le nouveau système est toujours libre. (Précisément : si $(f_1, \ldots, f_k)$ est libre, pour tout i tel que $1 \leq i \leq k$, le système $(f_1, \ldots, f_{i-1}, f_{i+1}, \ldots, f_k)$ l'est encore).

Démonstration : Soit $(\lambda_1, \ldots, \lambda_{i-1}, \lambda_{i+1}, \ldots, \lambda_k)$ des scalaires tels que

$$\lambda_1 f_1 + \cdots + \lambda_{i-1} f_{i-1} + \lambda_{i+1} f_{i+1} + \cdots + \lambda_k f_k = 0.$$

Si on pose alors $\lambda_i = 0$, on a $\lambda_i f_i = 0$, donc

$$\begin{aligned}
\lambda_1 f_1 + \cdots + \lambda_k f_k &= (\lambda_1 f_1 + \cdots + \lambda_{i-1} f_{i-1} + \lambda_{i+1} f_{i+1} + \cdots + \lambda_k f_k) + \lambda_i f_i \\
&= 0 + 0 = 0.
\end{aligned}$$

Vu la liberté du "gros" système $(f_1, \ldots, f_k)$, on en déduit $\lambda_1 = \cdots = \lambda_k = 0$, d'où *a fortiori*

$$\lambda_1 = \cdots = \lambda_{i-1} = \lambda_{i+1} = \cdots = \lambda_k = 0.$$

•

Proposition 3-5-14 : Si on ajoute un vecteur à un système générateur, le nouveau système est toujours générateur. (Précisément : si $(g_1, \ldots, g_k)$ est générateur d'un sous-espace E d'un certain $\mathbf{R}^n$, pour tout vecteur e de E, le système $(g_1, \ldots, g_k, e)$ l'est également).

Démonstration : Soit donc e un vecteur de E et posons $F = \mathbf{R}g_1 + \cdots + \mathbf{R}g_k + \mathbf{R}e$.

Il nous faut montrer l'égalité $E = F$, montrons donc la double inclusion.

* Montrons que $E \subset F$. Soit u un vecteur de E. Comme $(g_1, \dots, g_k)$ engendre E, il existe des scalaires $\alpha_1, \dots, \alpha_k$ tels que $u = \alpha_1 g_1 + \dots + \alpha_k g_k$. On a alors aussi $u = \alpha_1 g_1 + \dots + \alpha_k g_k + 0e$, donc u est aussi combinaison linéaire de $(g_1, \dots, g_k, e)$.
* Montrons que $F \subset E$. Tous les g_i sont dans E, ainsi que e, donc aussi toutes leurs combinaisons linéaires. •

La proposition suivante est extrêmement utile dans les exercices pratiques :
Proposition 3-5-15 : Soit $n \geq 0$ fixé,
(a) Pour tout $e \in \mathbf{R}^n$, (e) est libre si et seulement si $e \neq 0$.
(b) Pour tous $e_1, e_2 \in \mathbf{R}^n$, (e_1, e_2) est libre si et seulement si $e_1 \neq 0$ et $e_2 \neq \lambda e_1$ pour tout scalaire λ.
(c) Pour tout $k \geq 1$ et tous $e_1, \dots, e_k \in \mathbf{R}^n$, $(e_1, \dots, e_k)$ est libre si et seulement si $(e_1, \dots, e_{k-1})$ est libre et e_k n'est pas combinaison linéaire de $(e_1, \dots, e_{k-1})$.
Démonstration : On observera que (b) n'est qu'un cas particulier de (c), énuméré à part pour aider à s'en souvenir. Le seul point à montrer sérieusement est donc le point (c). Attaquons-le. Il s'agit d'une équivalence, il est raisonnable de vérifier successivement les deux implications.
* Preuve de $\Rightarrow$ (par contraposition). Supposons donc que $(e_1, \dots, e_{k-1})$ est lié ou que e_k est combinaison linéaire de $(e_1, \dots, e_{k-1})$, et montrons que $(e_1, \dots, e_k)$ est lié.

- Dans la première éventualité, la proposition 3-5-13 (sous sa forme contraposée) nous prouve que $(e_1, \dots, e_k)$ est lié.
- Dans la seconde éventualité, il existe donc des scalaires $\alpha_1, \dots, \alpha_{k-1}$ permettant d'écrire $e_k = \alpha_1 e_1 + \dots + \alpha_{k-1} e_{k-1}$. On en déduit aussitôt que $\alpha_1 e_1 + \dots + \alpha_{k-1} e_{k-1} + (-1) e_k = 0$. Ceci fournit une combinaison linéaire de $e_1, \dots, e_k$ nulle bien que tous ses coefficients ne le soient pas (le dernier ne l'est pas) : le système $(e_1, \dots, e_k)$ est donc lié.

L'implication est donc prouvée.
* Preuve de $\Leftarrow$ (par contraposition également !). Supposons donc que $(e_1, \dots, e_k)$ est lié, et prouvons que $(e_1, \dots, e_{k-1})$ est lié ou que e_k est combinaison linéaire de $(e_1, \dots, e_{k-1})$.
Vu l'hypothèse, il existe des scalaires λ_i non tous nuls tels que :

$$(*) \qquad \lambda_1 e_1 + \dots + \lambda_k e_k = 0.$$

On va distinguer deux cas, selon que λ_k est nul ou non.

- Si $\lambda_k = 0$, la relation $(*)$ se réduit à $\lambda_1 e_1 + \dots + \lambda_{k-1} e_{k-1} = 0$. De plus, les λ_i pour $1 \leq i \leq k-1$ ne sont pas tous nuls. Le système $(e_1, \dots, e_{k-1})$ est donc lié.
- Si $\lambda_k \neq 0$, la relation $(*)$ peut être regroupée sous la forme :

$$e_k = -\frac{\lambda_1}{\lambda_k} e_1 + \dots + -\frac{\lambda_{k-1}}{\lambda_k} e_{k-1}.$$

On constate alors que e_k est combinaison linéaire de $e_1, \dots, e_{k-1}$.

Les deux implications sont prouvées, donc l'équivalence.
Je n'écris pas la preuve de (a) qui n'est autre que (c) lorsque $k = 1$ (il faut alors comprendre que $(e_1, \dots, e_{k-1})$ signifie ()). C'est l'étude idiote d'un cas dégénéré... •

Bien que les définitions en soient assez dissemblables, les concepts de système générateur et de système libre sont plus apparentés qu'on ne pourrait le croire. Le parallèle sera frappant sur cette
Proposition 3-5-16 : Soit $n \geq 0$ fixé et E un sous-espace vectoriel de $\mathbf{R}^n$. Soit $(e_1, \dots, e_k)$ un système de vecteurs de E. Alors :
(a) $(e_1, \dots, e_k)$ engendre E si et seulement si pour tout $v \in E$, il existe **au moins un** k-uplet de scalaires $(\alpha_1, \dots, \alpha_k)$ tel que $v = \alpha_1 e_1 + \dots + \alpha_k e_k$.
(b) $(e_1, \dots, e_k)$ est libre si et seulement si pour tout $v \in E$, il existe **au plus un** k-uplet de scalaires $(\alpha_1, \dots, \alpha_k)$ tel que $v = \alpha_1 e_1 + \dots + \alpha_k e_k$.
(c) $(e_1, \dots, e_k)$ est une base de E si et seulement si pour tout $v \in E$, il existe **exactement un** k-uplet de scalaires $(\alpha_1, \dots, \alpha_k)$ tel que $v = \alpha_1 e_1 + \dots + \alpha_k e_k$.
Démonstration :
* Il n'y a quasiment rien à prouver dans (a), qui découle (presque) directement de la définition de "système générateur".

* La preuve de (b) est le morceau où il faut un peu travailler. Il s'agit d'une équivalence, montrons successivement les deux implications.

• Preuve de $\Rightarrow$. Supposons le système $(e_1, \ldots, e_k)$ libre. Soit v un vecteur de E, et supposons que les scalaires $\alpha_1, \ldots, \alpha_k$ et $\beta_1, \ldots, \beta_k$ permettent d'écrire

$$\begin{aligned} & v = \alpha_1 e_1 + \cdots + \alpha_k e_k \\ \text{et aussi} \qquad & v = \beta_1 e_1 + \cdots + \beta_k e_k. \end{aligned}$$

En soustrayant ces deux inégalités, on obtient alors :

$$0 = (\alpha_1 - \beta_1)e_1 + \cdots + (\alpha_k - \beta_k)e_k.$$

La liberté de $(e_1, \ldots, e_k)$ fournit alors $\alpha_1 - \beta_1 = \cdots = \alpha_k - \beta_k = 0$, ou, avec des mots, l'unicité de l'écriture de v.

• Preuve de $\Leftarrow$ (par contraposition). Supposons le système $(e_1, \ldots, e_k)$ lié. Il existe donc des scalaires non tous nuls $\lambda_1, \ldots, \lambda_k$ tels que $\lambda_1 e_1 + \cdots + \lambda_k e_k = 0$. Mais on peut aussi écrire 0 d'une autre façon, à savoir $0 = 0e_1 + \cdots + 0e_k$ (c'est bien une "autre" façon puisque les λ_i ne sont pas tous nuls). Il existe donc un vecteur de E qui possède plus d'une écriture.

L'équivalence est donc prouvée.

* L'énoncé (c) n'est que la synthèse des deux autres.

•

6 - Coordonnées et matrices des vecteurs

La proposition qui précède donne dès lors un sens à la

Définition 3-6-46 : Soit $n \geq 0$ fixé, E un sous-espace de $\mathbf{R}^n$, $(e_1, \ldots, e_k)$ une base de E et v un vecteur de E. On appelle **coordonnées** de v dans la base $\underline{e}$ les réels uniquement déterminés $\alpha_1, \ldots, \alpha_k$ tels que :

$$v = \alpha_1 e_1 + \cdots + \alpha_k e_k.$$

Notation 3-6-30 : Pour des motivations qui apparaitront plus tard, on prendra dès maintenant l'habitude de ranger les coordonnées $\alpha_1, \ldots, \alpha_k$ de v dans une grande colonne entre deux parenthèses, sous la forme :

$$\mathrm{mat}_{\underline{e}}(v) = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix}.$$

On peut déjà donner une justification de l'intérêt de cette notation : elle permet de ne pas confondre un vecteur $(x_1, \ldots, x_n)$ de $\mathbf{R}^n$, noté horizontalement et avec des virgules, et l'objet formé en regroupant ses coordonnées dans telle ou telle base de $\mathbf{R}^n$.

7 - Informations non généralisables à tout espace vectoriel

Cette section contient quelques définitions propres à $\mathbf{R}^n$ (qui pourront se généraliser à $\mathbf{K}^n$, pour $\mathbf{K}$ corps commutatif, mais pas plus loin).

Tout d'abord, pour éviter les confusions avec les "coordonnées" dans telle ou telle base, j'aime bien mettre les points sur les ı en ajoutant un peu de vocabulaire.

Définition 3-7-47 : Pour $(x_1, \ldots, x_n)$ élément de $\mathbf{R}^n$ j'appellerai **composantes** de $(x_1, \ldots, x_n)$ les réels $x_1, \ldots, x_n$.

Il existe dans $\mathbf{R}^n$ une base particulièrement simple :

Proposition 3-7-17 : Le système $\big((1,0,\ldots,0,0),(0,1,\ldots,0,0),\ldots,(0,0,\ldots,1,0),(0,0,\ldots,0,1)\big)$ est une base de $\mathbf{R}^n$.

Démonstration : Est-ce vraiment la peine de la faire ? Un vecteur $(x_1, \ldots, x_n)$ de $\mathbf{R}^n$ admet l'écriture $\alpha_1(1,0,\ldots,0,0)+\alpha_2(0,1,\ldots,0,0)+\ldots+\alpha_{n-1}(0,0,\ldots,1,0)+\alpha_n(0,0,\ldots,0,1) = (\alpha_1, \ldots, \alpha_n)$ si et seulement si pour chaque i entre 1 et n, on a $x_i = \alpha_i$. Il possède donc une et une seule écriture, propriété qui caractérise les bases.

•

On notera au passage que dans cette base, les coordonnées de $\underline{x}$ sont exactement ses composantes, et on prendra garde que c'est justement très spécifique et ne fonctionne dans aucune autre base de $\mathbf{R}^n$!
Définition 3-7-48 : La base $\big((1,0,\ldots,0,0),(0,1,\ldots,0,0),\ldots,(0,0,\ldots,1,0),(0,0,\ldots,0,1)\big)$ de $\mathbf{R}^n$ est appelée la **base canonique** de $\mathbf{R}^n$.

8 - Opérations sur les sous-espaces

Proposition 3-8-18 : Soit $n \geq 0$ fixé et E, F deux sous-espaces de $\mathbf{R}^n$. Alors $E \cap F$ est également un sous-espace de $\mathbf{R}^n$.
Démonstration : Elle est facile et peu intéressante. Vérifions le plus vite possible les propriétés (*1*) et (*2*) de la caractérisation des sous-espaces.
* (*1*) est vraie parce que 0 est à la fois dans E et dans F.
* (*2*) Soit u et v deux vecteurs de $E \cap F$, et λ un scalaire. Alors $u + \lambda v \in E$ parce que E est un sous-espace, et $u + \lambda v \in F$ parce que F est un sous-espace. Et donc $u + \lambda v \in E \cap F$.
•

Attention ! Ce qui marche avec l'intersection **ne marche pas du tout** avec la réunion. Il suffit de ne pas perdre la géométrie de vue et de faire un dessin (deux droites de $\mathbf{R}^2$!) pour s'en convaincre.

C'est pourquoi on introduit une autre opération, la somme des sous-espaces, qui remplace au pied levé l'inefficace réunion.
Définition 3-8-49 : Soit $n \geq 0$ fixé et A, B deux parties de $\mathbf{R}^n$. On appelle **somme** de A et B la partie de $\mathbf{R}^n$ formée des vecteurs v pour lesquels il existe $a \in A$ et $b \in B$ tels que $v = a + b$.
Notation 3-8-31 : La somme de A et B sera notée $A + B$.

On remarquera que j'ai donné la définition pour des parties quelconques, car ça ne coûte pas plus cher, mais qu'on ne s'en servira que pour des sous-espaces vectoriels.
Proposition 3-8-19 : Soit $n \geq 0$ fixé et E, F deux sous-espaces de $\mathbf{R}^n$. Alors $E + F$ est également un sous-espace de $\mathbf{R}^n$.
Démonstration : Encore une vérification ennuyeuse...
* (*1*) est vraie parce que 0 peut s'écrire $0 + 0$, et que dans cette écriture, on peut voir le premier 0 comme un élément de E et le second comme un élément de F.
* (*2*) Soit u_1 et u_2 deux vecteurs de $E + F$, et λ un scalaire. Alors , comme $u_1 \in E + F$, on peut prendre des vecteurs $e_1 \in E$ et $f_1 \in F$ tels que $u_1 = e_1 + f_1$. On procède de même avec u_2.

On a alors : $u_1 + \lambda u_2 = (e_1 + f_1) + \lambda(e_2 + f_2) = (e_1 + \lambda e_2) + (f_1 + \lambda f_2)$. Dans cette nouvelle écriture, $e_1 + \lambda e_2 \in E$, car E est un sous-espace, et $f_1 + \lambda f_2 \in F$, car F est un sous-espace. Donc $u_1 + \lambda u_2 \in E + F$. •

Remarque : Il est ennuyeux et facile de vérifier que $+$, vu comme opération sur l'ensemble des sous-ensembles de $\mathbf{R}^n$, est associative et commutative. Ce qu'on utilisera implicitement fort fréquemment sans le dire.
Remarque : On a déjà introduit une notation $+$ entre sous-espaces, dans la notation pour le sous-espace engendré par un système de vecteurs. Heureusement, il n'y a pas de bavure : les deux notations sont bien cohérentes, très précisément on a bien $\mathrm{Vect}(e_1, \ldots, e_k) = \mathrm{Vect}(\mathrm{e}_1) + \cdots + \mathrm{Vect}(\mathrm{e}_k)$. Ce qui se vérifie aussitôt dès qu'on s'est donné la peine de l'énoncer.

Chapitre 4 - Dimension

1 - Le nœud des démonstrations

J'ai essayé de regrouper dans cette section les idées les plus ingénieuses de la preuve. Ainsi on traverse un passage ardu, mais tout ira plus facilement par la suite.

Commençons par un premier lemme (le "lemme d'échange"), assez facile à prouver, mais dont l'énoncé n'est pas de ceux qui viennent spontanément à l'esprit.

Lemme 4-1-1 : Soit $n \geq 0$ fixé et E un sous-espace de $\mathbf{R}^n$; soit $(f_1, \ldots, f_p)$ un système libre de vecteurs de E (avec $p \geq 1$) et $(g_1, \ldots, g_q)$ un système générateur de E. Alors il existe un indice j (où $1 \leq j \leq q$) tel que $(f_1, \ldots, f_{p-1}, g_j)$ soit encore un système libre.

Démonstration : (par l'absurde). Supposons que l'hypothèse "$(f_1, \ldots, f_p)$ est libre" soit vraie, mais que la conclusion soit fausse, c'est-à-dire que pour chaque j, $(f_1, \ldots, f_{p-1}, g_j)$ soit un système lié. Comme $(f_1, \ldots, f_p)$ est libre,$(f_1, \ldots, f_{p-1})$ est lui-même libre. Pour chaque indice j fixé, on a alors simultanément "$(f_1, \ldots, f_{p-1})$ est libre" et "$(f_1, \ldots, f_{p-1}, g_j)$ n'est pas libre". En utilisant la proposition 3-5-15, on en déduit que pour chaque j, g_j est une combinaison linéaire de $(f_1, \ldots, f_{p-1})$. Il existe donc des réels $a_{i,1}$ tels que :

$$g_1 = a_{1,1}f_1 + a_{2,1}f_2 + \cdots + a_{p-2,1}f_{p-2} + a_{p-1,1}f_{p-1}$$

puis des réels $a_{i,2}$ tels que :

$$g_2 = a_{1,2}f_1 + a_{2,2}f_2 + \cdots + a_{p-2,2}f_{p-2} + a_{p-1,2}f_{p-1}$$

et ainsi de suite jusqu'à :

$$\begin{aligned} g_{q-1} &= a_{1,q-1}f_1 + a_{2,q-1}f_2 + \cdots + a_{p-2,q-1}f_{p-2} + a_{p-1,q-1}f_{p-1} \\ g_q &= a_{1,q}f_1 + a_{2,q}f_2 + \cdots + a_{p-2,q}f_{p-2} + a_{p-1,q}f_{p-1} \end{aligned}$$

Maintenant il est temps d'utiliser le fait que $\underline{g}$ est génératrice de E : tout vecteur de E peut s'écrire comme combinaison linéaire de $g_1, \ldots, g_q$, et en particulier c'est le cas de f_p. Il existe donc des scalaires $b_1, \ldots, b_q$ tels que

$$f_p = b_1 g_1 + \cdots + b_q g_q.$$

En reportant alors les expressions des g_j dans cette égalité, on obtient (mentalement...) une expression de f_p comme combinaison linéaire de $f_1, \ldots, f_{p-1}$.

Ceci contredit l'hypothèse selon laquelle $(f_1, \ldots, f_p)$ est libre.

•

On va déduire de ce "lemme d'échange" un deuxième lemme, au résultat beaucoup plus parlant. (Il est toutefois inutile d'apprendre ce deuxième lemme, car on prouvera un peu plus loin des résultats encore plus performants).

Lemme 4-1-2 : Soit $n \geq 0$ fixé et E un sous-espace de $\mathbf{R}^n$; soit $(f_1, \ldots, f_p)$ un système libre de vecteurs de E (avec $p \geq 1$) et $(g_1, \ldots, g_q)$ un système générateur de E. Alors $p \leq q$.

Démonstration : Si $p = 0$, le résultat est évident. Sinon commençons par appliquer le lemme d'échange. Nous voilà devant un nouveau système libre $(f_1, \ldots, f_{p-1}, g_j)$. Qu'en faire ?

Une première initiative va être de donner un nom à l'indice j qui vient d'apparaître —d'autres vont arriver et il ne faut pas se noyer sous les lettres. Notons le $\varphi(p)$.

Une deuxième initiative est de faire tourner le système. Il est clair, à partir de la définition de la liberté, que le système $(g_{\varphi(p)}, f_1, \ldots, f_{p-1})$ est encore libre. Appliquons le lemme d'échange à ce nouveau système libre, et toujours au système générateur $\underline{g}$. Cette fois, nous notons $\varphi(p-1)$ l'indice du g_j qui a choisi la liberté, et voilà devant nous un nouveau système libre $(g_{\varphi(p)}, f_1, \ldots, f_{p-2}, g_{\varphi(p-1)})$ qu'on fait lui aussi tourner sur lui-même pour produire un nouveau système libre $(g_{\varphi(p-1)}, g_{\varphi(p)}, f_1, \ldots, f_{p-2})$. On réapplique le lemme d'échange à ce système, et ainsi de suite... Sans écrire la récurrence formelle, on est convaincu de pouvoir aboutir à un système libre $(g_{\varphi(1)}, \ldots, g_{\varphi(p)})$.

Maintenant on peut considérer φ comme une application définie sur $\{1,\ldots,p\}$, et à valeurs dans $\{1,\ldots,q\}$ (l'ensemble des indices des g_j). De plus, comme le système $(g_{\varphi(1)},\ldots,g_{\varphi(p)})$ est libre, il ne peut contenir de répétition, et donc pour $i \neq j$, $g_{\varphi(i)} \neq g_{\varphi(j)}$, et *a fortiori* $\varphi(i) \neq \varphi(j)$. L'application φ est donc une injection de l'ensemble fini $\{1,\ldots,p\}$ dans l'ensemble fini $\{1,\ldots,q\}$. Ceci entraîne que le premier a moins d'éléments que le second (au sens large), c'est-à-dire que $p \leq q$.

•

2 - Dimension. Première approche, où reste un trou

Théorème 4-2-2 : Soit $n \geq 0$ fixé et E un sous-espace de $\mathbf{R}^n$. Toutes les bases de E sont formées du même nombre de vecteurs.
Démonstration : Soit $(e_1,\ldots,e_k)$ et $(f_1,\ldots,f_l)$ deux bases de E.
En appliquant le lemme qui précède au système libre $\underline{e}$ et au système générateur $\underline{f}$, on obtient $k \leq l$.
En appliquant le lemme qui précède au système libre $\underline{f}$ et au système générateur $\underline{e}$, on obtient $l \leq k$.

•

Corollaire : Soit $n \geq 0$ fixé. Toutes les bases de $\mathbf{R}^n$ sont des n-uplets.
Démonstration : On connaît en effet une base de $\mathbf{R}^n$: la base canonique, formée de n vecteurs. Toutes les autres sont donc dans la même situation.

•

Il serait alors tentant de définir ici la dimension d'un sous-espace E comme le nombre d'éléments de l'une quelconque de ces bases. Mais en l'état de nos connaissances, la définition serait buggée. Nous n'avons en effet pas encore montré que E possède au moins une base, et il reste encore du travail à faire pour cela...

3 - Systèmes libres maximaux et générateurs minimaux

Les notions introduites dans cette section sont un peu subtiles, mais déjà d'usage plus fréquent que le lemme d'échange. En toute honnêteté, vous devriez pouvoir survivre quelque temps même si vous les assimilez mal, mais ça ne peut vous faire de mal de les connaître. Évidemment, ce sursaut de franchise est une occasion de rappeler qu'à peu près tout le reste est indispensable...
Définition 4-3-50 : Soit $n \geq 0$ fixé et E un sous-espace de $\mathbf{R}^n$. Un système libre $(f_1,\ldots,f_p)$ de vecteurs de E est dit **maximal** (dans E) lorsqu'on ne peut y ajouter un vecteur de E sans le rendre lié. (Plus formellement : lorsque pour tout v de E le système $(f_1,\ldots,f_p,v)$ est lié).
Définition 4-3-51 : Soit $n \geq 0$ fixé et E un sous-espace de $\mathbf{R}^n$. Un système générateur de E $(g_1,\ldots,g_q)$ E est dit **minimal** lorsqu'on ne peut lui enlever un vecteur sans lui faire perdre sa capacité génésique. (Plus formellement : lorsque pour tout indice j (avec $1 \leq j \leq q$) le système $(g_1,\ldots,g_{j-1},g_{j+1},\ldots,g_p)$ n'est pas générateur).
Proposition 4-3-20 : Soit $n \geq 0$ fixé et E un sous-espace de $\mathbf{R}^n$. Tout système libre maximal dans E est une base de E.
Démonstration : Soit $(f_1,\ldots,f_p)$ un tel système libre maximal. Montrons qu'il est générateur de E. Soit v un vecteur de E. Par l'hypothèse de maximalité, $(f_1,\ldots,f_p,v)$ est lié. Comme par ailleurs $(f_1,\ldots,f_p)$ est libre, on en déduit —par la proposition 3-5-15— que v est une combinaison linéaire de $(f_1,\ldots,f_p)$.

•

Et, symétriquement :
Proposition 4-3-21 : Soit $n \geq 0$ fixé et E un sous-espace de $\mathbf{R}^n$. Tout système générateur de E minimal est une base de E.
Démonstration : Soit $(g_1,\ldots,g_q)$ un tel système. Supposons que ce système ne soit pas libre. Il existerait alors des scalaires non tous nuls $\lambda_1,\ldots,\lambda_q$ tels que $\lambda_1 g_1 + \cdots + \lambda_q g_q = 0$. Soit i un indice tel que $\lambda_i \neq 0$; on peut alors écrire :

$$g_i = -\frac{\lambda_1}{\lambda_i}g_1 - \cdots - \frac{\lambda_{i-1}}{\lambda_i}g_{i-1} - \frac{\lambda_{i+1}}{\lambda_i}g_{i+1} - \cdots - \frac{\lambda_q}{\lambda_i}g_q.$$

On va voir que le système $(g_1,\ldots,g_{i-1},g_{i+1},\ldots,g_q)$ est encore générateur —contredisant la minimalité. Soit en effet v un vecteur de E. Comme $\underline{g}$ est générateur, il existe des coefficients $\alpha_1,\ldots,\alpha_q$ tels que $v = \alpha_1 g_1 + \cdots + \alpha_q g_q$.

Reportons dans cette écriture de v l'expression de g_i dont nous disposons : on obtient

$$\alpha_1 g_1 + \cdots + \alpha_{i-1} g_{i-1} + \alpha_i\big(-\frac{\lambda_1}{\lambda_i} g_1 - \cdots - \frac{\lambda_{i-1}}{\lambda_i} g_{i-1} - \frac{\lambda_{i+1}}{\lambda_i} g_{i+1} - \cdots - \frac{\lambda_q}{\lambda_i} g_q\big) + \alpha_{i+1} g_{i+1} + \cdots + \alpha_q g_q$$

$$= (\alpha_1 - \frac{\alpha_i \lambda_1}{\lambda_i}) g_1 + \cdots + (\alpha_{i-1} - \frac{\alpha_i \lambda_{i-1}}{\lambda_i}) g_{i-1} + (\alpha_{i+1} - \frac{\alpha_i \lambda_{i+1}}{\lambda_i}) g_{i+1} + \cdots + (\alpha_q - \frac{\alpha_i \lambda_q}{\lambda_i}) g_q$$

On a réussi à écrire v comme combinaison linéaire de $(g_1, \ldots, g_{i-1}, g_{i+1}, \ldots, g_q)$: cette famille est donc génératrice, ce qui contredit la minimalité qu'on avait supposée.

•

4 - Existence de bases pour les sous-espaces de $\mathbf{R}^n$

Avant d'aboutir au résultat annoncé, on va montrer un résultat qui méritera l'honneur d'être appelé "théorème" : le théorème de la base incomplète.

Théorème 4-4-3 : Soit $n \geq 0$ fixé et E un sous-espace de $\mathbf{R}^n$. Tout système libre de vecteurs de E peut être prolongé en une base de E par l'adjonction de nouveaux vecteurs de E (éventuellement aucun !) (Plus formellement : soit $(f_1, \ldots, f_p)$ un système libre de E. Il existe un entier $l \geq 0$ et des vecteurs $f_{p+1}, \ldots, f_{p+l}$ tels que $(f_1, \ldots, f_{p+l})$ soit une base de E.)

Démonstration : De deux choses l'une : ou bien le système considéré est libre maximal, et alors par la section précédente c'est déjà une base, ou bien il ne l'est pas, et on peut lui adjoindre un vecteur f_{p+1} en en faisant un système libre. Si ce nouveau système est maximal, c'est une base, et on a fini. Sinon on peut lui adjoindre un nouveau vecteur f_{p+2}. On peut ensuite continuer...

Reste à prouver qu'on s'arrêtera un jour. Si ce n'était pas le cas, on arriverait à posséder un système libre à $n+1$ composantes $(f_1, \ldots, f_{n+1})$. Mais dans $\mathbf{R}^n$ ce système serait libre avec $n+1$ composantes, tandis que la base canonique est génératrice avec n vecteurs.

Ceci contredit le "nœud des démonstrations".

•

Très symétriquement :

Théorème 4-4-4 : Soit $n \geq 0$ fixé et E un sous-espace de $\mathbf{R}^n$. De tout système générateur de E on peut extraire une base de E par suppression de certains vecteurs du système (éventuellement aucun !) (Je ne donne pas de version plus formelle, les trouvant trop peu lisibles).

Démonstration : Elle est basée sur le même principe : si mon système est minimal, c'est une base et j'ai fini. Sinon, j'en retranche un vecteur ; si le nouveau système est minimal, c'est une base et j'ai fini. Sinon je retranche un nouveau vecteur et ainsi de suite.

Reste à prouver qu'on s'arrête. Mais ici c'est stupide, parce qu'on ne peut évidemment plus rien soustraire quand on arrive, éventuellement, à 0 vecteur !

•

On remarquera la "fausse symétrie" entre les deux théorèmes : bien que leurs énoncés soient analogues, le premier repose sur les lemmes un peu subtils du début du chapitre, le second s'en passe fort bien.

Il reste à en déduire le résultat annoncé dans l'en-tête de la section

Théorème 4-4-5 : Soit $n \geq 0$ fixé et E un sous-espace de $\mathbf{R}^n$. Alors E possède des bases.

Démonstration : Le système () est un système libre de E. Si, si, c'est vrai, je vous le jure, relisez la définition dans le cas dégénéré... Si on n'y croit pas, on pourra se contenter de prendre un vecteur non nul e de E et observer que (e) est un système libre. (Mais bien sûr, ça ne marche pas dans le cas stupidement dégénéré où $E = \{0\}$!).

Appliquons alors le théorème de la base incomplète à ce système libre. On obtient une base de E.

•

On voit encore ici la cassure de la symétrie : il serait tentant de préférer montrer ce théorème à partir de celui sur les systèmes générateurs, qui est plus facile à prouver que le théorème de la base incomplète, mais ce serait voué à l'échec car on ne dispose pas de système générateur évident de E.

5 - Dimension des sous-espaces de $\mathbf{R}^n$

On a enfin tout le matériel pour énoncer la

Définition 4-5-52 : Soit $n \geq 0$ fixé et E un sous-espace de $\mathbf{R}^n$. On appelle **dimension** de E le nombre de composantes de n'importe quelle base de E.

Cette définition est cohérente : tout espace possède au moins une base, et l'usage de n'importe quelle base donnera le même résultat.
Notation 4-5-32 : La dimension de E sera notée $\dim E$.

Allez, j'offre le nom de "théorème" au résultat qui suit, bien que sa démonstration ne contienne guère d'idées nouvelles, mais pour que vous voyiez bien qu'il mérite d'être retenu (en pratique, il servira si souvent en TD que vous le connaitrez sans conscience de le connaître !)
Théorème 4-5-6 : Soit $n \geq 0$ fixé et E un sous-espace de $\mathbf{R}^n$.
(a) Tout système libre dans E est formée d'au plus $\dim E$ vecteurs.
(b) Tout système générateur de E est formée d'au moins $\dim E$ vecteurs.
(c) Un système libre dans E formée d'exactement $\dim E$ vecteurs est une base de E.
(d) Un système générateur de E formée d'exactement $\dim E$ vecteurs est une base de E.
Démonstration :
(a) D'après le théorème de la base incomplète, le système libre considéré peut se prolonger en une base de E. Cette base possède alors $\dim E$ vecteurs. Donc le système de départ en possédait moins (au sens large).
(b) C'est la même chose à partir du théorème analogue sur les systèmes générateurs.
(c) Le système libre considéré peut être prolongé en une base de E ; après prolongement, il est formé d'autant de vecteurs qu'avant prolongement (à savoir $\dim E$). Donc le prolongement est le prolongement par rien, et le système originel était déjà une base.
(d) Même raisonnement avec ablation de vecteurs.

•

Et maintenant, deux sous-espaces emboîtés à la fois !
Théorème 4-5-7 : Soit $n \geq 0$ fixé et E, F deux sous-espaces de $\mathbf{R}^n$.
(a) Si $E \subset F$, $\dim E \leq \dim F$.
(b) Si de plus l'inclusion est stricte, l'inégalité aussi.
Démonstration :

(a) Prenons une base de E qui est donc formée de $\dim E$ vecteurs ; c'est un système libre dans F. Donc par le (a) du théorème précédent, $\dim E \leq \dim F$.

(b) (Par contraposition) Supposons donc $E \subset F$ avec $\dim E = \dim F$. Prenons une base de E. C'est aussi un système libre dans F, et elle possède $\dim F$ éléments, donc c'est une base de F. L'espace engendré par ce système est donc à la fois E et F, d'où $E = F$.

•

6 - Une formule de Grassmann

Théorème 4-6-8 : Soit $n \geq 0$ fixé et E, F deux sous-espaces de $\mathbf{R}^n$.

$$\dim(E + F) = \dim E + \dim F - \dim(E \cap F).$$

Démonstration :
Notons $k = \dim E$, $l = \dim F$ et $t = \dim(E \cap F)$.
Partons d'une base $(g_1, \ldots, g_t)$ de $E \cap F$.
Par le théorème de la base incomplète appliquée à $E \cap F$ et E, il existe des vecteurs $e_1, \ldots, e_a$ (en notant $a = k - t$) tels que $(g_1, \ldots, g_t, e_1, \ldots, e_a)$ soit une base de E.
De même (en notant $b = l - t$), on fabrique une base $(g_1, \ldots, g_t, f_1, \ldots, f_b)$ de F.
J'affirme que $(g_1, \ldots, g_t, e_1, \ldots, e_a, f_1, \ldots, f_b)$ est une base de $E + F$. Il me reste à le prouver.
* Montrons tout d'abord que c'est un système générateur de $E + F$. Notons G le sous-espace de $\mathbf{R}^n$ engendré par $(g_1, \ldots, g_t, e_1, \ldots, e_a, f_1, \ldots, f_b)$. On va montrer la double inclusion entre $E + F$ et G.

  • Montrons que $E + F \subset G$. Soit $v \in E + F$ un vecteur ; il existe donc des vecteurs $e \in E$ et $f \in F$ tels que $v = e + f$.
  Comme $(g_1, \ldots, g_t, e_1, \ldots, e_a)$ engendre E, il existe des scalaires $\gamma_1, \ldots, \gamma_g, \alpha_1, \ldots, \alpha_a$ tels que

$$e = \gamma_1 g_1 + \cdots + \gamma_t g_t + \alpha_1 e_1 + \cdots + \alpha_a e_a.$$

  De même, il existe des scalaires $\delta_1, \ldots, \delta_t, \beta_1, \ldots, \beta_b$ tels que

$$f = \delta_1 g_1 + \cdots + \delta_t g_t + \beta_1 f_1 + \cdots + \beta_b f_b.$$

Dès lors

$$\begin{aligned} v = e + f &= \gamma_1 g_1 + \cdots + \gamma_t g_t + \alpha_1 e_1 + \cdots + \alpha_a e_a + \delta_1 g_1 + \cdots + \delta_t g_t + \beta_1 f_1 + \cdots + \beta_b f_b \\ &= (\gamma_1 + \delta_1) g_1 + \cdots + (\gamma_t + \delta_t) g_t + \alpha_1 e_1 + \cdots + \alpha_a e_a + \beta_1 f_1 + \cdots + \beta_b f_b \end{aligned}$$

et donc $v \in G$.
• Réciproquement, montrons que $G \subset E + F$. Soit $v \in G$.
Il existe donc des scalaires $\gamma_1, \ldots, \gamma_t, \alpha_1, \ldots, \alpha_a, \beta_1, \ldots, \beta_b$ tels que

$$\begin{aligned} v =& \gamma_1 g_1 + \cdots + \gamma_t g_t + \alpha_1 e_1 + \cdots + \alpha_a e_a + \beta_1 f_1 + \cdots + \beta_b f_b \\ &= \big(\gamma_1 g_1 + \cdots + \gamma_t g_t + \alpha_1 e_1 + \cdots + \alpha_a e_a\big) + \big(\beta_1 f_1 + \cdots + \beta_b f_b\big). \end{aligned}$$

Dans ce parenthésage, on voit qu'on a écrit v comme somme d'un vecteur de E et d'un vecteur de F, donc $v \in E + F$.
On a donc montré que $(g_1, \ldots, g_t, e_1, \ldots, e_a, f_1, \ldots, f_b)$ engendre $E + F$.
* Montrons maintenant que $(g_1, \ldots, g_t, e_1, \ldots, e_a, f_1, \ldots, f_b)$ est un système libre.
Soit des scalaires $\nu_1, \ldots, \nu_t, \lambda_1, \ldots, \lambda_a, \mu_1, \ldots, \mu_b$ tels que :

$$\nu_1 g_1 + \cdots + \nu_t g_t + \lambda_1 e_1 + \cdots + \lambda_a e_a + \mu_1 f_1 + \cdots + \mu_b f_b = 0.$$

Il faut ici être un peu ingénieux, car cette égalité ne laisse directement utiliser aucune des hypothèses de liberté dont on dispose pour l'instant.
Regroupons la différemment :

$$\nu_1 g_1 + \cdots + \nu_t g_t + \lambda_1 e_1 + \cdots + \lambda_a e_a = -\mu_1 f_1 + \cdots + -\mu_b f_b$$

et appelons h ce nouveau vecteur.
Vu sa première expression, h est un vecteur de E ; vu sa deuxième expression, c'est un vecteur de F. Ainsi h est un vecteur de $E \cap F$.
En tant que vecteur de $E \cap F$, il peut être écrit dans le système $(g_1, \ldots, g_t)$, base de $E \cap F$; soit donc des scalaires $\gamma_1, \ldots, \gamma_t$ tels que

$$h = \gamma_1 g_1 + \cdots + \gamma_t g_t.$$

Mettons côte à côte deux expressions de h :

$$\begin{aligned} h &= \gamma_1 g_1 + \cdots + \gamma_t g_t + 0 f_1 + \cdots + 0 f_b \\ &= 0 g_1 + \cdots + 0 g_t + -\mu_1 f_1 + \cdots + -\mu_b f_b. \end{aligned}$$

Mais le système $(g_1, \ldots, g_t, f_1, \ldots, f_b)$ est libre dans F, donc ces deux écritures doivent coïncider. On en déduit que $\gamma_1 = \cdots = \gamma_t = 0$ (ce qui ne sert à rien), mais aussi que $\mu_1 = \cdots = \mu_b = 0$, ce qui faisait partie de notre but. Dès lors tout finit très vite : on déduit de $\mu_1 = \cdots = \mu_b = 0$ que $h = 0$, donc que

$$\nu_1 g_1 + \cdots + \nu_t g_t + \lambda_1 e_1 + \cdots + \lambda_a e_a = 0$$

et, en utilisant cette fois la liberté de $(g_1, \ldots, g_t, e_1, \ldots, e_a)$ dans E que $\nu_1 = \cdots = \nu_t = \lambda_1 = \cdots = \lambda_a = 0$.
La liberté est prouvée.

On a alors fini, reste à s'en apercevoir ! La base de $E + F$ qu'on vient de calculer est formée de $t + a + b$ vecteurs. On en conclut que :

$$\dim(E + F) = t + a + b = t + (m - t) + (n - t) = m + n - t = \dim E + \dim F - \dim(E \cap F).$$

•

7 - Sommes directes

Si on peut tenter de justifier par analogie la notion qui va être définie, la somme directe est à la somme ce que la base est au système générateur.

Définition 4-7-53 : Soit $n \geq 0$ fixé, $k \geq 0$ fixé, et $E_1, \dots, E_k$ des sous-espaces de $\mathbf{R}^n$. En notant leur sommme $F = E_1 + \cdots + E_k$, on dira que les sous-espaces $E_1, \dots, E_k$ sont en **somme directe** lorsque :

pour tout $v \in F$, il existe un k-uplet **unique** de vecteurs $v_1, \dots, v_k$ avec $v_1 \in E_1, \dots, v_k \in E_k$ tel que

$$v = v_1 + \cdots + v_k.$$

Notation 4-7-33 : Quand F est somme directe de $E_1, \dots, E_k$, on note $F = E_1 \oplus \cdots \oplus E_k$.

Théorème 4-7-9 : Soit $n \geq 0$ fixé, $k \geq 0$ fixé, et $E_1, \dots, E_k$ des sous-espaces de $\mathbf{R}^n$. Les sous-espaces $E_1, \dots, E_k$ sont en somme directe si et seulement si

$$\dim(E_1 + \cdots + E_k) = \dim E_1 + \cdots + \dim E_k.$$

Démonstration :

On va montrer successivement les deux implications.

Pour cela, dans un sens comme dans l'autre, on notera $d_1, \dots, d_k$ les dimensions respectives de $E_1, \dots, E_k$, et on considèrera des bases respectives $(e_1^{(1)}, \dots, e_{d_1}^{(1)}), \dots, (e_1^{(k)}, \dots, e_{d_k}^{(k)})$ des sous-espaces $E_1, \dots, E_k$.

On va s'intéresser au système $(e_1^{(1)}, \dots, e_{d_1}^{(1)}, e_1^{(2)}, \dots, e_{d_2}^{(2)}, \dots, e_1^{(k)}, \dots, e_{d_k}^{(k)})$ (obtenu en réunissant toutes les bases à notre disposition) . Il est clair (?) que ce système engendre F (en tous cas, si ce n'est pas clair, je n'ai pas envie de l'écrire...).

* Supposons $E_1, \dots, E_k$ en somme directe, et montrons l'identité entre les dimensions.

On va montrer que le gros système $(e_1^{(1)}, \dots, e_{d_1}^{(1)}, e_1^{(2)}, \dots, e_{d_2}^{(2)}, \dots, e_1^{(k)}, \dots, e_{d_k}^{(k)})$ est une base de F. On en déduira aussitôt que la dimension de F est bien égale à l'entier $d_1 + \cdots + d_k$.

On sait déjà qu'il est générateur, montrons qu'il est libre. Soit $(\lambda_1^{(1)}, \dots, \lambda_{d_1}^{(1)}, \lambda_1^{(2)}, \dots, \lambda_{d_2}^{(2)}, \dots, \lambda_1^{(k)}, \dots, \lambda_{d_k}^{(k)})$ des scalaires tels que

$$\lambda_1^{(1)} e_1^{(1)} + \cdots + \lambda_{d_1}^{(1)} e_{d_1}^{(1)} + \lambda_1^{(2)} e_1^{(2)} + \cdots + \lambda_{d_2}^{(2)} e_{d_2}^{(2)} + \cdots + \lambda_1^{(k)} e_1^{(k)} + \cdots + \lambda_{d_k}^{(k)} e_{d_k}^{(k)} = 0.$$

Notons $v_1 = \lambda_1^{(1)} e_1^{(1)} + \cdots + \lambda_{d_1}^{(1)} e_{d_1}^{(1)}, \dots, v_k = \lambda_1^{(k)} e_1^{(k)} + \cdots + \lambda_{d_k}^{(k)} e_{d_k}^{(k)}$. On a alors deux écritures du vecteur nul (vecteur de F) comme somme de vecteurs des $E_1, \dots, E_k$:

$$\begin{aligned} 0 &= 0 + 0 + \cdots + 0 \\ &= v_1 + v_2 + \cdots + v_k \end{aligned}$$

Par l'hypothèse selon laquelle F est somme directe, l'écriture de 0 à partir de $E_1, \dots, E_k$ est unique. On en déduit que $v_1 = \cdots = v_k = 0$.

Maintenant, en utilisant la liberté des bases de chacun des espaces sommés, on en déduit ensuite finalement que :

$$\lambda_1^{(1)} = \cdots = \lambda_{d_1}^{(1)} = \lambda_1^{(2)} = \cdots = \lambda_{d_2}^{(2)} = \cdots = \lambda_1^{(k)} = \cdots = \lambda_{d_k}^{(k)} = 0.$$

La dimension de F est bien la dimension annoncée.

* Réciproquement, supposons l'identité entre les dimensions, et montrons que $E_1, \dots, E_k$ sont en somme directe.

Cette fois le gros système $(e_1^{(1)}, \dots, e_{d_1}^{(1)}, e_1^{(2)}, \dots, e_{d_2}^{(2)}, \dots, e_1^{(k)}, \dots, e_{d_k}^{(k)})$ est formé de $\dim F$ vecteurs, et on sait déjà qu'il engendre F. C'est donc une base de F.

Soit maintenant v un vecteur de F. Considérons deux écritures

$$\begin{aligned} v &= v_1 + \cdots + v_k \\ &= w_1 + \cdots + w_k \end{aligned}$$

de v, dans lequelles $v_1 \in E_1, \dots, v_k \in E_k$, $w_1 \in E_1, \dots, w_k \in E_k$.

On peut alors décomposer chacun des vecteurs v_i ou w_i dans la base à notre disposition du E_i correspondant, soit $v_i = \alpha_1^{(i)} e_1^{(i)} + \cdots + \alpha_{d_i}^{(i)} e_{d_i}^{(i)}$, $w_i = \beta_1^{(i)} e_1^{(i)} + \cdots + \beta_{d_i}^{(i)} e_{d_i}^{(i)}$.

Reportons ces écritures dans les deux écritures du vecteur v.

On obtient :

$$\begin{aligned} v &= \alpha_1 e_1^{(1)} + \cdots + \alpha_{d_1} e_{d_1}^{(1)} + \alpha_1 e_1^{(2)} + \cdots + \alpha_{d_2} e_{d_2}^{(2)} + \cdots + \alpha_1 e_1^{(k)} + \cdots + + \alpha_{d_k} e_{d_k}^{(k)} \\ &= \beta_1 e_1^{(1)} + \cdots + \beta_{d_1} e_{d_1}^{(1)} + \beta_1 e_1^{(2)} + \cdots + \beta_{d_2} e_{d_2}^{(2)} + \cdots + \beta_1 e_1^{(k)} + \cdots + + \beta_{d_k} e_{d_k}^{(k)}. \end{aligned}$$

Le “gros” système étant une base, on en déduit l'égalité des α et des β (unicité des coordonnées de v dans ce gros système), donc celle de chaque v_i au w_i correspondant.
L'écriture de v était bien unique.

•

On en déduit aussitôt, dans le cas particulier de somme directe de deux sous-espaces seulement la :
Proposition 4-7-22 : Soit $n \geq 0$ fixé, E_1 et E_2 des sous-espaces de $\mathbf{R}^n$. Les sous-espaces E_1 et E_2 sont en somme directe si et seulement si $E_1 \cap E_2 = \{0\}$.
Démonstration : Par le théorème précédent, E_1 et E_2 sont en somme directe si et seulement si on a : $\dim(E_1 + E_2) = \dim E_1 + \dim E_2$. En lisant la formule de Grassmann, on s'aperçoit que ceci équivaut exactement à $\dim(E_1 \cap E_2) = 0$, donc à $E_1 \cap E_2 = \{0\}$.

•

Attention ! D'expérience, trop d'étudiants oublient que cette proposition ne concerne que la somme directe de **deux** sous-espaces.Une généralisation à plus de deux sous-espaces existe, mais est d'énoncé malcommode, et d'usage encore plus malcommode (et je me garde bien de la donner). Considérez donc, c'est bien plus sûr, qu'il n'y a pas d'énoncé analogue pour plus de deux espaces —et gardez-vous bien d'utiliser tout énoncé fantaisiste de votre invention !

Chapitre 5 - Limites

1 - Opérations sur les fonctions

D'expérience un enseignant s'aperçoit souvent qu'il a négligé de préciser ce qu'était la somme ou le produit de deux fonctions... Je ne suis pas loin de l'avoir encore oublié, mais je reviens en arrière pour ajouter cette section.

Définition 5-1-54 : Soit f et g deux fonctions réelles d'une variable réelle, définies sur un même ensemble $\mathcal{D}$. La **somme** des fonctions f et g est la fonction $f+g$ définie sur le même ensemble $\mathcal{D}$ par : pour tout $t \in \mathcal{D}$, $(f+g)(t) = f(t) + g(t)$. Et de même on définirait la **différence**, le **produit**, le **quotient** (sous l'hypothèse supplémentaire : g ne s'annule pas sur $\mathcal{D}$)... en espérant ne pas en avoir oublié. (Je constate en tapant la suite que j'avais oublié de signaler que $f \leq g$ signifie : pour tout $t \in \mathcal{D}$, $f(t) \leq g(t)$, omission réparée...)

2 - Point adhérent à une partie de R

En préalable aux définitions, un peu de vocabulaire utile pour cerner précisément dans quelles circonstances il est légitime chercher à calculer une limite (ainsi, cela a un sens de se demander quelle est la limite de $t \ln t$ quand t tend vers 0, mais il est stupide de se demander quelle est la limite de $t \ln t$ quand t tend vers -1, qui est "trop loin" du domaine de définition du logarithme).

Ce sera une définition sèche —ni commentaires, ni énoncés à démontrer.

Définition 5-2-55 : Soit $\mathcal{D}$ une partie de $\mathbf{R}$ et a un réel. On dit que a est **adhérent** à $\mathcal{D}$ lorsque pour tout réel $\eta > 0$, l'intervalle $[a-\eta, a+\eta]$ rencontre $\mathcal{D}$; en d'autres termes lorsque pour tout $\eta > 0$, il existe un $t \in \mathcal{D}$ tel que $|t-a| \leq \eta$.

Pour pouvoir donner des définitions les plus analogues possibles concernant les limites quand t tend vers $+\infty$ ou quand t tend vers $-\infty$, introduisons un concept analogue à celui de "point adhérent" mais relatif à l'infini :

Définition 5-2-56 : Soit $\mathcal{D}$ une partie de $\mathbf{R}$, on dit que $\mathcal{D}$ est **non majorée** lorsque pour tout réel A, l'intervalle $[A, +\infty[$ rencontre $\mathcal{D}$; en d'autres termes lorsque pour tout réel A, il existe un $t \in \mathcal{D}$ tel que $A \leq t$. De même, on dit que $\mathcal{D}$ est **non minorée** lorsque pour tout réel A, l'intervalle $]-\infty, A]$ rencontre $\mathcal{D}$; en d'autres termes lorsque pour tout réel A, il existe un $t \in \mathcal{D}$ tel que $t \leq A$.

3 - Définition des limites

Définition 5-3-57 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$, et soit a un réel adhérent à $\mathcal{D}_f$; soit l un réel. On dit que $f(t)$ **tend vers** l quand t tend vers a lorsque :

$$\text{pour tout } \epsilon > 0, \text{ il existe } \eta > 0 \text{ tel que pour tout } t \in \mathcal{D}_f, \big((|t-a| \leq \eta) \Rightarrow (|f(t) - l| \leq \epsilon)\big).$$

On pourrait d'ailleurs donner cette définition même pour des a non adhérents à $\mathcal{D}_f$, mais elle serait stupide ($f(t)$ tendrait vers n'importe quoi quand t tend vers a).

On recommence pour les limites en $+\infty$ et en $-\infty$.

Définition 5-3-58 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$; on suppose que $\mathcal{D}_f$ n'est pas majoré ; soit l un réel. On dit que $f(t)$ **tend vers** l quand t tend vers $+\infty$ lorsque :

$$\text{pour tout } \epsilon > 0, \text{ il existe un réel } A \text{ tel que pour tout } t \in \mathcal{D}_f, \big((A \leq t) \Rightarrow (|f(t) - l| \leq \epsilon)\big).$$

De même, sous l'hypothèse $\mathcal{D}_f$ non minoré, on dira que $f(t)$ **tend vers** l quand t tend vers $-\infty$ lorsque :

$$\text{pour tout } \epsilon > 0, \text{ il existe un réel } A \text{ tel que pour tout } t \in \mathcal{D}_f, \big((t \leq A) \Rightarrow (|f(t) - l| \leq \epsilon)\big).$$

Il nous reste à avaler ce que signifie "tendre vers $+\infty$" (ou $-\infty$) pour une fonction f.

Définition 5-3-59 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$, et soit a un réel adhérent à $\mathcal{D}_f$. On dit que $f(t)$ **tend vers** $+\infty$ quand t tend vers a lorsque :

$$\text{pour tout réel B, il existe } \eta > 0 \text{ tel que pour tout } t \in \mathcal{D}_f, \big((|t-a| \leq \eta) \Rightarrow (B \leq f(t))\big).$$

De même on dit que $f(t)$ **tend vers** $-\infty$ quand t tend vers a lorsque :

$$\text{pour tout réel B, il existe } \eta > 0 \text{ tel que pour tout } t \in \mathcal{D}_f, \big((|t-a| \leq \eta) \Rightarrow (f(t) \leq B)\big).$$

Il faudrait maintenant que je définisse ce que signifie "$f(t)$ tend vers $+\infty$ quand t tend vers $+\infty$" et ainsi de suite... C'est là que j'abandonne la souris et les copier-coller physiques pour tenter de vous inviter à pratiquer le copier-coller mental : ces définitions s'obtiennent en mélangeant avec intelligence les morceaux déjà écrits, et en pratique vous devriez y arriver.

Ayant pris la précaution d'être ainsi paresseux, je peux désormais sans risque de m'écrouler donner encore d'autres définitions... étant entendu qu'elles doivent pouvoir se recoller les unes aux autres, ou être modifiées quand des $-$ remplacent les $+$...

Définition 5-3-60 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$, et soit a un réel adhérent à $\mathcal{D}_f \setminus \{a\}$; soit l un réel. On dit que $f(t)$ **tend vers** l quand t tend vers a, $t \neq a$ lorsque la restriction de f à $\mathcal{D}_f \setminus \{a\}$ tend vers l quand t tend vers a.

Définition 5-3-61 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$, et soit a un réel adhérent à $\mathcal{D}_f \cap]a, +\infty[$; soit l un réel. On dit que $f(t)$ **tend vers** l quand t tend vers a **à droite** (ou "quand t tend vers a, $a < t$") lorsque la restriction de f à $\mathcal{D}_f \cap]a, +\infty[$ tend vers l quand t tend vers a. Plus explicitement, $f(t)$ tend vers l quand t tend vers a, $a < t$ lorsque :

$$\text{pour tout } \epsilon > 0, \text{ il existe } \eta > 0 \text{ tel que pour tout } t \in \mathcal{D}_f, \big((a < t \leq a + \eta) \Rightarrow (|f(t) - l| \leq \epsilon)\big).$$

4 - Opérations sur les limites finies

Cette section devra être complétée par une deuxième ajoutant des résultats sur les limites éventuellement infinies... Je tente de donner toutes les démonstrations ici, car dans la suite j'en serai certainement las.

Premières pièces du puzzle (limites de l'identité et des constantes). De vraies évidences !

Proposition 5-4-23 : Soit a un réel, notons c la fonction constante prenant la valeur constante également notée c. Alors :

$$t \text{ tend vers } a \text{ quand } t \text{ tend vers } a$$

$$c \text{ tend vers } c \text{ quand } t \text{ tend vers } a.$$

Démonstration : Soit $\epsilon > 0$ fixé. Prenons $\eta = \epsilon$, qui est bien un réel strictement positif. Alors quand $|t-a| \leq \eta$, $|t-a| \leq \epsilon = \eta$, ce qui prouve la première affirmation, et quand $|t-a| \leq \eta$, $|c-c| = 0 \leq \epsilon$, ce qui prouve la deuxième affirmation.

•

En anticipant d'un chapitre, on vient juste de montrer que l'identité d'une part, les constantes d'autre part, sont des fonctions continues sur $\mathbf{R}$.

L'addition

Proposition 5-4-24 : Soit f_1 et f_2 deux fonctions réelles d'une variable réelle, définies sur un même ensemble $\mathcal{D}$, et soit a un réel adhérent à $\mathcal{D}$.

On suppose que $f_1(t)$ tend vers l_1 et $f_2(t)$ tend vers l_2 quand t tend vers a. Alors $(f_1+f_2)(t)$ tend vers l_1+l_2 quand t tend vers a.

Démonstration : Soit $\epsilon > 0$ fixé ; appliquons la définition de "tendre vers" à f_1 et à $\frac{\epsilon}{2}$: il existe donc un η_1 tel que pour $t \in \mathcal{D}$, si $|t-a| \leq \eta_1$, alors $|f_1(t) - l_1| \leq \frac{\epsilon}{2}$. De même il existe un η_2 tel que pour $t \in \mathcal{D}$, si $|t-a| \leq \eta_2$, alors $|f_2(t) - l_2| \leq \frac{\epsilon}{2}$.

Soit η le plus petit des deux réels strictement positifs η_1 et η_2. Dès qu'un $t \in \mathcal{D}$ vérifie $|t-a| \leq \eta$, il vérifie donc à la fois $|t-a| \leq \eta_1$ et $|t-a| \leq \eta_2$, donc à la fois $|f_1(t) - l_1| \leq \frac{\epsilon}{2}$ et $|f_2(t) - l_2| \leq \frac{\epsilon}{2}$.

On en déduit alors que

$$|(f_1+f_2)(t) - (l_1+l_2)| = |(f_1(t) - l_1) + (f_2(t) - l_2)| \leq |f_1(t) - l_1| + |f_2(t) - l_2| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

•

La multiplication
La proposition ressemble à la précédente, la preuve est tout de même significativement plus compliquée...
Proposition 5-4-25 : Soit f_1 et f_2 deux fonctions réelles d'une variable réelle, définies sur un même ensemble $\mathcal{D}$, et soit a un réel adhérent à $\mathcal{D}$.
On suppose que $f_1(t)$ tend vers l_1 et $f_2(t)$ tend vers l_2 quand t tend vers a. Alors $(f_1f_2)(t)$ tend vers l_1l_2 quand t tend vers a.
Démonstration : Soit ϵ fixé ; on va appliquer la définition de limite à f_1 et à f_2, mais pour des réels significativement plus bizarres que les tout simples $\frac{\epsilon}{2}$ de la preuve précédente.

Évidemment, la lecture de la preuve dans l'ordre logique ne permet pas de comprendre tout de suite pourquoi on a pêché ces bizarres réels : ils sont en fait sélectionnés pour que tout s'arrange à la fin ; dans une façon "dynamique" de rédiger la preuve, il serait agréable de les laisser en blanc puis de combler les blancs en arrivant en bas de la preuve —c'est ce que je fais lorsque je l'écris au tableau, c'est irréalisable hélas dans une version papier...

Parachutons donc l'introduction des réels strictement positifs $\epsilon_1 = \frac{\epsilon}{2(|l_2|+1)}$ puis $\epsilon_2 = \frac{\epsilon}{2(|l_1|+\epsilon_1)}$.

Appliquons la définition de "tendre vers" d'une part à f_1 et ϵ_1 et d'autre part à f_2 et ϵ_2 : on produit ainsi deux réels strictement positifs η_1 et η_2 tels que pour $t \in \mathcal{D}$, si $|t-a| \leq \eta_1$, alors $|f_1(t) - l_1| \leq \epsilon_1$ et pour $t \in \mathcal{D}$, si $|t-a| \leq \eta_2$, alors $|f_2(t) - l_2| \leq \epsilon_2$.

Comme dans la preuve précédente, soit η le plus petit des deux réels strictement positifs η_1 et η_2. Dès qu'un $t \in \mathcal{D}$ vérifie $|t-a| \leq \eta$, il vérifie donc à la fois $|f_1(t) - l_1| \leq \epsilon_1$ et $|f_2(t) - l_2| \leq \epsilon_2$.

On en déduit alors que

$$\begin{aligned}
|(f_1f_2)(t) - l_1l_2| &= |f_1(t)f_2(t) - f_1(t)l_2 + f_1(t)l_2 - l_1l_2| \\
&\leq |f_1(t)f_2(t) - f_1(t)l_2| + |f_1(t)l_2 - l_1l_2| \\
&= |f_1(t)||f_2(t) - l_2| + |f_1(t) - l_1||l_2| \\
&\leq |f_1(t)|\epsilon_2 + \epsilon_1|l_2| \\
&= |f_1(t)|\frac{\epsilon}{2(|l_1|+\epsilon_1)} + \frac{\epsilon}{2(|l_2|+1)}|l_2| \\
&\leq \epsilon\Big(\frac{|f_1(t)|}{2(|l_1|+\epsilon_1)} + \frac{1}{2}\Big) \\
&= \epsilon\Big(\frac{|(f_1(t)-l_1)+l_1|}{2(|l_1|+\epsilon_1)} + \frac{1}{2}\Big) \\
&\leq \epsilon\Big(\frac{|f_1(t)-l_1|+|l_1|}{2(|l_1|+\epsilon_1)} + \frac{1}{2}\Big) \\
&\leq \epsilon\Big(\frac{\epsilon_1+|l_1|}{2(|l_1|+\epsilon_1)} + \frac{1}{2}\Big) \\
&= \epsilon.
\end{aligned}$$

•

Nous savons donc faire des soustractions : $f-g$ est obtenu par addition de f et du produit de g par la fonction constante -1.

Le passage à l'inverse
Proposition 5-4-26 : Soit a un point de $\mathbf{R}^*$. Alors $1/t$ tend vers $1/a$ quand t tend vers a.

Démonstration : Soit $\epsilon > 0$ fixé. Soit η le plus petit des deux réels strictement positifs $\dfrac{|a|^2\epsilon}{2}$ et $\dfrac{|a|}{2}$. Soit alors un $t \in \mathbf{R}^*$ tel que $|t-a| \leq \eta$.

Tentons de majorer $|\dfrac{1}{t} - \dfrac{1}{a}| = \dfrac{|a-t|}{|t||a|}$.

Pour ce faire, il va falloir majorer le numérateur et minorer le dénominateur.

Le numérateur est facilement majoré par $|a-t| \leq \eta \leq \dfrac{|a|^2\epsilon}{2}$.

Le dénominateur est à peine moins aisément minoré par

$$|t| = |(t-a)+a| \geq |a| - |t-a| \geq |a| - \eta \geq |a| - \frac{|a|}{2} = \frac{|a|}{2}, \text{ donc } |t||a| \geq \frac{|a|^2}{2}.$$

Donc $|\dfrac{1}{t} - \dfrac{1}{a}| = \dfrac{|a-t|}{|t||a|} \leq \dfrac{|a|^2\epsilon/2}{|a|^2/2} = \epsilon$.

•

En anticipant encore de quelques pages, on vient de montrer que la fonction t donne $1/t$ est continue sur $\mathbf{R}^*$.

La composition

Proposition 5-4-27 : Soit f et g deux fonctions d'une variable réelle, définies respectivement sur les ensembles $\mathcal{D}_f$ et $\mathcal{D}_g$. Soit a et b deux réels, respectivement adhérents à $\mathcal{D}_f$ et à $\mathcal{D}_g$, et l un autre réel. On suppose que la composée $g \circ f$ existe, que $f(t)$ tend vers b quand t tend vers a et que $g(u)$ tend vers l quand u tend vers b.

Alors $(g \circ f)(t)$ tend vers l quand t tend vers a.

Démonstration :

Soit un $\epsilon > 0$ fixé. En appliquant la définition de "tend vers" à g et ϵ, on produit un réel $\eta_1 > 0$ tel que pour tout $u \in \mathcal{D}_g$ vérifiant $|u-b| \leq \eta_1$, on ait $|g(u)-l| \leq \epsilon$. Recommençons en appliquant cette fois la définition de "tend vers" à f et η_1 : on obtient un nouveau réel $\eta > 0$ tel que pour tout $t \in \mathcal{D}_f$ vérifiant $|t-a| \leq \eta$, on ait $|f(t)-b| \leq \eta_1$. Mais alors, en posant $u = f(t)$, on a $|u-b| \leq \eta_1$ et donc $|g(u)-l| \leq \epsilon$, c'est-à-dire $|(g \circ f)(t) - l| \leq \epsilon$.

•

Remarque : en bonne rigueur ce n'est pas "$g \circ f$" qui est étudié en général, puisque f a pour ensemble d'arrivée $\mathbf{R}$ et g a un ensemble de départ plus petit. Il fallait bien sûr comprendre que la condition exprimée sous forme volontairement un peu imprécise ("$g \circ f$ existe") devait être interprétée comme signifiant : $f(\mathcal{D}_f) \subset \mathcal{D}_g$, et cela a donc un sens de restreindre l'ensemble d'arrivée de f à $\mathcal{D}_g$, produisant ainsi une nouvelle application abusivement notée f et telle que $g \circ f$ existe.

En appliquant ce résultat à $g(t) = 1/t$, on en déduit que si f tend vers l non nulle (et f ne s'annule pas), $1/f$ tend vers $1/l$, puis en utilisant la multiplication, que la limite d'un quotient est le quotient des limites.

Avec tous ces outils, on sait déjà calculer pas mal de limites... d'autres outils apparaîtront au second semestre, il me semble toutefois utile de donner dès à présent

Le principe des gendarmes

Proposition 5-4-28 : Soit f_1, f_2 et f_3 trois fonctions réelles d'une variable réelle, définies sur un même ensemble $\mathcal{D}$, et soit a un réel adhérent à $\mathcal{D}$.

On suppose que pour tout $t \in \mathcal{D}$,

$$f_1(t) \leq f_2(t) \leq f_3(t),$$

que $f_1(t)$ tend vers l et $f_3(t)$ tend vers l (la même !) quand t tend vers a. Alors $f_2(t)$ tend aussi vers l quand t tend vers a.

Démonstration : Soit $\epsilon > 0$ fixé. En appliquant la définition de "tend vers" à f_1 et à f_3, on produit deux réels strictement positifs η_1 et η_3 tels que pour $t \in \mathcal{D}$, si $|t-a| \leq \eta_1$, on ait $|f_1(t)-l| \leq \epsilon$, et que si $|t-a| \leq \eta_3$, on ait $|f_3(t)-l| \leq \epsilon$. Soit alors η le plus petit des deux réels η_1 et η_3. Alors pour $|t-a| \leq \eta$, comme on a simultanément $|t-a| \leq \eta_1$ et $|t-a| \leq \eta_3$, on a donc à la fois $|f_1(t)-l| \leq \epsilon$ et $|f_3(t)-l| \leq \epsilon$, donc —en perdant volontairement la moitié de l'information— on a à la fois $l - f_1(t) \leq \epsilon$ et $f_3(t) - l \leq \epsilon$, soit $l - \epsilon \leq f_1(t)$ et $f_3(t) \leq l + \epsilon$. On en déduit que

$$l - \epsilon \leq f_1(t) \leq f_2(t) \leq f_3(t) \leq l + \epsilon$$

et donc $|f_2(t) - l| \leq \epsilon$.

•

Restrictions

Les résultats qui suivent ne sont généralement pas explicités, pourtant, ils servent franchement très souvent... implicitement. Pourquoi ne pas les donner donc ? À vous de juger s'ils méritent d'être retenus.

Proposition 5-4-29 : Soit f une fonction d'une variable réelle, définie sur un ensemble $\mathcal{D}_f$. Soit a un point adhérent à $\mathcal{D}_f$.

a) Soit $\mathcal{D}$ une partie de $\mathcal{D}_f$, à laquelle a est adhérent. Si $f(t)$ tend vers l quand t tend vers a, $f_{|\mathcal{D}}(t)$ tend aussi vers l quand t tend vers a.

b) Soit $\mathcal{D}_1$ et $\mathcal{D}_2$ deux parties de $\mathcal{D}_f$ telles que $\mathcal{D}_f = \mathcal{D}_1 \cup \mathcal{D}_2$; on suppose a adhérent tant à $\mathcal{D}_1$ qu'à $\mathcal{D}_2$. Si $f_{|\mathcal{D}_1}(t)$ et $f_{|\mathcal{D}_2}(t)$ tendent tous deux vers l quand t tend vers a, $f(t)$ tend aussi vers l quand t tend vers a.

c) Soit I un **intervalle ouvert** contenant a. Alors a est adhérent à $I \cap \mathcal{D}_f$, et si $f_{I\cap\mathcal{D}_f}(t)$ tend vers l quand t tend vers a, alors $f(t)$ tend aussi vers l quand t tend vers a.

Démonstration :

a) C'est le plus facile : pour $\epsilon > 0$ fixé, le $\eta > 0$ qui a servi pour f peut servir de nouveau tel quel pour $f_{|\mathcal{D}}$.

b) Ce n'est pas bien méchant. Soit $\epsilon > 0$ fixé ; la définition de "tend vers" appliquée à $f_{|\mathcal{D}_1}$ fournit un $\eta_1 > 0$, l'application à $f_{|\mathcal{D}_2}$ fournit un $\eta_2 > 0$. Le plus petit des deux réels η_1 et η_2 convient alors pour f.

c) C'est le morceau le plus sérieux, car il faudra comprendre comment utiliser l'hypothèse selon laquelle I est un intervalle ouvert

Comme I est un intervalle ouvert, il existe "évidemment" un réel strictement positif δ tel que $[a-\delta, a+\delta] \subset I$. Si on n'y croit pas, on lira le paragraphe suivant, si on y croit on le sautera...

Notons $I =]\alpha_-, \alpha_+[$, où $\alpha_- < a < \alpha_+$ (et où α_- peut être le symbole $-\infty$, α_+ peut être le symbole $+\infty$). On définira δ de la façon suivante : si l'intervalle I est un segment, δ sera le plus petit des deux réels strictement positifs $\dfrac{a-\alpha_-}{2}$ et $\dfrac{\alpha_+ - a}{2}$; si I est une demi-droite s'étendant vers la gauche, on prendra $\delta = \dfrac{\alpha_+ - a}{2}$; si I est une demi-droite s'étendant vers la droite, on prendra $\delta = \dfrac{a - \alpha_-}{2}$; enfin si $I = \mathbf{R}$ (cas sans intérêt, mais qu'il faut bien traiter aussi pour être complet) on prendra $\delta = 1$.

Vérifions tout d'abord que a est bien adhérent à $I \cap \mathcal{D}_f$. Soit un $\eta_1 > 0$ fixé. Notons η le plus petit des deux réels strictement positifs η_1 et δ. En appliquant la définition de "point adhérent" à $\mathcal{D}_f$ et à η il existe un $t \in [a-\eta, a+\eta] \cap \mathcal{D}_f$. Ainsi $t \in \mathcal{D}_f$ et $t \in [a-\eta, a+\eta]$; comme $\eta \leq \delta$ et $\eta \leq \eta_1$ on obtient $t \in [a-\delta, a+\delta] \subset I$ et $t \in [a-\eta_1, a+\eta_1]$. Dès lors $t \in [a-\eta_1, a+\eta_1] \cap (I \cap \mathcal{D}_f)$ qui' n'est donc pas vide.

L'argument justifiant l'énoncé sur les limites est le même : fixons un $\epsilon > 0$ et soit $\eta_1 > 0$ obtenu en appliquant la définition de "tend vers" à $f_{I\cap\mathcal{D}_f}$, notons de nouveau η le plus petit des deux réels strictement positifs η_1 et δ. Alors dès que $|t-a| \leq \eta$, avec $t \in \mathcal{D}_f$, on obtient $|t-a| \leq \delta$, donc $t \in I$ et donc $t \in I \cap \mathcal{D}_f$ et $|t-a| \leq \eta_1$. On en déduit que $|f_{|I\cap\mathcal{D}_f}(t) - l| \leq \epsilon$, soit $|f(t) - l| \leq \epsilon$.

•

Remarques : Le b) sert un peu tout le temps. On en déduit par exemple que si f admet une limite à droite et une limite à gauche égales quand t tend vers a, elle admet une limite quand t tend vers a ($t \neq a$) ; ou —dans la version analogue pour $a = +\infty$— que si une fonction d'une variable entière n admet une limite quand n tend vers $+\infty$, n pair, et la même limite quand n tend vers $+\infty$, n impair, elle admet une limite quand n tend vers $+\infty$.

Le c) sera par exemple utilisé pour calculer la dérivée d'une fonction à partir d'informations portant sur la seule restriction de cette fonction à un intervalle **ouvert**.

5 - Opérations sur les limites éventuellement infinies

Je triche ! Je triche ! Écrire cette section m'ennuyant trop, je la saute...

6 - Limites et inégalités

Ce sera l'unique occasion d'utiliser la définition du concept de "point adhérent"...

Proposition 5-6-30 : Soit f_1 et f_2 deux fonctions réelles d'une variable réelle, définies sur un même ensemble $\mathcal{D}$, et soit a un réel adhérent à $\mathcal{D}$.
On suppose que $f_1(t)$ tend vers l_1 et $f_2(t)$ tend vers l_2 quand t tend vers a.
On suppose enfin que pour tout $t \in \mathcal{D}$, $f_1(t) \leq f_2(t)$.
Alors $l_1 \leq l_2$.
Démonstration : Soit un $\epsilon > 0$. Appliquons la définition de "tend vers" à f_1 et à f_2 et pour $\frac{\epsilon}{3}$: elle fournit un η_1 et un η_2 tels que pour $t \in \mathcal{D}$, si $|t-a| \leq \eta_1$, on ait $|f_1(t) - l_1| \leq \frac{\epsilon}{3}$ et que si $|t-a| \leq \eta_2$, on ait $|f_2(t) - l_2| \leq \frac{\epsilon}{3}$. Prenons η le plus petit des deux réels strictement positifs η_1 et η_2 ; si $|t-a| \leq \eta$, on a $|f_1(t) - l_1| \leq \frac{\epsilon}{3}$ et donc, en perdant volontairement de l'information, $l_1 - f_1(t) \leq \frac{\epsilon}{3}$, soit $l_1 - \frac{\epsilon}{3} \leq f_1(t)$. En attaquant maintenant du côté de f_2, on obtient $|f_2(t) - l_1| \leq \frac{\epsilon}{3}$, puis par perte volontaire d'information $f_2(t) - l_2 \leq \frac{\epsilon}{3}$, soit $f_2(t) \leq l_2 + \frac{\epsilon}{3}$. On a ainsi, pour $t \in \mathcal{D}$ vérifiant $|t-a| \leq \eta$:

$$l_1 - \frac{\epsilon}{3} \leq f_1(t) \leq f_2(t) \leq l_2 + \frac{\epsilon}{3}.$$

On croit en déduire $l_1 - \frac{\epsilon}{3} \leq l_2 + \frac{\epsilon}{3}$ et ce n'est pas faux ; il faut toutefois souligner qu'en ce point précis on se sert de l'existence d'un t au moins vérifiant les conditions initiales, et que c'est parce que a est supposé adhérent à $\mathcal{D}$ qu'un tel t existe. On obtient donc $l_1 - l_2 \leq 2\frac{\epsilon}{3} < \epsilon$.

Ceci étant vrai pour tout $\epsilon > 0$, on finit par conclure que $l_1 - l_2 \leq 0$, soit $l_1 \leq l_2$.

•

7 - Le mot limite

Hé oui, j'ai attendu le dernier moment avant de quitter le chapitre... Mais il est tout de même temps d'énoncer la toute simple
Proposition 5-7-31 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$, et soit a un réel adhérent à $\mathcal{D}_f$. Alors $f(t)$ tend vers au plus un réel quand t tend vers a.
Démonstration : Supposons que $f(t)$ tende simultanément vers l_1 et vers l_2 quand t tend vers a. Appliquons la proposition précédente à $f_1 = f_2 = f$: comme $f_1 \leq f_2$, $l_1 \leq l_2$. Et réciproquement en utilisant $f_2 \leq f_1$. •

Cet énoncé permet de donner la
Définition 5-7-62 : Lorsque $f(t)$ tend vers l quand t tend vers a, on dit que l est la **limite** de f en a.
Notation 5-7-34 : Cette limite est notée $\lim_{t \to a} f(t)$, ou plus légèrement $\lim_a f$.

Évidemment, après ce bel effort, il faudrait recommencer avec les cas où a est infini, puis avec l'éventualité de limites valant $+\infty$ ou $-\infty$. Je m'octroie une dispense.
Remarque : La notation lim, quoique si usuelle que je ne puisse me l'interdire, me paraît dangereuse : elle ne doit pas faire perdre de vue qu'une limite **peut ne pas exister** ! Il est si tentant de travailler sur $\lim_a f$ sans démontrer préalablement son existence... La notation avec des flèches me paraît d'expérience mériter pour cette raison d'être vivement recommandée.

8 - Un exemple à méditer

Voici sans doute l'exemple le plus visuel de fonction n'admettant aucune limite :
Exemple : $\cos t$ n'admet pas de limite quand t tend vers $+\infty$.
Démonstration : Supposons que $l = \lim_{t \to +\infty} \cos t$ existe.

Appliquons la définition de "limite" à $\epsilon = \frac{1}{2}$: il existe donc un A réel tel que pour tout $t \geq A$, on ait :

$$|\cos t - l| \leq \frac{1}{2} \qquad (*).$$

Prenons un entier $k \geq 0$ tel que $k \geq \frac{A}{2\pi}$, ou en d'autres termes $2k\pi \geq A$ — et *a fortiori* $(2k+1)\pi \geq A$.

Appliquons l'inégalité $(*)$ à $t = 2k\pi$: on obtient $|1 - l| \leq \dfrac{1}{2}$; appliquons la à $t = (2k+1)\pi$, on obtient $|-1 - l| \leq \dfrac{1}{2}$.

On en déduit alors que :

$$|1 - (-1)| = |(1 - l) + (l - (-1))| \leq |1 - l| + |l - (-1)| \leq \frac{1}{2} + \frac{1}{2}$$

c'est-à-dire $2 \leq 1$. C'est absurde ! •

Chapitre 6 - Fonctions continues

Juste quelques mots ; le sujet sera essentiellement traité au second semestre, mais un théorème m'étant indispensable pour continuer, il me faut le citer dès maintenant.

1 - La définition

Définition 6-1-63 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$. Soit a un point de $\mathcal{D}_f$, supposé adhérent à $\mathcal{D}_f \setminus \{a\}$. On dit que f est **continue en** a lorsque $f(t)$ tend vers $f(a)$ quand t tend vers a ($t \neq a$).
Remarque : La définition usuellement trouvée dans les livres (la bonne, celle que vous reverrez en licence) ne demande pas que a soit adhérent à $\mathcal{D}_f \setminus \{a\}$ et se contente de demander que $f(t)$ tende vers $f(a)$ quand t tend vers a. Malgré sa plus grande simplicité formelle, elle me semble un peu plus délicate à manipuler pour un débutant et je l'ai donc légèrement adaptée *ad usum delphini*.

De même qu'il existe un concept de limite à droite, il existe un concept de continuité à droite (et bien sûr de même à gauche...)
Définition 6-1-64 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$. Soit a un point de $\mathcal{D}_f$, supposé adhérent à $\mathcal{D}_f \cap]a, +\infty[$. On dit que f est **continue à droite en** a lorsque $f(t)$ tend vers $f(a)$ quand t tend vers a ($t > a$).
Proposition 6-1-32 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$. Soit a un point de $\mathcal{D}_f$, à la fois adhérent à $\mathcal{D}_f \cap]a, +\infty[$ et à $\mathcal{D}_f \cap]-\infty, a[$. Alors f est continue en a si et seulement si f est simultanément continue à gauche et à droite en a.
Démonstration : Cela découle de la remarque suivant la proposition 5-4-29 : la limite quand $t \to a$ ($t \neq a$) peut être étudiée comme synthèse d'une étude à droite et d'une étude à gauche.

•

Définition 6-1-65 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$. On dit que f est **continue** (ou "continue sur $\mathcal{D}_f$") lorsqu'elle est continue en chaque point de $\mathcal{D}_f$.

2 - Opérations sur les fonctions continues

Rien que des résultats tellement évidents qu'il est à peine besoin de les lire...
Proposition 6-2-33 :

a) L'application identique et les constantes sont continues sur $\mathbf{R}$; l'application $t \mapsto 1/t$ est continue sur $\mathbf{R}^*$.

b) Soit f et g des fonctions réelles d'une variable réelle, respectivement définies sur les ensembles $\mathcal{D}_f$ et $\mathcal{D}_g$, et soit $a \in \mathcal{D}_f$. On suppose que $g \circ f$ existe, que a est adhérent à $\mathcal{D}_f \setminus \{a\}$, que f est continue en a, que $f(a)$ est adhérent à $\mathcal{D}_g \setminus \{f(a)\}$ et que g est continue en $f(a)$. Alors $g \circ f$ est continue en a.

c) Soit f et g des fonctions réelles d'une variable réelle, respectivement définies sur les ensembles $\mathcal{D}_f$ et $\mathcal{D}_g$. On suppose que $g \circ f$ existe et que f et g sont continues. Alors $g \circ f$ est continue.

d) Soit f et g des fonctions réelles d'une variable réelle, définies sur un même ensemble $\mathcal{D}$, et soit $a \in \mathcal{D}$. On suppose que a est adhérent à $\mathcal{D}_f \setminus \{a\}$ et que f et g sont continues en a. Alors $f + g$ et fg sont continues en a.

e) Soit f et g des fonctions réelles d'une variable réelle, définies sur un même ensemble $\mathcal{D}$. On suppose que f et g sont continues. Alors $f + g$ et fg sont continues.

Démonstration :
a) a déjà été remarqué par anticipation dans le chapitre sur les limites...
b) demande un peu plus de soin qu'on ne pourrait le croire, du fait que j'ai choisi de donner une définition malcommode de la continuité... Mais c'est un investissement car l'idée de la preuve resservira — et là de façon incontournable— quand il faudra montrer la dérivabilité d'une composée de fonctions dérivables.

Nous devons montrer que $|(g \circ f)(t) - (g \circ f)(a)| \leq \epsilon$ quand $t \to a$. Notons en vue d'allégement des formules $b = f(a)$, et fixons un $\epsilon > 0$; la définition de "tend vers" appliquée à g fournit un $\eta_1 > 0$ tel que pour tout $u \in \mathcal{D}_g \setminus \{b\}$ vérifiant $|u - b| \leq \eta_1$, on ait $|g(u) - l| \leq \epsilon$. Recommençons en appliquant cette fois la définition de "tend vers" à f et η_1 : on obtient un nouveau réel $\eta > 0$ tel que pour tout $t \in \mathcal{D}_f \setminus \{a\}$ vérifiant $|t-a| \leq \eta$, on ait $|f(t)-b| \leq \eta_1$.

Soit alors un $t \in \mathcal{D}_f \setminus \{a\}$ vérifiant $|t-a| \leq \eta$, donc $|f(t)-b| \leq \eta_1$. On est amené à distinguer deux cas :

* Si $f(t) \neq b$. En notant $u = f(t)$, on dispose alors d'un $u \in \mathcal{D}_g \setminus \{b\}$ qui vérifie $|u-b| \leq \eta_1$; il vérifie donc $|g(u)-g(b)| \leq \epsilon$, c'est-à-dire $|(g \circ f)(t) - (g \circ f)(a)| \leq \epsilon$.

* Si $f(t) = b$, on a alors $(g \circ f)(t) - (g \circ f)(a) = g(b) - g(b) = 0$, et donc $|(g \circ f)(t) - (g \circ f)(a)| \leq \epsilon$.

Dans les deux cas, l'inégalité est bien démontrée.

c) n'est qu'une conséquence immédiate du b).

d) n'est que la conséquence des théorèmes d'addition et de multiplication des limites.

e) n'est que la conséquence immédiate du d).

•

Remarque : Je n'ai pas jugé utile d'expliciter les énoncés analogues pour la soustraction et la division ; ils se déduisent immédiatement de la liste qui précède : la soustraction en multipliant la constante -1 à la fonction g, puis en additionnant f et $-g$, la division en composant $t \mapsto 1/t$ et g, puis en multipliant f et $1/g$.

3 - Comportement vis-à-vis des restrictions

Proposition 6-3-34 : Soit f une fonction réelle d'une variable réelle définie sur un ensemble $\mathcal{D}_f$; soit I un **intervalle ouvert** inclus dans $\mathcal{D}_f$ et soit a un point de I. Alors f est continue en a si et seulement si la restriction $f_{|I}$ est continue en a.

Démonstration :

* Preuve de $\Rightarrow$. C'est le sens facile, qui n'utilise pas le fait que I est un intervalle ouvert : si f est continue en a, $f(t) \to f(a)$ quand $t \to a$ $(t \neq a)$; en appliquant à f le a) de la proposition 5-4-29 (non pas à f mais à $f_{|\mathcal{D}_f \setminus \{a\}}$ remarqueront les puristes) et vu l'hypothèse simplificatrice $I \subset \mathcal{D}_f$ qui simplifie $I \cap \mathcal{D}_f$ en I, on obtient bien que $f_{|I}(t)$ tend aussi vers $f(a)$ $(=f_{|I}(a))$ quand $t \to a$ $(t \neq a)$.

* Preuve de $\Leftarrow$. C'est le morceau sérieux, utilisant le fait que I est un intervalle ouvert ; c'est le même principe que l'autre implication, mais cette fois en utilisant le c) de la proposition 5-4-29.

•

La raison qui justifie l'introduction de cette proposition est qu'on ne se gêne pas pour donner des énoncés tels que "tan est continue sur $]-\frac{\pi}{2}, \frac{\pi}{2}[$" qui n'est pourtant pas l'ensemble de définition de tan. Cet énoncé est sans ambiguïté parce que $]-\frac{\pi}{2}, \frac{\pi}{2}[$ est un intervalle ouvert, et que les deux sens qu'on peut lui donner à savoir "la restriction de tan à $]-\frac{\pi}{2}, \frac{\pi}{2}[$ est continue" et "tan est continue en chaque point de $]-\frac{\pi}{2}, \frac{\pi}{2}[$" sont équivalents. Mais il n'en serait pas de même pour un sous-ensemble qui ne serait pas un intervalle ouvert ! Cette difficulté justifie que je donne, en garde-fou plus que comme une chose à apprendre la

Définition 6-3-66 : Soit f une fonction réelle d'une variable réelle définie sur un ensemble $\mathcal{D}_f$ et soit I un intervalle ouvert inclus dans $\mathcal{D}_f$. On dira que f est **continue sur** I lorsque f est continue en chaque point de I.

Et on s'interdira d'utiliser l'expression ambiguë "continu sur A" pour un ensemble $A \subset \mathcal{D}_f$ qui ne serait pas un intervalle ouvert.

4 - Un théorème à démonstration laissée en suspens

Le cours sur les fonctions dérivables fera un usage crucial du théorème suivant, dont la démonstration sera donnée au second semestre :

Théorème 6-4-10 : Soit f une fonction réelle continue d'une variable réelle définie sur un segment **fermé** [a,b]. Alors il existe un $c_- \in [a,b]$ et un $c_+ \in [a,b]$ tel que pour tout $t \in [a,b]$ on ait :

$$f(c_-) \leq f(t) \leq f(c_+).$$

Démonstration : Au second semestre, vous dis-je !

•

Chapitre 7 - Fonctions dérivables

1 - La définition

Définition 7-1-67 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$. Soit a un point de $\mathcal{D}_f$, supposé adhérent à $\mathcal{D}_f \setminus \{a\}$. On dit que f est **dérivable en** a lorsque $\dfrac{f(t)-f(a)}{t-a}$ admet une limite (finie) quand t tend vers a ($t \neq a$).
Définition 7-1-68 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$, et soit $\mathcal{E}$ l'ensemble des points où f est dérivable. L'application de $\mathcal{E}$ vers $\mathbf{R}$ qui associe à un réel a le réel $\lim\limits_{\substack{t\to a\\ t\neq a}} \dfrac{f(t)-f(a)}{t-a}$ est appelée la **dérivée** de f.
Notation 7-1-35 : La dérivée de f est notée f'.
Définition 7-1-69 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$. Soit a un point de $\mathcal{D}_f$, supposé adhérent à $\mathcal{D}_f \cap]a,+\infty[$. On dit que f est **dérivable à droite en** a lorsque $\dfrac{f(t)-f(a)}{t-a}$ admet une limite (finie) quand t tend vers a ($t > a$).
Définition 7-1-70 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$, et soit $\mathcal{E}_d$ l'ensemble des points où f est dérivable à droite. L'application de $\mathcal{E}_d$ vers $\mathbf{R}$ qui associe à un réel a le réel $\lim\limits_{\substack{t\to a\\ t> a}} \dfrac{f(t)-f(a)}{t-a}$ est appelée la **dérivée à droite** de f.
Notation 7-1-36 : La dérivée à droite de f est notée f'_d (ou f'_+), la dérivée à gauche étant notée f'_g (ou f'_-).
Proposition 7-1-35 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$. Soit a un point de $\mathcal{D}_f$, à la fois adhérent à $\mathcal{D}_f \cap]a,+\infty[$ et à $\mathcal{D}_f \cap]-\infty,a[$. Alors f est dérivable en a si et seulement si f est simultanément dérivable à gauche et à droite en a et vérifie $f'_g(a) = f'_d(a)$.
Démonstration : Comme pour la proposition analogue concernant la continuité, c'est une application de la proposition 5-4-29.

•

Définition 7-1-71 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$. On dit que f est **dérivable** (ou "dérivable sur $\mathcal{D}_f$") si f est dérivable en chaque point de $\mathcal{D}_f$.

2 - Dérivabilité et continuité

Il faut ne pas lire à l'envers la
Proposition 7-2-36 : Soit f une fonction réelle d'une variable réelle définie sur l'ensemble $\mathcal{D}_f$ et soit a un point adhérent à $\mathcal{D}_f$. Si f est dérivable en a, alors elle est continue en a.
Démonstration : Supposons f dérivable en a et écrivons :

$$f(t)-f(a) = \frac{f(t)-f(a)}{t-a}(t-a).$$

Dans cette expression, le facteur $\dfrac{f(t)-f(a)}{t-a}$ tend vers la limite (finie) $f'(a)$ tandis que le facteur $t-a$ tend vers 0 quand t tend vers a ($t \neq a$). Par multiplication des limites, on en déduit que $f(t)-f(a) \to 0$, donc que $f(t) \to f(a)$ quand $t \to a$ ($t \neq a$).

•

Tout étudiant un peu sérieux aura au minimum en tête en suivant ce chapitre la fonction valeur absolue de $\mathbf{R}$ vers $\mathbf{R}$ qui est continue en 0 (parce que continue à droite et continue à gauche) mais pas dérivable en 0.

3 - Opérations sur les fonctions dérivables

Proposition 7-3-37 :

a) L'application identique et les constantes sont dérivables sur $\mathbf{R}$ de dérivées respectives 1 et 0 ; l'application $t \mapsto 1/t$ est dérivable sur $\mathbf{R}^*$, de dérivée $t \mapsto -1/t^2$.

b) Soit f et g des fonctions réelles d'une variable réelle, respectivement définies sur les ensembles $\mathcal{D}_f$ et $\mathcal{D}_g$, et soit $a \in \mathcal{D}_f$. On suppose que $g \circ f$ existe, que a est adhérent à $\mathcal{D}_f \setminus \{a\}$, que f est dérivable en a, que $f(a)$ est adhérent à $\mathcal{D}_g \setminus \{f(a)\}$ et que g est dérivable en $f(a)$. Alors $g \circ f$ est dérivable en a. On a la formule :

$$(g \circ f)'(a) = g'[f(a)]f'(a).$$

c) Soit f et g des fonctions réelles d'une variable réelle, respectivement définies sur les ensembles $\mathcal{D}_f$ et $\mathcal{D}_g$. On suppose que $g \circ f$ existe et que f et g sont dérivables. Alors $g \circ f$ est dérivable.

d) Soit f et g des fonctions réelles d'une variable réelle, définies sur un même ensemble $\mathcal{D}$, et soit $a \in \mathcal{D}$. On suppose que a est adhérent à $\mathcal{D} \setminus \{a\}$ et que f et g sont dérivables en a. Alors $f+g$ et fg sont dérivables en a, avec les formules :

$$(f+g)'(a) = f'(a) + g'(a) \qquad (fg)'(a) = f'(a)g(a) + f(a)g'(a).$$

e) Soit f et g des fonctions réelles d'une variable réelle, définies sur un même ensemble $\mathcal{D}$. On suppose f et g dérivables. Alors $f+g$ et fg sont dérivables.

Démonstration :

(a) Les affirmations concernant l'identité ou les constantes sont évidentes ; pour la fonction $t \mapsto \frac{1}{t}$, soit un $a \in \mathbf{R}^*$; calculons

$$\frac{\frac{1}{t} - \frac{1}{a}}{t-a} = \frac{\frac{a-t}{at}}{t-a} = -\frac{1}{at}$$

Il est alors clair que $-\dfrac{1}{at} \to -\dfrac{1}{a^2}$ quand $t \to a$ $(t \neq a)$.

(b) Il est plus délicat qu'il n'y paraît, car il y a un lieu névralgique de la preuve où il est difficile de résister à l'envie de diviser par zéro... Nous résisterons.

Il y a à considérer le quotient :

$$\frac{g[f(t)] - g[f(a)]}{t-a} \quad \text{quand } t \to a$$

et il est tentant d'écrire :

$$\frac{g[f(t)] - g[f(a)]}{t-a} = \frac{g[f(t)] - g[f(a)]}{f(t) - f(a)} \frac{f(t) - f(a)}{t-a}$$

mais ça ne marche évidemment pas si $f(t) - f(a)$ est nul.

Heureusement, la proposition 5-4-29 sera totalement efficace pour régler cette difficulté.

Pour pouvoir l'utiliser de façon claire, distinguons deux cas, celui qui marche bien et celui où il faut réfléchir.

* Premier cas : s'il existe un réel $\delta > 0$ tel que sur $[a-\delta, a+\delta]$ la fonction f ne prenne la valeur $f(a)$ qu'au seul point a.

Dans ce cas, notons $I =]a-\delta, a+\delta[$, qui est un intervalle ouvert. Sur cet intervalle (pour $t \neq a$), on peut diviser par $f(t) - f(a)$ et écrire valablement :

$$\frac{g[f(t)] - g[f(a)]}{t-a} = \frac{g[f(t)] - g[f(a)]}{f(t) - f(a)} \frac{f(t) - f(a)}{t-a}.$$

Maintenant, quand $t \to a$ $(t \neq a)$, $f(t) \to f(a)$ (continuité des fonctions dérivables), et quand $u \to f(a)$ $(u \neq f(a))$, $\dfrac{g[u] - g[f(a)]}{u - f(a)} \to g'[f(a)]$ (définition d'une dérivée) donc

$$\frac{g[f(t)] - g[f(a)]}{f(t) - f(a)} \to g'[f(a)].$$

(composition des limites) Par ailleurs $\dfrac{f(t) - f(a)}{t - a} \to f'(a)$ (définition d'une dérivée), d'où le résultat (multiplication des limites). (Observation pour les gens pointilleux : le c) de la proposition 5-4-29 est discrètement utilisé, car on n'obtient en bonne rigueur le résultat que pour des fonctions restreintes à I et on doit remonter à un énoncé sur $\mathcal{D}_f$).

Dans ce premier cas, tout marche donc comme sur des roulettes.

* Second cas : si pour tout réel $\delta > 0$ il existe au moins une valeur $t \neq a$ dans $[a - \delta, a + \delta]$ telle que $f(t) = f(a)$.

Dans ce second cas, notons $\mathcal{D}_1$ l'ensemble des t de $\mathcal{D}_f$ tels que $f(t) = f(a)$. L'hypothèse ouvrant ce second cas affirme exactement que a est adhérent à $\mathcal{D}_1$. On peut donc calculer $f'(a)$ en utilisant non f mais la restriction de f à $\mathcal{D}_1$ (proposition 5-4-29 a)), qui vaut constamment $f(a)$, donc en dérivant une fonction constante. On en déduit déjà que $f'(a) = 0$.

Distinguons à partir de là deux sous-cas.

• Premier sous-cas (stupide) S'il existe un réel $\delta > 0$ tel que sur $[a - \delta, a + \delta]$ la fonction f prenne constamment la valeur $f(a)$.

Dans ce cas, sur l'intervalle $I =]a - \delta, a + \delta[$, $g[f(t)] - g[f(a)]$ vaut constamment 0, donc le quotient qu'on doit étudier aussi : il tend donc vers 0 quand $t \to a$, $t \neq a$ (en bonne rigueur pour des fonctions restreintes à I, qui est un intervalle ouvert, donc aussi pour les fonctions initiales). Et 0, c'est bien $g'[f(a)]f'(a)$ puisque $f'(a) = 0$.

• Second sous-cas (le vrai sous-cas à problème) Si pour tout $\delta > 0$ il existe un $t \neq a$ dans $[a - \delta, a + \delta]$ tel que $f(t) \neq f(a)$.

Dans ce cas, adjoignons à notre notation $\mathcal{D}_1$, ensemble des t de $\mathcal{D}_f$ tels que $f(t) = f(a)$ la notation $\mathcal{D}_2$, ensemble des t de $\mathcal{D}_f$ tels que $f(t) \neq f(a)$. La nouvelle hypothèse assure que a est également adhérent à $\mathcal{D}_2$. Maintenant la proposition affirmée se prouve comme dans le premier cas sur $\mathcal{D}_2$ et comme dans le sous-cas précédent sur $\mathcal{D}_1$. Il n'y a plus qu'à relire le b) de la proposition 5-4-29 pour conclure.

Le b) est donc vrai aussi dans le second cas, donc dans tous les cas. Ouf !

(c) Là on souffle : ce n'est que l'application du b) en tous les points de $\mathcal{D}_f$.

(d) Pour la somme, c'est vraiment trop facile, passons au produit.

On suppose f et g dérivables en a, on écrit alors, pour $t \in \mathcal{D}$:

$$\begin{aligned}\frac{(fg)(t) - (fg)(a)}{t - a} &= \frac{f(t)g(t) - f(t)g(a) + f(t)g(a) - f(a)g(a)}{t - a} \\ &= \frac{f(t)g(t) - f(t)g(a)}{t - a} + \frac{f(t)g(a) - f(a)g(a)}{t - a}. \\ &= f(t)\frac{g(t) - g(a)}{t - a} + g(a)\frac{f(t) - f(a)}{t - a}\end{aligned}$$

Dans cette expression, la limite de chaque terme est limpide : $f(t)$ tend vers $f(a)$ par continuité des fonctions dérivables, les quotients tendent vers les dérivées respectives de f et g en a ; le gros quotient admet donc bien une limite, qui est donnée par la formule bien connue.

(e) C'est simplement le d) lorsqu'il est vrai en tous points de $\mathcal{D}$.

•

4 - Comportement vis-à-vis des restrictions

Comme pour les fonctions continues, on a la

Proposition 7-4-38 : Soit f une fonction réelle d'une variable réelle définie sur un ensemble $\mathcal{D}_f$; soit I un **intervalle ouvert** inclus dans $\mathcal{D}_f$ et soit a un point de I. Alors f est dérivable en a si et seulement si la restriction $f_{|I}$ est dérivable en a (et ils ont bien sûr les mêmes dérivées).

Démonstration : On commence à s'en lasser de ces arguties autour de la proposition 5-4-29. Disons que c'est exactement pareil que l'énoncé analogue pour les fonctions continues, et n'en parlons plus.

•

Comme pour les fonctions continues, donnons la

Définition 7-4-72 : Soit f une fonction réelle d'une variable réelle définie sur un ensemble $\mathcal{D}_f$ et soit I un intervalle ouvert inclus dans $\mathcal{D}_f$. On dira que f est **dérivable sur** I lorsque f est dérivable en chaque point de I.

Et interdisons l'expression ambiguë "dérivable sur A" pour un ensemble $A \subset \mathcal{D}_f$ qui ne serait pas un intervalle ouvert.

La différence qui existe avec les fonctions continues, c'est qu'avec les fonctions dérivables, l'ambiguïté interdite recèle de vrais pièges, dans des vraies fonctions de vrais exercices, comme le montrera l'exemple qui termine cette section.

Pour pouvoir utiliser ces notions sans les comprendre, donnons un énoncé évident si on a compris, mais d'usage sans doute plus aisé que celui qui précède.

Proposition 7-4-39 : Soit f et g deux fonctions réelles d'une variable réelle et I un intervalle **ouvert**, qu'on supposera inclus dans les deux ensembles de définition de f et de g. On suppose qu'en tout point t de I, $f(t) = g(t)$.

Si g est dérivable en un point a de I, alors f aussi, et $f'(a) = g'(a)$.

Démonstration : D'après la proposition qui précède, si g est dérivable en a, $g_{|I}$ aussi. Mais par hypothèse $g_{|I} = f_{|I}$. On applique de nouveau la proposition qui précède et on déduit que f est dérivable en a avec la même dérivée.

•

Cela ne marche pas du tout si I n'est pas un intervalle ouvert ! Si f et g sont les fonctions de $\mathbf{R}$ vers $\mathbf{R}$ respectivement définies par $f(t) = |t|$ et $g(t) = t$, f et g coïncident sur $\mathbf{R}^+$, g est dérivable en 0, et pourtant f n'est pas dérivable en 0.

5 - Extrema : première couche

Cette première couche se réduit à accumuler quelques définitions...

Définition 7-5-73 : Soit f une fonction réelle d'une variable réelle définie sur un ensemble $\mathcal{D}_f$, et soit c un point de $\mathcal{D}_f$. On dit que f admet un **maximum** (ou **maximum global** en c si on craint les confusions) lorsque pour tout $x \in \mathcal{D}_f$, $f(x) \leq f(c)$.

Définition 7-5-74 : Soit f une fonction réelle d'une variable réelle définie sur un ensemble $\mathcal{D}_f$, et soit c un point de $\mathcal{D}_f$. On dit que f admet un **maximum strict** (ou **maximum global strict** si on craint les confusions) en c lorsque pour tout $x \in \mathcal{D}_f$, $x \neq c$, $f(x) < f(c)$.

Définition 7-5-75 : Soit f une fonction réelle d'une variable réelle définie sur un ensemble $\mathcal{D}_f$, et soit c un point de $\mathcal{D}_f$. On dit que f admet un **maximum local** en c lorsqu'il existe un réel $\delta > 0$ tel que pour tout $x \in \mathcal{D}_f$ tel que $|x - c| \leq \delta$, on ait l'inégalité $f(x) \leq f(c)$.

Définition 7-5-76 : Soit f une fonction réelle d'une variable réelle définie sur un ensemble $\mathcal{D}_f$, et soit c un point de $\mathcal{D}_f$. On dit que f admet un **maximum local strict** en c lorsqu'il existe un réel $\delta > 0$ tel que pour tout $x \in \mathcal{D}_f$ $(x \neq c)$ tel que $|x - c| \leq \delta$, on ait l'inégalité $f(x) < f(c)$.

On définit de même évidemment les minimums ; "extremum" signifiera "maximum ou minimum".

Ce vocabulaire possède un petit inconvénient : le mot "maximum" (que je n'ai pas formellement défini) désigne plutôt la valeur $f(c)$ alors qu'on a plus souvent envie de parler de c, qui n'a pas de nom bien établi... Il faudra se résigner à des périphrases plus ou moins habiles —et à tolérer bien des impropriétés dans les copies...

Avec ce vocabulaire, le théorème 6-4-10 se réécrit ainsi : toute fonction continue réelle définie sur un segment fermé de $\mathbf{R}$ admet un maximum et un minimum en des points du segment.

Pour faire le lien avec la dérivation, encore un mot :

Définition 7-5-77 : Soit f une fonction réelle d'une variable réelle définie sur un ensemble $\mathcal{D}_f$, et soit c un point de $\mathcal{D}_f$. On dit que c est un **point critique** de f lorsque f est dérivable en c et $f'(c) = 0$.

Un résultat facile et d'usage banal est le suivant :

Proposition 7-5-40 : Soit f une fonction réelle d'une variable réelle définie sur un **intervalle ouvert** I, et soit c un point de $\mathcal{D}_f$. Si f admet un extremum local en c et est dérivable au point c, alors c est un point critique de f.

Démonstration : (écrite dans le cas d'un maximum en c). Vu l'hypothèse simplificatrice sur l'ensemble de définition de f, cela a un sens de calculer tant la dérivée à droite que la dérivée à gauche de f au point c.

Soit $\delta > 0$ un réel fourni par la définition d'"extremum local" ; pour $c < t \leq c + \delta$, le quotient $\dfrac{f(t) - f(c)}{t - c}$ est négatif, car $f(t) \leq f(c)$ et $t - c > 0$. Par passage à la limite, sa limite est aussi négative, et cette limite est la dérivée à droite de f en c. De l'autre côté, pour $c - \delta \leq t < c$, le quotient $\dfrac{f(t) - f(c)}{t - c}$ est positif, donc la dérivée à gauche de f en c est positive. La dérivée $f'(c)$ est donc à la fois positive et négative, donc nulle.
•

On remarquera qu'une hypothèse restrictive sur l'ensemble de définition est indispensable : la fonction f définie sur $[0, 1]$ par $f(t) = t$ admet un maximum en 1 bien que sa dérivée ne s'y annule pas.

On veillera aussi à ne pas mémoriser l'implication à l'envers ! Un bon réflexe est de se souvenir que $f(t) = t^3$ fournit un contre-exemple significatif dans ce genre de questions : effectivement, elle admet un point critique mais pas d'extremum (même local) en 0.

Pour résumer ce que nous savons faire, en mettant bout à bout cette proposition et diverses évidences :

$$\begin{array}{ccc}
f \text{ admet un maximum strict en } c & \Rightarrow & f \text{ admet un maximum local strict en } c \\
\Downarrow & & \Downarrow \\
f \text{ admet un maximum en } c & \Rightarrow & f \text{ admet un maximum local en } c \\
& & \Downarrow \\
& & \text{(ne marche que sur un intervalle } \textbf{ouvert} \text{ !)} \\
& & c \text{ est critique pour } f
\end{array}$$

6 - Le théorème de Rolle

Avant d'aller plus loin, une convention de langage, (qui s'avère astreignante, mais tant pis...) Pour deux réels a et b pas forcément dans cet ordre, quand je parlerai du **segment** $[a, b]$, ce sera l'intervalle $[a, b]$ si $a \leq b$ et l'intervalle $[b, a]$ si $b \leq a$. On ne dispose malheureusement que d'une unique notation pour les intervalles ou les segments, donc je serai condamné à taper le mot segment chaque fois que je ne veux pas supposer $a \leq b$ (pour le lecteur, ce devrait être moins pénible). De façon analogue, quand je dirai "c est entre a et b", cela signifie "c est dans le segment $[a, b]$" (c'est-à-dire : si $a \leq b$, $a \leq c \leq b$, tandis que si $b \leq a$, $b \leq c \leq a$.)

Théorème 7-6-11 : Soit f une fonction réelle définie sur un segment $[a, b]$ (avec $a \neq b$), continue sur ce segment et dérivable sur le segment $]a, b[$.

On suppose en outre que $f(a) = f(b)$. Alors il existe un c strictement entre a et b tel que $f'(c) = 0$.

Démonstration : L'idée de la preuve est de se placer en un maximum de f ; en un maximum de f la dérivée de f est nulle. C'est en gros l'idée, mais si on se borne à cela, on n'utilise pas l'hypothèse $f(a) = f(b)$, ce qui sent l'erreur ! C'est que si le maximum se produit en une borne, il n'y a pas de garantie que f' s'annule en ce point, et il faut attaquer par les minimums... Faisons cela formellement en travaillant à la fois sur les maxima et les minima.

En appliquant le théorème $6 - 4 - 10$ à la fonction continue f sur le segment fermé $[a, b]$, on obtient deux points d et e du segment $[a, b]$ tels que f admette un minimum (global) en d et un maximum (global) en e.

On distingue alors deux cas :

* Premier cas : l'un au moins des deux réels d ou e est dans le segment ouvert $]a, b[$.

Dans ce cas, prenons c dans le segment $]a, b[$ égal soit à d soit à e. Puisque f admet un extremum en c sur le segment **ouvert** $]a, b[$, on en déduit que $f'(c) = 0$. La preuve est finie.

* Deuxième cas : les deux réels d et e sont tous les deux dans $\{a, b\}$.

Dans ce cas, notons quelques instants k la valeur commune de $f(a)$ et de $f(b)$ (c'est ici qu'on utilise discrètement l'hypothèse $f(a) = f(b)$). Puisque d est égal à a ou à b, on obtient $f(d) = k$. De même, $f(e) = k$. Mais pour tout t du segment $[a, b]$, $f(d) \leq f(t) \leq f(e)$, donc $k \leq f(t) \leq k$, donc $f(t) = k$. La fonction f est donc constante, et sa dérivée est donc nulle partout sur le segment $[a, b]$. N'importe quel c du segment ouvert $]a, b[$ convient donc.
•

7 - Le théorème des accroissements finis

Lorsque l'on pert l'information $f(a) = f(b)$, on peut néanmoins obtenir un résultat d'énoncé à peine un peu plus lourd que le théorème de Rolle, tout aussi intuitif, et qui s'en déduit immédiatement. C'est le

Théorème 7-7-12 : Soit f une fonction réelle définie sur un segment $[a, b]$ (avec $a \neq b$), continue sur ce segment et dérivable sur le segment $]a, b[$.

Alors il existe un c strictement entre a et b tel que $f'(c) = \dfrac{f(b)-f(a)}{b-a}$.

Démonstration : L'idée est de construire une fonction auxiliaire g liée à f par une formule très simple mais qui ait le bon goût de vérifier l'hypothèse supplémentaire $g(a) = g(b)$, et donc d'accepter l'application du théorème de Rolle.

On va poser

$$g(t) = f(t) - \frac{f(b)-f(a)}{b-a}(t-a)$$

(l'usage de t au lieu de $t-a$ suffirait à faire marcher la preuve, mais l'usage de $t-a$ prépare à des constructions plus compliquées au chapitre suivant).

On vérifie aussitôt que $g(a) = f(a) - 0 = f(a)$ tandis que $g(b) = f(b) - \dfrac{f(b)-f(a)}{b-a}(b-a) = f(b) - f(b) + f(a)$. On applique alors le théorème de Rolle à g pour obtenir un réel c dans le segment $]a,b[$ tel que $g'(c) = 0$, soit $f'(c) - \dfrac{f(b)-f(a)}{b-a} = 0$.

•

8 - Dérivées et sens de variation

Les résultats de cette section sont couramment utilisés sans s'en apercevoir (quand on remplit un tableau de variations, notamment). Il est tout de même bien utile de se souvenir qu'ils concernent tous des fonctions définies sur **un intervalle**. Je souligne tout de suite que la toute simple fonction "signe" définie sur $\mathbf{R}^*$ par $\text{signe}(t) = -1$ si $t < 0$ et $\text{signe}(t) = 1$ si $t > 0$ est dérivable partout (où elle est définie), de dérivée nulle, sans pour autant être constante !

Passons aux énoncés

Proposition 7-8-41 : Soit f une fonction dérivable définie sur un **intervalle** I. f est croissante sur I si et seulement si $f' \geq 0$ sur I.

Démonstration :

* Vérification de $\Rightarrow$.

Supposons f croissante sur I et soit t_0 un point de I. Considérons le quotient $\dfrac{f(t)-f(t_0)}{t-t_0}$ pour $t \neq t_0$ élément de I. Comme f est supposée croissante, $f(t)-f(t_0)$ est de même signe (au sens large) que $t-t_0$, donc le quotient considéré est positif (au sens large). Sa limite quand $t \to t_0$ ($t \neq t_0$) est donc elle-même positive, soit $f'(t_0) \geq 0$. (Ce sens serait vrai même sur un I plus ou moins biscornu).

* Vérification de $\Leftarrow$.

Supposons $f' \geq 0$ sur I, et soit $s < t$ deux éléments de I. Comme I est un intervalle, l'intervalle $[s,t]$ est inclus dans I et on peut lui appliquer le théorème des accroissements finis. Il existe donc un $c \in]s,t[$ tel que $\dfrac{f(t)-f(s)}{t-s} = f'(c)$. Donc $f(t)-f(s) = f'(c)(t-s)$ est le produit de deux réels positifs et est lui-même positif. Ainsi $f(s) \leq f(t)$. Ceci prouve la croissance de f.

•

Proposition 7-8-42 : Soit f une fonction dérivable définie sur un **intervalle** I. f est constante sur I si et seulement si $f' = 0$ sur I.

Démonstration : f est constante si et seulement si elle est à la fois croissante et décroissante, donc si et seulement si f' est à la fois positive et négative, soit si et seulement si $f' = 0$.

•

Pour la croissance stricte, les choses ne marchent pas aussi bien, un seul sens est vrai...

Proposition 7-8-43 : Soit f une fonction dérivable définie sur un **intervalle** I. Si $f' > 0$ sur I, alors f est strictement croissante sur I.

Démonstration : C'est la même que pour l'implication $\Leftarrow$ de la proposition concernant la croissance : la nouveauté est qu'ici on peut affirmer que $f'(c) > 0$ et donc déduire que $f(s) < f(t)$.

•

Chapitre 8 - Applications linéaires

Dans tout ce chapitre, en vu d'une relecture possible dès que les définitions des espaces vectoriels "les plus généraux" auront été données, on conviendra provisoirement que "soit E un espace vectoriel" ou "soit E un espace vectoriel de dimension finie" est provisoirement une abréviation de "soit $k \geq 0$ un entier et soit E un sous-espace vectoriel de $\mathbf{R}^k$" ; "soit E_1 un sous-espace vectoriel inclus dans E" sera provisoirement un alourdissement de "soit F un espace vectoriel inclus dans E".

1 - Des définitions

Définition 8-1-78 : Soit E et F deux espaces vectoriels et u une application de E vers F. On dit que u est une **application linéaire** lorsque
(i) Pour tous x, y de E, $u(x+y) = u(x) + u(y)$.
(ii) Pour tout λ réel, et tout x de E, $u(\lambda x) = \lambda u(x)$.

On peut aussi —si l'on n'a pas peur d'être lourd et qu'on souhaite mettre en relief la ressemblance de la notion avec celle qu'on verra bientôt pour les groupes— parler de **morphisme d'espaces vectoriels**.
Définition 8-1-79 : Une application linéaire bijective est dite **isomorphisme** (on précisera "d'espaces vectoriels" si on est dans un contexte faisant redouter une ambiguïté).
Notation 8-1-37 : L'ensemble des applications linéaires de E vers F sera noté $\mathcal{L}(E, F)$.
Définition 8-1-80 : Une application linéaire dont l'espace de départ est égal à l'espace d'arrivée sera appelée un **endomorphisme**.
Définition 8-1-81 : Un endomorphisme bijectif pourra être appelé un **automorphisme** (je ne trouve pas ce terme très utile, et m'abstiens généralement de l'utiliser).
Notation 8-1-38 : On notera $\mathcal{L}(E)$ pour $\mathcal{L}(E, E)$.

De même qu'on peut facilement caractériser les sous-espaces vectoriels avec une petite variante, on peut facilement caractériser les applications linéaires avec une petite variante :
Proposition 8-1-44 : Une application u d'un espace vectoriel E vers un espace vectoriel F est linéaire si et seulement si :

pour tous x, y de E et tout λ réel, $u(\lambda x + y) = \lambda u(x) + u(y)$.

Démonstration :
* Preuve de $\Rightarrow$. Supposons f linéaire, et soit x, y dans E et λ réel. On a alors :

$$u(\lambda x + y) = u(\lambda x) + u(y) = \lambda u(x) + u(y).$$

* Preuve de $\Leftarrow$. Supposons que f vérifie la caractérisation de l'énoncé de la proposition. Soit x, y réels. On a alors $u(x+y) = u(1x+y) = 1u(x)+u(y) = u(x)+u(y)$. On en déduit ensuite que $u(0+0) = u(0)+u(0)$ donc $u(0) = 0$; soit alors x dans E et λ un réel ; on a alors $u(\lambda x) = u(\lambda x+0) = \lambda u(x)+u(0) = \lambda u(x)+0 = \lambda u(x)$. u est donc bien linéaire.

•

2 - Opérations sur les applications linéaires

Définition 8-2-82 : Soit f et g deux applications définies sur un même ensemble A et à valeurs dans un même espace vectoriel F. La **somme** de f et g est l'application $f+g$ définie pour chaque x de A par : $(f+g)(x) = f(x) + g(x)$.

Les étudiants consciencieux remarqueront que cette définition fait double emploi entre celle donnée pour ouvrir le premier chapitre d'analyse... C'est (presque) vrai, mais j'ai préféré un doublon, les contextes étant très différents.
Définition 8-2-83 : Soit f une application définie sur un ensemble A et à valeurs dans un espace vectoriel F et soit λ un réel. On ne donne pas de nom très précis à l'application λf définie pour chaque x de A par : $(\lambda f)(x) = \lambda f(x)$.
Proposition 8-2-45 : Soit E, F, G trois espaces vectoriels, f, f_1, f_2 applications de E vers F, g_1, g_2 applications de F vers G et g application **linéaire** de F vers G.

Alors

$$(g_1 + g_2) \circ f = g_1 \circ f + g_2 \circ f$$

$$g \circ (f_1 + f_2) = g \circ f_1 + g \circ f_2.$$

Démonstration : Ce n'est que vérifications stupides. Pour $x \in E$, $[(g_1 + g_2) \circ f](x) = (g_1 + g_2)[f(x)] = (g_1 \circ f)(x) + (g_2 \circ f)(x) = (g_1 \circ f + g_2 \circ f)(x)$, tandis que $[g \circ f_1 + g \circ f_2](x) = (g \circ f_1)(x) + (g \circ f_2)(x) = g[f_1(x)]+g[f_2(x)] = g[f_1(x)+f_2(x)]$ (c'est ici le seul endroit où on utilise la linéarité de l'une des applications, g en l'occurence), donc $[g \circ f_1 + g \circ f_2](x) = g[(f_1 + f_2)(x)]$.
•

(Il est inutile de faire de gros efforts pour se souvenir où la linéarité est indispensable et où elle ne l'est pas, en pratique cette proposition sera utilisée pour des applications toutes linéaires...)
On utilisera régulièrement sans même s'en rendre compte l'évidente
Proposition 8-2-46 : Soit E un espace vectoriel et E_1 un sous-espace de E ; soit F un espace vectoriel et F_1 un sous-espace F. Si u de E vers F est linéaire et si la restriction de u à E_1 et F_1 existe, elle est elle-même linéaire.
Démonstration : C'est creux et évident.
•

Enfin, on utilisera peut-être de ci de là la très facile à énoncer
Proposition 8-2-47 : Soit E et F deux espaces vectoriels et u une application linéaire bijective de E vers F. Alors la réciproque u^{-1} est elle-même linéaire.
Démonstration : Soit y_1, y_2 deux éléments de F et λ un réel. On doit montrer l'égalité des deux vecteurs $x = \lambda u^{-1}(y_1) + u^{-1}(y_2)$ et $x' = u^{-1}(\lambda y_1 + y_2)$. Pour cela, comparons tout d'abord $u(x)$ et $u(x')$. On calcule, en utilisant la linéarité de u, $u(x) = u[\lambda u^{-1}(y_1) + u^{-1}(y_2)] = \lambda u[u^{-1}(y_1)] + u[u^{-1}(y_2)] = \lambda y_1 + y_2$; on calcule de façon totalement évidente $u(x') = u[u^{-1}(\lambda y_1 + y_2)] = \lambda y_1 + y_2$. On a trouvé la même chose, donc $u(x) = u(x')$; comme u est injective, on en déduit que $x = x'$. L'application u^{-1} est donc linéaire.
•

3 - Applications linéaires et bases

L'énoncé suivant est fondamental, en ce qu'il nous permettra de définir la matrice d'une application linéaire.
Proposition 8-3-48 : Soit E un espace vectoriel de dimension finie et F un espace vectoriel. Soit $(e_1, \ldots, e_k)$ une base de E et $(f_1, \ldots, f_k)$ un système de vecteurs de F. Il existe une et une seule application linéaire telle que

$$u(e_1) = f_1, \ldots, u(e_k) = f_k.$$

Démonstration : On va d'abord montrer qu'il existe au plus une telle u linéaire en explicitant une formule qu'elle vérifie forcément ; puis on prouvera son existence en vérifiant que cette formule définit bien une application linéaire de E vers F vérifiant la propriété souhaitée.

Soit donc u une application linéaire de E vers F vérifiant $u(e_1) = f_1, \ldots, u(e_k) = f_k$. Soit x un vecteur de E. Puisqu'on a supposé que $(e_1, \ldots, e_k)$ est une base de E, on peut parler des coordonnées de x dans $(e_1, \ldots, e_k)$ et les noter $\alpha_1, \ldots, \alpha_k$. On a alors $u(x) = u(\alpha_1 e_1 + \cdots + \alpha_k e_k) = \alpha_1 u(e_1) + \cdots + \alpha_k u(e_k) = \alpha_1 f_1 + \cdots + \alpha_k f_k$, donc u est déterminée sur chacun des vecteurs de E ; elle est donc unique si elle veut bien exister.

Pour l'existence, on a alors — si on ne l'avait déjà — idée de la bonne formule : pour x dans E de coordonnées $\alpha_1, \ldots, \alpha_k$ dans $(e_1, \ldots, e_k)$, posons $u(x) = \alpha_1 f_1 + \cdots + \alpha_k f_k$. La vérification de la linéarité de u est peu passionnante à lire, encore moins à écrire, et est très très facile.
•

4 - Noyau et injectivité

Proposition 8-4-49 : Soit E et F deux espaces vectoriels et u une application linéaire de E vers F. Soit F_1 un sous-espace vectoriel de F. Alors $u^{-1}(F_1)$ est un sous-espace vectoriel de E.
Démonstration :

* Vérifions tout d'abord que $u^{-1}(F_1)$ n'est pas vide : c'est évident car $u(0) = 0$ et $0 \in F_1$ donc $0 \in u^{-1}(F_1)$.

* Soit maintenant x_1 et x_2 dans $u^{-1}(F_1)$ et λ un réel. Intéressons nous au vecteur λx_1+x_2 en considérant son image par u : $u(\lambda x_1+x_2)=\lambda u(x_1)+u(x_2)$ puisqu'u est linéaire ; dans cette expression $u(x_1)$ et $u(x_2)$ sont tous deux dans F_1 puisque x_1 et x_2 sont tous deux dans $u^{-1}(F_1)$, donc, étant donné que F_1 est un sous-espace vectoriel, $u(\lambda x_1+x_2)=\lambda u(x_1)+u(x_2)\in F_1$. On conclut comme on le souhaitait que $\lambda x_1+x_2\in u^{-1}(F_1)$. •

Définition 8-4-84 : Soit E et F deux espaces vectoriels et u une application linéaire de E vers F. Le sous-espace vectoriel $u^{-1}(\{0\})$ de E est appelé le **noyau** de u.
Notation 8-4-39 : Le noyau de u est noté $\operatorname{Ker} u$.

Pour ceux qui trouveraient trop difficiles d'absorber la notation u^{-1}, ils pourront tout simplement retenir que $\operatorname{Ker} u=\{x\in E\,|\,u(x)=0\}$.
Proposition 8-4-50 : Soit E et F deux espaces vectoriels et u une application linéaire de E vers F. u est injective si et seulement si $\operatorname{Ker} u=\{0\}$.
Démonstration :

Sans surprise, vérifions successivement les deux implications.
Preuve de $\Rightarrow$.
Supposons u injective.

On a remarqué que $u(0)=0$, et donc que $\{0\}\subset\operatorname{Ker} u$. Réciproquement, si $x\in\operatorname{Ker} u$, $f(x)=f(0)=0$, et comme u est injective, $x=0$. D'où l'égalité $\{0\}=\operatorname{Ker} u$.

Preuve de $\Leftarrow$.
Supposons $\operatorname{Ker} u=\{0\}$.

Soit x_1 et x_2 deux éléments de E vérifiant $u(x_1)=u(x_2)$. On a alors $u(x_1-x_2)=u(x_1)-u(x_2)=0-0=0$, donc $x_1-x_2\in\operatorname{Ker} u$, donc $x_1-x_2=0$, donc $x_1=x_2$. u est bien injective. •

5 - Image et surjectivité

Proposition 8-5-51 : Soit E et F deux espaces vectoriels et u une application linéaire de E vers F. Soit E_1 un sous-espace vectoriel de E. Alors $u(E_1)$ est un sous-espace vectoriel de F.
Démonstration :

* Vérifions tout d'abord que $u(E_1)$ n'est pas vide : c'est évident car $u(0)=0$ et $0\in E_1$ donc $0\in u(E_1)$.

* Soit maintenant y_1 et y_2 dans $u(E_1)$ et λ un réel. Intéressons nous au vecteur λy_1+y_2 ; puisque $y_1\in u(E_1)$ et $y_2\in u(E_2)$, il existe des vecteurs x_1 et x_2 dans E_1 tels que $y_1=u(x_1)$ et $y_2=u(x_2)$; on a alors $\lambda y_1+y_2=\lambda u(x_1)+u(x_2)=u(\lambda x_1+x_2)$ (en utilisant la linéarité de u). Comme E_1 est un sous-espace vectoriel, $\lambda x_1+x_2\in E_1$, et donc $\lambda y_1+y_2=u(\lambda x_1+x_2)\in u(E_1)$. •

Définition 8-5-85 : Soit E et F deux espaces vectoriels et u une application linéaire de E vers F. Le sous-espace vectoriel $u(E)$ de F est appelé l'**image** de u.
Notation 8-5-40 : L'image de u est notée $\operatorname{Im} u$.
Proposition 8-5-52 : Soit E et F deux espaces vectoriels et u une application linéaire de E vers F. u est surjective si et seulement si $\operatorname{Im} u=F$.
Démonstration : Il n'y a rien à montrer, c'est totalement tautologique. •

6 - La formule du rang

Théorème 8-6-13 : Soit E un espace vectoriel de dimension finie et F un espace vectoriel ; soit u une application linéaire de E vers F. On a alors :

$$\dim\operatorname{Ker} u+\dim\operatorname{Im} u=\dim E.$$

Démonstration : Commençons par prendre une base $(e_1,\ldots,e_k)$ de $\operatorname{Ker} u$, et, en utilisant le théorème de la base incomplète, complétons la en une base $(e_1,\ldots,e_k,e_{k+1},\ldots,e_n)$ de E. Avec ces notations, $\dim\operatorname{Ker} u=k$ et $\dim E=n$; on doit donc arriver à prouver que $\dim\operatorname{Im} u=n-k$. Pour cela, on va prouver que $(u(e_{k+1}),\ldots,u(e_n))$ est une base de $\operatorname{Im} u$.

Montrons dans un premier temps que $(u(e_{k+1}),\ldots,u(e_n))$ est un système générateur de $\operatorname{Im} u$. Soit y un vecteur de $\operatorname{Im} u$. Il existe donc un $x\in E$ tel que $y=u(x)$. On peut écrire x dans la base $(e_1,\ldots,e_k,e_{k+1},\ldots,e_n)$ de E, soit $x=\alpha_1e_1+\cdots+\alpha_ke_k+\alpha_{k+1}e_{k+1}+\cdots+\alpha_ne_n$. Donc

$$\begin{aligned} y = u(x) = & u[\alpha_1 e_1 + \cdots + \alpha_k e_k + \alpha_{k+1} e_{k+1} + \cdots + \alpha_n e_n] \\ & = \alpha_1 u(e_1) + \cdots + \alpha_k u(e_k) + \alpha_{k+1} u(e_{k+1}) + \cdots + \alpha_n u(e_n) \\ & = 0 + \cdots 0 + \alpha_{k+1} u(e_{k+1}) + \cdots + \alpha_n u(e_n) \end{aligned}$$

On a bien réussi à écrire x comme combinaison linéaire de $(u(e_{k+1}), \dots, u(e_n))$.

Montrons dans un second temps que $(u(e_{k+1}), \dots, u(e_n))$ est libre. Soit des scalaires $\lambda_{k+1}, \dots, \lambda_n$ tels que $\lambda_{k+1} u(e_{k+1}) + \cdots + \lambda_n u(e_n) = 0$, soit $u(\lambda_{k+1} e_{k+1} + \cdots + \lambda_n e_n) = 0$, soit $\lambda_{k+1} e_{k+1} + \cdots + \lambda_n e_n \in \operatorname{Ker} u$. Il existe dès lors des scalaires $\lambda_1, \dots, \lambda_k$ tels que $\lambda_{k+1} e_{k+1} + \cdots + \lambda_n e_n = \lambda_1 e_1 + \cdots + \lambda_k e_k$. La liberté du gros système $(e_1, \dots, e_k, e_{k+1}, \dots, e_n)$ entraîne alors la nullité de tous les λ_i et en particulier celles qui nous intéressent, à savoir la conclusion $\lambda_{k+1} = \cdots = \lambda_n = 0$.

•

7 - Critères de bijectivité

Lorsque la dimension de l'espace de départ et celle de l'espace d'arrivée sont égales, on dispose de critères de bijectivité particulièrement confortables ; on pourra faire un rapprochement avec la façon dont on peut prouver la bijectivité d'une application entre deux ensembles finis ayant le même nombre d'éléments. Les critères qui suivent seront tout particulièrement utiles pour des endomorphismes.

Théorème 8-7-14 : Soit E et F deux espaces vectoriels de dimension finie et soit u une application linéaire de E vers F. On suppose que $\dim E = \dim F$. Alors :

* Si u est injective, elle est bijective.

* Si u est surjective, elle est bijective.

Démonstration :

Notons n la dimension commune de E et F. Alors u est injective $\iff$ $\operatorname{Ker} u = \{0\}$ $\iff$ $\dim \operatorname{Ker} u = 0$ $\iff$ $n - \dim \operatorname{Im} u = 0$ $\iff$ $\dim \operatorname{Im} u = \dim F$ $\iff$ $\operatorname{Im} u = F$ $\iff$ u est surjective.

•

Le critère qui suit sert assez peu souvent, mais mérite néanmoins d'être connu —et mémorisé à long terme— car il peut significativement simplifier une démonstration.

Proposition 8-7-53 : Soit E et F deux espaces vectoriels de dimension finie et soit u une application linéaire de E vers F. On suppose que $\dim E = \dim F$. Alors :

* S'il existe une application v de F vers E telle que $u \circ v = Id_F$, alors u est bijective.

* S'il existe une application v de F vers E telle que $v \circ u = Id_E$, alors u est bijective.

En d'autres termes : le critère de bijectivité par existence d'un inverse peut dans ce cas particulier n'être vérifié que dans un seul sens.

Démonstration :

* Supposons qu'il existe v de F vers E telle que $u \circ v = Id_F$. Soit alors un $y \in F$; comme $y = u[v(y)]$, y est image de quelqu'un par u. Ceci prouve que $\operatorname{Im} u = F$, donc que u est surjective ; par le théorème précédent elle est donc bijective.

* Supposons qu'il existe v de F vers E telle que $v \circ u = Id_E$. Soit x un élément de $\operatorname{Ker} u$. Alors $x = v[u(x)] = v(0) = 0$. Donc $\operatorname{Ker} u = \{0\}$, donc u est injective, donc, par le théorème précédent, u est bijective.

•

Chapitre 9 - Les deux formules de Taylor

Ce chapitre contient deux théorèmes bien distincts et à ne pas confondre. Le premier d'entre eux généralise la formule des accroissements finis, le second généralise la définition de dérivée.

1 - Un peu de vocabulaire

Les concepts définis ci-dessous sont bien connus de tous, et il est vivement recommandé de ne pas regarder avec trop d'attention les définitions ci-dessous, qui m'ont demandé la plus grande attention (toujours ces problèmes de restrictions...) mais n'en méritent pas autant de votre part. (N'oubliez tout de même pas de connaître la terminologie "**de classe $\mathcal{C}^n$**").

Définition 9-1-86 : (présentée sous forme récursive, qui mèle cette définition, la suivante et la notation qui les suit) Soit $n \geq 2$ un entier. Soit f une fonction réelle d'une variable réelle, définie sur un ensemble $\mathcal{D}_f$ et soit a un point de $\mathcal{D}_f$. On dit que f est n **fois dérivable en** a lorsqu'il existe un intervalle ouvert I contenant a tel que f soit $n-1$ fois dérivable en tous les points de $I \cap \mathcal{D}_f$, et que $f^{(n-1)}$ est elle-même dérivable en a. (On convient pour initier les récurrences que"1 fois dérivable" est synonyme de "dérivable", et, si on aime les cas dégénérés, on pourra même convenir que "0 fois dérivable" s'applique à n'importe quelle fonction en n'importe quel point et appliquer cette définition dès $n = 1$.).

Définition 9-1-87 : Soit $n \geq 2$ un entier. Soit f une fonction réelle d'une variable réelle, définie sur un ensemble $\mathcal{D}_f$ et soit $\mathcal{E}_n$ l'ensemble des points où f est n fois dérivable. L'application de $\mathcal{E}_n$ vers $\mathbf{R}$ qui associe à un réel a la valeur en a de la dérivée de $f^{(n-1)}$ est appelée la **dérivée n-ème de f**. (On convient pour initier les récurrences que la "dérivée 1-ème de f" est f', et même, si on aime les cas dégénérés, que sa dérivée 0-ème est elle-même).

Notation 9-1-41 : La dérivée n-ème d'une fonction f est notée $f^{(n)}$. (Pour les petites valeurs de n on peut utiliser des apostrophes : f'' pour $f^{(2)}$, etc...)

Définition 9-1-88 : On dit qu'une fonction réelle d'une variable réelle est n **fois dérivable** sur son ensemble de définition (ou sur un intervalle ouvert contenu dans celui-ci) lorsqu'elle est n fois dérivable en tout point de cet ensemble (ou intervalle ouvert).

Définition 9-1-89 : On dit qu'une fonction réelle d'une variable réelle est n **fois continûment dérivable** sur son ensemble de définition (ou sur un intervalle ouvert contenu dans celui-ci) (ou "**de classe $\mathcal{C}^n$**") lorsqu'elle y est n fois dérivable et que sa dérivée n-ème est continue. On complète cette définition en définissant "**de classe $\mathcal{C}^0$**" comme synonyme de "continu" et "**de classe $\mathcal{C}^\infty$**" comme signifiant "de classe $\mathcal{C}^n$ pour tout $n \geq 0$".

2 - Le théorème de Taylor-Lagrange

Le lemme qui suit n'est pas à retenir. Il est destiné à mettre au maximum en relief l'analogie entre le théorème des accroissements finis et le théorème de Taylor-Lagrange. Le plan de la preuve est le même : le lemme est une variante améliorée du théorème de Rolle, puis on passe du lemme au théorème par utilisation d'une fonction auxiliaire pas trop compliquée.

Lemme 9-2-3 : Soit f une fonction réelle définie sur le segment $[a, b]$ ($a \neq b$) et soit $n \geq 0$ un entier. On suppose f de classe $\mathcal{C}^n$ sur le segment $[a, b]$ et $n+1$ fois dérivable sur le segment $]a, b[$.

On suppose en outre que $f(a) = f(b)$ et que $f'(a) = \ldots = f^{(n)}(a) = 0$. Alors il existe un c strictement entre a et b tel que $f^{(n+1)}(c) = 0$.

Démonstration : C'est une récurrence sur l'entier n.

* Cas où $n = 0$. La condition énumérative "$f'(a) = \ldots = f^{(n)}(a) = 0$" est alors une condition vide, et l'énoncé est exactement celui du théorème de Rolle. Le résultat est donc déjà connu.

* Soit un $n \geq 1$ fixé ; supposons le lemme vrai pour la valeur $n-1$ et montrons le pour la valeur n. Soit f comme dans l'énoncé. On peut dans un premier temps lui appliquer le théorème de Rolle, obtenant ainsi un point c_1 strictement compris entre a et b tel que $f'(c_1) = 0$. On remarque alors que $f'(a) = f'(c_1)$, ce qui invite à appliquer le lemme à l'ordre $n-1$ à la fonction f' sur le segment $[a, c_1]$ (si on est pointilleux, on dira "à la restriction de f' au segment $[a, c_1]$"). Cette fonction est en effet de classe $\mathcal{C}^{n-1}$ sur le segment fermé

$[a, c_1]$, et n fois dérivable sur le segment ouvert $]a, c_1[$. Il existe donc un point c strictement compris entre a et c_1 (et *a fortiori* strictement compris entre a et b) tel que $f'^{(n)}(c) = 0$, soit $f^{(n+1)}(c) = 0$.

•

Théorème 9-2-15 : Soit f une fonction réelle définie sur le segment $[a, b]$ ($a \neq b$) et soit $n \geq 0$ un entier. On suppose f de classe $\mathcal{C}^n$ sur le segment $[a, b]$ et $n+1$ fois dérivable sur le segment $]a, b[$.

Alors il existe un c strictement entre a et b tel que

$$f(b) = f(a) + f'(a)\frac{(b-a)}{1!} + f''(a)\frac{(b-a)^2}{2!} + \cdots + f^{(n)}(a)\frac{(b-a)^n}{n!} + f^{(n+1)}(c)\frac{(b-a)^{n+1}}{(n+1)!}.$$

Démonstration : Le principe est le même que pour le théorème des accroissements finis : la fonction f ne vérifie *a priori* pas toutes les égalités $f(a) = f(b)$ ni $f'(a) = \ldots = f^{(n)}(a) = 0$; on la remplace par une fonction vérifiant toutes ces égalités ; on applique le "super-Rolle" qui précède, et on conclut.

On pourrait parachuter d'un seul coup une fonction auxiliaire qui marche. Cela donne une démonstration vérifiable mais un peu mystérieuse. En décomposant la difficulté en deux morceaux, on verra —je l'espère— un peu mieux comment s'organisent les calculs.

Dans un premier temps, montrons le théorème de Taylor-Lagrange pour les fonctions g vérifiant l'hypothèse supplémentaire (évidemment des plus restrictives !) : $g'(a) = g''(a) = \cdots = g^{(n)}(a) = 0$. Pour pouvoir appliquer "super-Rolle" il manque seulement l'égalité des valeurs prises par g en a et en b.

Pour obtenir cette égalité, on va modifier g par l'addition d'une fonction de la forme $\lambda(t-a)^{n+1}$; une telle expression a en effet le bon goût d'avoir des dérivées nulles en a jusqu'à la n-ème incluse, donc de ne pas perturber ce qui fonctionnait bien chez g tout en prenant des valeurs différentes en a et en b, donc en pouvant amender la tare originelle de g.

Posons, conformément à ce programme :

$$g_1(t) = g(t) + \frac{(t-a)^{n+1}}{(b-a)^{n+1}}\left(g(a) - g(b)\right).$$

Ainsi, pour tout t du segment fermé $[a, b]$:

$$g_1'(t) = g'(t) + (n+1)\frac{(t-a)^n}{(b-a)^{n+1}}\left(g(a) - g(b)\right)$$

donc $g_1'(a) = g'(a) + 0 = 0$.

Puis, pour tout t du segment fermé $[a, b]$:

$$g_1''(t) = g''(t) + n(n+1)\frac{(t-a)^{n-1}}{(b-a)^{n+1}}\left(g(a) - g(b)\right)$$

donc on garde bien encore la bonne propriété $g_1''(a) = 0$.

Tout continue pour le mieux jusqu'à la dérivée n-ème de g, encore nulle en a. Sur notre élan, nous pouvons même calculer (mais pour les seuls t du segment ouvert $]a, b[$) :

$$g_1^{(n+1)}(t) = g^{(n+1)}(t) + (n+1)!\frac{1}{(b-a)^{n+1}}\left(g(a) - g(b)\right).$$

La nullité des n premières dérivées de g_1 au point a conjointement avec $g_1(a) = g_1(b)$ nous permet d'appliquer le lemme de "super-Rolle" ; on obtient donc un c strictement entre a et b tel que

$$g_1^{(n+1)}(c) = 0$$

c'est-à-dire

$$g^{(n+1)}(c) + (n+1)!\frac{1}{(b-a)^{n+1}}\left(g(a) - g(b)\right) = 0$$

ou encore, en regroupant le tout différemment :

$$g(b) = g(a) + \frac{g^{(n+1)}(c)(b-a)^{n+1}}{(n+1)!}$$

qui est bien la formule de Taylor-Lagrange pour g (n'oublions pas que les dérivées de g en a sont nulles jusqu'à la n-ème...)

Il nous reste à montrer la formule pour la fonction f de l'énoncé dont aucune dérivée ne s'annule *a priori* en a. On va la modifier en ajoutant des facteurs polynomiaux de degré compris entre 1 et n dont la finalité est de modifier les dérivées au point a.

Posons donc, conformément à ce programme :

$$g(t) = f(t) - f'(a)(t-a) - f''(a)\frac{(t-a)^2}{2!} - \cdots - f^{(n)}(a)\frac{(t-a)^n}{n!}.$$

On en déduit, pour tout t du segment fermé $[a,b]$:

$$g'(t) = f'(t) - f'(a) - f''(a)(t-a) - \cdots - f^{(n)}(a)\frac{(t-a)^{n-1}}{(n-1)!}$$

et en particulier $g'(a) = f'(a) - f'(a) - 0 - \cdots - 0 = 0$.

On continue ainsi jusqu'à la dérivée n-ème et on pousse le calcul jusqu'à la dérivée $n+1$-ème (ceci n'étant valide que pour t dans le segment ouvert $]a,b[$:

$$g^{(n+1)}(t) = f^{n+1}(t).$$

On peut alors appliquer le théorème de Taylor-Lagrange à g, qui vérifie l'hypothèse restrictive sous laquelle il est déjà connu.

On obtient l'existence d'un c tel que :

$$g(b) = g(a) + \frac{g^{(n+1)}(c)(b-a)^{n+1}}{(n+1)!}$$

soit, en allant repêcher l'expression de g d'une part, l'expression de $g^{(n+1)}$ d'autre part :

$$f(b) - f'(a)(b-a) - f''(a)\frac{(b-a)^2}{2!} - \cdots - f^{(n)}(a)\frac{(b-a)^n}{n!} = f(a) - 0 - \cdots - 0 + \frac{f^{(n+1)}(c)(b-a)^{n+1}}{(n+1)!}.$$

•

3 - Le théorème de Taylor-Young

Ce théorème n'utilise qu'un seul point a, et donne une information précieuse sur le comportement quand t tend vers a d'une fonction de la variable réelle t supposée n fois dérivable au point a quand t tend vers a $(t \neq a)$.

Il est intéressant de partir de la remarque suivante, qui n'est qu'une conséquence de la définition même de dérivée :

Remarque : Soit f une fonction réelle d'une variable réelle définie sur un ensemble $\mathcal{D}_f$ et soit a un point de $\mathcal{D}_f$. On suppose que $f'(a) = 0$. Alors :

$$\frac{f(t) - f(a)}{t-a} \to 0 \text{ quand } t \to a \ (t \neq a).$$

Il n'y a rien à prouver, c'est la définition même de la dérivée !

Le théorème de Taylor-Young n'est guère qu'une reformulation du lemme qui suit, qui généralise la remarque qui précède :

Lemme 9-3-4 : Soit f une fonction réelle d'une variable réelle définie sur un intervalle I et soit a un point de I ; soit $n \geq 1$ un entier. On suppose que $f'(a) = \cdots = f^{(n)}(a) = 0$. Alors :

$$\frac{f(t)-f(a)}{(t-a)^n} \to 0 \text{ quand } t \to a \ (t \neq a).$$

Démonstration : C'est, sans surprise, une récurrence sur l'entier n.
* Cas où $n = 1$: c'est la remarque qui précède, il n'y a rien à prouver !
* Soit un $n \geq 2$ fixé ; supposons le lemme vrai pour la valeur $n-1$ et montrons le pour la valeur n.

Soit donc une fonction f qui ait ses n premières dérivées nulles en a. On peut appliquer l'hypothèse de récurrence à la fonction f' qui a ses $n-1$ premières dérivées nulles en a et obtenir :

$$\frac{f'(t)}{(t-a)^{n-1}} \to 0 \text{ quand } t \to a \ (t \neq a)$$

(la formule ne contient pas de terme $f'(a)$ puisqu'on a supposé $f'(a) = 0$).

Pour $t \in I$ $(t \neq a)$, commençons alors à examiner le quotient $\dfrac{f(t)-f(a)}{(t-a)^n} = \dfrac{f(t)-f(a)}{t-a}\dfrac{1}{(t-a)^{n-1}}$.

Comme on a supposé $n \leq 2$, $f''(a)$ existe, donc il existe un intervalle ouvert J contenant a tel que f' existe sur $I \cap J$; en utilisant plus ou moins implitement la proposition 5-4-29, on travaillera pour t dans cet intervalle $I \cap J$. Dès lors que l'on prend t dans cet intervalle $(t \neq a)$, f est dérivable (donc continue) sur tout le segment fermé $[a, t]$. On peut donc lui appliquer le théorème des accroissements finis, et trouver un c_t strictement entre a et t tel que $\dfrac{f(t)-f(a)}{t-a} = f'(c_t)$, donc

$$\frac{f(t)-f(a)}{(t-a)^n} = f'(c_t)\frac{1}{(t-a)^{n-1}} = \frac{f'(c_t)}{(c_t-a)^{n-1}}\frac{(c_t-a)^{n-1}}{(t-a)^{n-1}}.$$

Remarquons maintenant que c_t est plus proche de a que t, ou, dit avec des formules, que $|c_t - a| \leq |t-a|$. Le quotient $\dfrac{(c_t-a)^{n-1}}{(t-a)^{n-1}}$ a donc une valeur absolue plus petite que 1 et on obtient la majoration :

$$0 \leq \left|\frac{f(t)-f(a)}{(t-a)^n}\right| \leq \left|\frac{f'(c_t)}{(c_t-a)^{n-1}}\right|.$$

Or $c_t \to a$ quand $t \to a$ $(t \neq a)$, et $\dfrac{f'(t)}{(t-a)^{n-1}} \to 0$ quand $t \to a$ $(t \neq a)$. Par composition des limites, on en déduit donc que $\dfrac{f'(c_t)}{(c_t-a)^{n-1}} \to 0$ quand $t \to a$ $(t \neq a)$. En appliquant alors le principe des gendarmes, on conclut que :

$$\left|\frac{f(t)-f(a)}{(t-a)^n}\right| \to 0 \text{ quand } t \to a \ (t \neq a).$$

•

En utilisant une fonction auxiliaire, on obtient, lorsqu'on supprime les hypothèses simplificatrices d'annulation de dérivées le

Théorème 9-3-16 : Soit f une fonction réelle d'une variable réelle définie sur un intervalle I et soit a un point de I ; soit $n \geq 1$ un entier. On suppose que f est (au moins) n fois dérivable au point a. Alors :

$$\frac{f(t)-f(a)-f'(a)(t-a)-f''(a)\dfrac{(t-a)^2}{2!}-\cdots-f^{(n)}(a)\dfrac{(t-a)^n}{n!}}{(t-a)^n} \to 0 \text{ quand } t \to a$$

Démonstration :

La bonne fonction auxiliaire est la même que celle utilisée vers la fin de la preuve de Taylor-Lagrange : on introduira

$$g(t) = f(t) - f'(a)(t-a) - f''(a)\frac{(t-a)^2}{2!} - \cdots - f^{(n)}(a)\frac{(t-a)^n}{n!}.$$

En récupérant les calculs faits plus haut, qui montrent que $g'(a) = \cdots = g^{(n)}(a) = 0$, on peut appliquer le lemme à g et le théorème tombe alors aussitôt.

•

Chapitre 10 - Équivalents

La notion de fonctions équivalentes est un outil simple d'une grande efficacité pour calculer des limites.

De plus la notion a un intérêt en tant que telle : savoir qu'une fonction f est équivalente à n donne n^3 quand n tend vers l'infini, cela donne en pratique une idée de l'ordre de grandeur de $f(1000000)$ (en pratique et non en théorie, d'ailleurs, car d'un point de vue théorique, 1000000 n'a rien de particulier et le comportement de f en ce point pourrait n'avoir rien de commun avec son comportement à l'infini !)

1 - La définition

Expliciter une définition correcte se révèle très désagréable : des problèmes se posent dès que les deux fonctions envisagées peuvent s'annuler, empêchant de faire la division qu'on souhaiterait.

De ce fait, la définition précise (et assez arbitraire) que je donne ne mérite pas d'être considérée longuement : elle sera exceptionnellement doublée d'une "définition approximative" qui me semble être celle qui doit être retenue.

Définition 10-1-90 : Soit a un nombre réel ; soit $\mathcal{D}$ une partie de $\mathbf{R}$ à laquelle a est adhérent, et soit f, g deux fonctions à valeurs réelles définies sur $\mathcal{D}$. On dit que f est **équivalente** à g quand $t \to a$ lorsqu'il existe un réel $\epsilon > 0$ et une fonction h de $[a-\epsilon, a+\epsilon] \cap \mathcal{D}$ vers $\mathbf{R}$ telle que pour t dans cet intervalle, $f(t) = h(t)g(t)$ et que $h(t)$ tende vers 1 quand $t \to a$.

Notation 10-1-42 : Lorque f est équivalente à g quand $t \to a$, on note "$f \sim g$ quand $t \to a$" (ou en abrégé $f \sim_a g$).

Remarques : * Il était difficile d'admettre que f et g aient des ensembles de définition distincts sans inconvénients ; de ce fait, quand on écrira : $\tan x \sim x$ quand $x \to 0$, il faudra bien sûr comprendre que la deuxième fonction mentionnée est la restriction de $x \mapsto x$ à l'ensemble de définition de la fonction tangente.

* Une autre définition est nécessaire pour le cas des équivalents à l'infini. Je ne lui fais pas l'honneur de la numéroter et me contente d'indiquer ce qui doit être modifié dans la définition précédente : en $+\infty$ on remplacera l'hypothèse "a adhérent à $\mathcal{D}$" par "$\mathcal{D}$ non majoré" et le passage qui parle d'ϵ par "il existe un réel A et une fonction h de $[A, +\infty[$ vers $\mathbf{R}$". Tous les résultats énoncés ci-dessous pour un a réel se transposent sans modifications à l'infini.

Comme promis, voici une version approximative, et utilisable en pratique, de la définition.

Version à retenir de la définition (fausse, mais qu'importe) : soit $\mathcal{D}$ une partie de $\mathbf{R}$ à laquelle a est adhérent, et soit f, g deux fonctions à valeurs réelles définies sur $\mathcal{D}$. On dit que f est **équivalente** à g quand $t \to a$ lorque $\dfrac{f}{g}(t) \to 1$ quand $t \to a$.

2 - Produire des limites à partir des équivalents

Proposition 10-2-54 : Soit $\mathcal{D}$ une partie de $\mathbf{R}$ et a un réel adhérent à $\mathcal{D}$; soit f une fonction de $\mathcal{D}$ vers $\mathbf{R}$. Alors pour toute constante c non nulle :

$$f(t) \sim c \qquad \text{quand } t \to a \iff f(t) \to c \qquad \text{quand } t \to a.$$

De plus, une fonction équivalente à une fonction qui tend vers 0 tend elle aussi vers 0 et une fonction équivalente à une fonction qui tend vers $+\infty$ tend aussi vers $+\infty$.

Démonstration : Tapant ce chapitre à la dernière minute, j'ai une tendance excessive à les considérer comme très faciles et les sauter.

•

3 - Propriétés élémentaires des équivalents

Proposition 10-3-55 : Comme son nom l'indique, pour $\mathcal{D}$ et a fixés, $\sim_a$ est une relation d'équivalence sur l'ensemble des fonctions de $\mathcal{D}$ vers $\mathbf{R}$.

Démonstration : Ennuyeuse comme la pluie, évidente avec la définition truquée et à peine plus longue avec la définition correcte...

•

Proposition 10-3-56 : Soit $\mathcal{D}$ une partie de $\mathbf{R}$ et a un réel adhérent à $\mathcal{D}$. Soit f, g, f_1 et g_1 des fonctions de $\mathcal{D}$ vers $\mathbf{R}$.
On suppose que $f \sim_a g$ et $f_1 \sim_a g_1$. Alors : a) $ff_1 \sim_a gg_1$; b) $\dfrac{1}{f} \sim_a \dfrac{1}{g}$;
c) Soit α un réel fixé, on suppose f à valeurs strictement positives sur $\mathcal{D}$. Alors, quitte à restreindre les ensembles de définitions, g est aussi à valeurs strictement positives et $f^\alpha \sim_a g^\alpha$;
d) Soit s_0 un réel, $\mathcal{D}_u$ une partie de $\mathbf{R}$ à laquelle s_0 est adhérent et u une fonction définie sur $\mathcal{D}_u$ et à valeurs dans $\mathcal{D}$ telle que $u(s) \to a$ quand $s \to s_0$. Alors $f[u(s)] \sim g[u(s)]$ quand $s \to s_0$.
Démonstration : Toujours facile et ennuyeux... •

Remarques : * du a) et du b) découle évidemment la possibilité de diviser les équivalents.
* le c) est un peu désagréablement exprimé, avec son "quitte à restreindre"... mais j'assume et n'éclaire pas davantage ce que ça veut dire.

Plutôt que d'écrire des démonstrations ennuyeuses, je préfère insister sur les points qui **ne marchent pas** :
* Les équivalents ne **s'additionnent pas** (et bien sûr ne se soustrayent pas).
* En utilisant le c), ne perdez pas de vue qu'il concerne un α réel (et donc constant) et qu'il ne **marche pas** pour une fonction $\alpha(t)$ à valeurs réelles : il se peut que $f(t) \sim_a g(t)$ mais que $[f(t)]^{\alpha(t)} \not\sim [g(t)]^{\alpha(t)}$.
* La composition **ne marche que dans un sens** (celui où les fonctions équivalentes sont "à gauche" dans la formule composée). Tout de suite un contre-exemple pour bien faire rentrer dans vos petites têtes le problème :
quand $x \to +\infty$, il est clair que $x^2 + x \sim x^2$, puisque $\dfrac{x^2+x}{x^2} = 1 + \dfrac{1}{x} \to 1$ quand $x \to +\infty$. Pourtant : $\dfrac{e^{x^2+x}}{e^{x^2}} = e^x$ ne tend pas vers 1 quand x to ∞ et donc $e^{x^2+x} \not\sim e^{x^2}$ en $+\infty$.
Les compositions avec l'exponentielle sont le piège le plus courant avec ce type de compositions, mais ce n'est pas le seul !
* Les équivalents **ne se laissent pas dériver** : si $f \sim_a g$ pour deux fonctions dérivables, rien n'assure que $f' \sim_a g'$.

4 - Un exemple d'utilisation de tout ce qui précède

Listons quelques équivalents classiques, qui découleront du chapitre suivant :
quand $x \to 0$, $\sin x \sim x$, $\operatorname{ch} x - 1 \sim x^2/2$, $\ln(1+x) \sim x$
et posons un
Exercice : prouver l'existence de la limite suivante, et la calculer : $\displaystyle\lim_{\substack{x\to 0\\ x<0}} \frac{x^2\sqrt{\operatorname{ch} x - 1}}{\sin(\tan^2 x)\ln(1+x)}$.
Solution : La question qui m'est posée possède une superbe barre de fractions qui la scinde en un haut et un bas. Les équivalents passant bien aux divisions, ceci invite à traiter séparément le haut et le bas.

Regardons le haut, soit $x^2\sqrt{\operatorname{ch} x - 1}$. C'est un produit : les équivalents se prêtent donc bien à son calcul. Quand $x \to 0$, on sait que $\operatorname{ch} x - 1 \sim x^2/2$. Donc $(\operatorname{ch} x - 1)^{1/2} \sim (x^2/2)^{1/2}$, c'est-à-dire (pour des $x < 0$) : $\sqrt{\operatorname{ch} x - 1} \sim -x/\sqrt{2}$. En multipliant les équivalents, on a donc montré que le numérateur $x^2\sqrt{\operatorname{ch} x - 1}$ est équivalent à $-x^3/\sqrt{2}$.

Regardons maintenant le bas, soit $\sin(\tan^2 x)\ln(1+x)$. On sait que quand $x \to 0$, $\ln(1+x) \sim x$; le premier morceau $\sin(\tan^2 x)$ reste à examiner. En utilisant la règle de composition dans le sens qui marche, et sans oublier de souligner préalablement qu'on peut légitimement l'utiliser parce que $\tan^2 x \to 0$ quand $x \to 0$, on voit d'abord que $\sin(\tan^2 x) \sim \tan^2 x$ quand $x \to 0$ (on peut l'exprimer si on trouve cela plus clair en posant $T = \tan^2 x$: puisque $T \to 0$, on a bien $\sin T \sim T$ quand $x \to 0$). Pour trouver un équivalent de tan, on remarque que comme $\cos x \to 1$ quand $x \to 0$, $\cos x \sim 1$ et donc $\tan x \sim x/1 = x$. En multipliant les équivalents, on a donc montré que le dénominateur , à savoir $\sin(\tan^2 x)\ln(1+x)$ est équivalent à x^3.

En divisant les équivalents, l'expression à étudier est donc équivalente à $-\dfrac{x^3}{\sqrt{2}}/x^3 = -\dfrac{1}{\sqrt{2}}$ quand x tend vers 0^-. Elle tend donc vers la constante $-\dfrac{1}{\sqrt{2}}$ quand x tend vers 0^-.

Chapitre 11 - Développements limités

Il s'agit de pallier à deux défauts des équivalents : le mauvais comportement vis-à-vis des additions et de la composition. Le but reste de déterminer des limites, ou peut-être des équivalents.

Les techniques de ce chapitre ont toutefois d'autres utilités indirectes : notamment elles nous permettront de calculer relativement facilement la dérivée 7-ème d'une fonction en un seul point sans avoir à dériver formellement sept fois une affreuse expression.

1 - Fonctions négligeables

Cette section ressemble étrangement à la définition des équivalents (aveu, j'ai copié-collé massivement) : les difficultés techniques sont encore sérieuses, une "définition simplifiée" nous suffira.

Définition 11-1-91 : Soit a un nombre réel ; soit $\mathcal{D}$ une partie de $\mathbf{R}$ à laquelle a est adhérent, et soit f, g deux fonctions à valeurs réelles définies sur $\mathcal{D}$. On dit que f est **négligeable** devant g quand $t \to a$ lorsqu'il existe un réel $\epsilon > 0$ et une fonction h de $[a - \epsilon, a + \epsilon] \cap \mathcal{D}$ vers $\mathbf{R}$ telle que pour t dans cet intervalle, $f(t) = h(t)g(t)$ et que $h(t)$ tende vers 0 quand $t \to a$.

Notation 11-1-43 : Lorque f est négligeable devant g quand $t \to a$, on note "$f \ll g$ quand $t \to a$" (ou en abrégé $f \ll_a g$). Cette notation sera abandonnée dans quelques lignes pour être remplacée par la très ésotérique (mais si pratique !) notation de Landau.

Remarques : * Quand les deux fonctions n'ont pas le même ensemble de définition, on restreint implicitement celle qui a le plus gros ensemble de départ.

* Une autre définition est nécessaire pour le cas de l'infini, exactement comme avec les équivalents.

Comme promis, voici une version approximative, et utilisable en pratique, de la définition.

Version à retenir de la définition (fausse, mais qu'importe) : soit $\mathcal{D}$ une partie de $\mathbf{R}$ à laquelle a est adhérent, et soit f, g deux fonctions à valeurs réelles définies sur $\mathcal{D}$. On dit que f est **négligeable** devant g quand $t \to a$ lorque $\dfrac{f}{g}(t) \to 0$ quand $t \to a$.

2 - La notation de Landau

Notation 11-2-44 : Lorsque f est négligeable devant g quand $t \to a$, on note :

$$f = o(g).$$

Il faut prendre garde que cette curieuse notation est un "faux" signe $=$: il lui manque un certain nombre de propriétés de l'égalité pour être utilisable comme elle.

Tout d'abord elle n'est pas réversible : ainsi quand $x \to 0$, $x^3 = o(x^2)$ et $x^5 = o(x^2)$ mais il serait bien hardi d'en déduire que $x^3 = x^5$.

Les choses vont se compliquer, car bien qu'à la lettre on n'ait défini que la seule expression "$f = o(g)$" (un o est immédiatement précédé d'un signe "$=$") on ne va pas se priver de faire des calculs qui vont déborder de cette définition. Ainsi on osera écrire une expression comme : $o(x) - o(x)$. Mais ceci ne fait pas 0.

L'étudiant est invité à ne pas s'inquiéter : la pratique de ces étrangetés se prend vite. S'il est curieux de comprendre plus, on ne lui reprochera pas : il pourra alors lire les paragraphes suivants ; s'il n'est pas curieux, on ne lui reprochera pas non plus et il fera glisser au plus vite son regard jusqu'à la section suivante.

On peut interpréter ces notations de façon correcte en définissant $o(g)$ comme l'ensemble des fonctions négligeables devant g. Quand on écrit $f = o(g)$, c'est un abus de langage pour $f \in o(g)$. Dès lors que $=$ n'est qu'un $\in$ déguisé, on n'est plus surpris qu'il ne soit pas réversible. On ajoutera que, par abus de langage classique, la notation $f(x)$ devra souvent être comprise comme représentant en réalité la fonction f et non le réel $f(x)$.

Si on est plus exigeant, on voudra alors comprendre le sens exact des $x^4 + o(x^4)$ voire $o(x) - o(x)$ qu'on va voir si souvent écrits. Pour cela, il faut avoir défini ce que veut dire le signe $+$ entre deux ensembles de fonctions, et cette définition est simple : si A est un ensemble de fonctions et B un autre, $A + B$ est l'ensemble des $f+g$ où $f \in A$ et $g \in B$ — de même avec toutes les autres opérations courantes. Cela étant posé, on comprend enfin pourquoi, pour x tendant vers 0, $o(x) - o(x)$ ne fait pas 0 : $o(x)$ contient de nombreuses fonctions, par exemple x^3 et x^5, donc $o(x) - o(x)$ en contient d'encore plus nombreuses, par exemple $x^3 - x^3 = 0$ mais aussi $x^3 - x^5$ ou $x^5 - x^3$.

Une fois ces manipulations ensemblistes comprises, on notera sans peine que certaines égalités sont à lire comme des inclusions : quand on écrit par exemple

$$\sin x = x + o(x^2) = x + o(x) \qquad \text{quand } x \to 0,$$

le premier = est un $\in$ qui s'est camouflé, tandis que le second est un $\subset$ déguisé.

L'étudiant le plus exigeant se plaindra peut-être de voir écrites des expressions comme $o(x+o(x))$ que les explications précédentes ne suffisent pas à expliquer. On lui répondra très brièvement que en convenant que pour A ensemble de fonctions $o(A)$ peut être défini comme l'ensemble des fonctions f qui sont négligeables devant un au moins des éléments de A et que cette définition supplémentaire permet, me semble-t-il, de finir de donner un sens à tous les calculs qui suivront.

3 - Produire des équivalents à partir des petits o

Le résultat suivant est de démonstration vide, mais essentiel car il explique l'utilité principale des développements limités :

Proposition 11-3-57 : Soit $\mathcal{D}$ une partie de $\mathbf{R}$ et a un réel adhérent à $\mathcal{D}$; soit f et g deux fonctions de $\mathcal{D}$ vers $\mathbf{R}$. Alors :

$$f(t) \sim g(t) \qquad \text{quand } t \to a \iff f(t) = g(t) + o[g(t)] \qquad \text{quand } t \to a.$$

Démonstration : La proposition me semble si importante que j'écris cette preuve bien qu'elle soit ennuyeuse :

quand $t \to a$, $f(t) \sim g(t)$ signifie qu'il existe un $\epsilon > 0$ et une fonction h de $[a-\epsilon, a+\epsilon] \cap \mathcal{D}$ vers $\mathbf{R}$ telle que pour t dans cet intervalle, $f(t) = h(t)g(t)$ et que $h(t)$ tende vers 1 quand $t \to a$.

D'un autre côté, $f(t) = g(t) + o[g(t)]$ signifie que $f-g$ est négligeable devant g, c'est-à-dire qu'il existe un $\epsilon > 0$ et une fonction k de $[a-\epsilon, a+\epsilon] \cap \mathcal{D}$ vers $\mathbf{R}$ telle que pour t dans cet intervalle, $f(t) - g(t) = k(t)g(t)$ et que $k(t)$ tende vers 1 quand $t \to a$.

Pour passer de l'un à l'autre, il suffit ainsi de poser $h = k+1$ (ou $k = h-1$).

•

4 - Propriétés élémentaires des petits o

Proposition 11-4-58 : $\mathcal{D}$ une partie de $\mathbf{R}$ et a un réel adhérent à $\mathcal{D}$. Soit f, g deux fonctions de $\mathcal{D}$ vers $\mathbf{R}$. Alors :

a) $f \times o(g) = o(fg)$;
b) $o(f) \times o(g) = o(fg)$;
c) $o(f) + o(f) = o(f)$;
d) pour tout réel λ, $o(\lambda f) = o(f)$;
e) $o[o(f)] = o(f)$;
f) $o[f + o(f)] = o(f)$;
g) soit s_0 un réel, $\mathcal{D}_u$ une partie de $\mathbf{R}$ à laquelle s_0 est adhérent et u une fonction définie sur $\mathcal{D}_u$ et à valeurs dans $\mathcal{D}$ telle que $u(s) \to a$ quand $s \to s_0$. Alors

$o(f) \circ u = o(f \circ u)$

(où le o de gauche est un o quand $t \to a$ et celui de droite quand $s \to s_0$).

Démonstration : Simples vérifications toutes évidentes, qui nécessitent toutefois de comprendre ce que veulent exactement dire toutes les expressions manipulées. Comme j'ai autorisé à sauter la lecture des explications à leur sujet, la démonstration ne peut donc être lue par tous. Beau prétexte pour ne pas l'écrire.

•

Si je n'en ai pas oublié, ces sept formules sont les seules utilisées dans les calculs courants sur les petits o.

Tout va mieux que pour les équivalents : on sait faire quelque chose en cas d'addition, et le jeu des égalités diminue les chances de blocage en cas de composition.

On prendra toutefois garde à ce que **on ne peut pas dériver** une relation entre petits o : si $f = o(g)$, il se peut que f' ne soit pas $o(g')$.

5 - Réécriture de la formule de Taylor-Young sous forme mémorisable

Maintenant que les notations de Landau sont connues, le théorème de Taylor-Young se réécrit :
Réécriture du théorème de Taylor-Young
Théorème 11-5-17 : Soit f une fonction réelle d'une variable réelle définie sur un intervalle I et soit a un point de I ; soit $n \geq 1$ un entier. On suppose que f est (au moins) n fois dérivable au point a. Alors, quand $t \to a$:

$$f(t) = f(a) + f'(a)(t-a) + \frac{f''(a)}{2!}(t-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(t-a)^n + o[(t-a)^n].$$

6 - Développements limités des fonctions classiques

Le formulaire regroupé page suivante est à savoir ; les démonstrations des formules ont été faites en cours :

Les développements limités à connaître
1 - La famille d'exponentielle
Quand $x \to 0$,

$$\begin{aligned}
e^x &= 1 + x + \frac{x^2}{2} + o(x^2) = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + o(x^n) = \sum_{k=0}^{n} \frac{x^k}{k!} + o(x^n) \\
\sin x &= x - \frac{x^3}{6} + o(x^4) = x - \frac{x^3}{3!} + \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + o(x^{2n+2}) \\
&= \sum_{k=0}^{n} (-1)^k \frac{x^{2k+1}}{(2k+1)!} + o(x^{2n+2}) \\
\cos x &= 1 - \frac{x^2}{2} + o(x^3) = 1 - \frac{x^2}{2} + \frac{x^4}{24} + o(x^5) \\
&= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + o(x^{2n+1}) = \sum_{k=0}^{n} (-1)^k \frac{x^{2k}}{(2k)!} + o(x^{2n+1}) \\
\operatorname{sh} x &= x + \frac{x^3}{6} + o(x^4) = x + \frac{x^3}{3!} + \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + o(x^{2n+2}) \\
&= \sum_{k=0}^{n} \frac{x^{2k+1}}{(2k+1)!} + o(x^{2n+2}) \\
\operatorname{ch} x &= 1 + \frac{x^2}{2} + o(x^3) = 1 + \frac{x^2}{2} + \frac{x^4}{24} + o(x^5) \\
&= 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + o(x^{2n+1}) = \sum_{k=0}^{n} \frac{x^{2k}}{(2k)!} + o(x^{2n+1})
\end{aligned}$$

2 - Le binôme
Pour α réel CONSTANT (ne contenant pas x), quand $x \to 0$,

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}x^n + o(x^n)$$

Il peut être bon de connaître spécifiquement les conséquences suivantes :

$$\begin{aligned}
\frac{1}{1+x} &= 1 - x + x^2 - x^3 + \ldots + (-1)^n x^n + o(x^n) \\
\frac{1}{1-x} &= 1 + x + x^2 + x^3 + \ldots + x^n + o(x^n) \\
\sqrt{1+x} &= 1 + \frac{1}{2}x - \frac{1}{8}x^2 + o(x^2)
\end{aligned}$$

3 - Logarithme et arctangente
Quand $x \to 0$,

$$\begin{aligned}
\ln(1+x) &= x - \frac{x^2}{2} + \frac{x^3}{3} + \cdots + (-1)^{n+1} \frac{x^n}{n} + o(x^n) \\
&= \sum_{k=1}^{n} (-1)^{k+1} \frac{x^k}{k} + o(x^n) \\
\operatorname{Arctan} x &= x - \frac{x^3}{3} + \frac{x^5}{5} + \cdots + (-1)^n \frac{x^{2n+1}}{2n+1} + o(x^{2n+2}) \\
&= \sum_{k=1}^{n} (-1)^k \frac{x^{2k+1}}{2k+1} + o(x^{2n+2})
\end{aligned}$$

Chapitre 12 - Groupes

L'étude abstraite de ce genre de structures peut vous sembler fascinante ou épuisante selon votre personnalité. L'inconvénient (que je crains) inévitable est que l'utilité des résultats démontrés peut difficilement être mise en relief immédiatement, car il faut passer un certain temps dans la théorie, puis de nouveau un certain temps dans des chapitres plus concrets où les résultats accumulés pourront être recyclés.

Soyez rassurés (ou effrayés ?), une bonne part des résultats énoncés sur les groupes finis (concept d'ordre, théorème de Lagrange...) auront l'occasion d'être mis en application dès le prochain chapitre d'arithmétique. Une première utilité de la théorie des groupes étant de formaliser et systématiser les calculs usuels qu'on sait pratiquer sur les ensembles de nombres.

L'autre point de vue sur lequel j'insiste est celui des groupes formés de bijections, mais malheureusement vous aurez peu l'occasion de les voir vraiment appliqués dans la suite du cours de maths de Deug. En revanche, je ne serais pas surpris que des connaissances sur les groupes de permutations (groupes de bijections des ensembles finis) soient utiles de ci ou de là en informatique... Et de toutes façons l'investissement sera rentabilisé dès que vous apprendrez plus de géomérie, cadre idéal d'usage des groupes de transformations.

1 - Opérations ; morphismes

Définition 12-1-92 : On appelle **opération** sur un ensemble E une application de $E \times E$ vers E.

En fait, bien que cette définition soit générale, on n'aurait pas l'idée d'appeler "opérations" toutes les applications de $E \times E$ vers E ; le vocable n'est utilisé que quand il est naturel de noter l'application par un symbole opératoire. Des exemples typiques d'opérations seront l'addition de $\mathbf{R}^2$ vers $\mathbf{R}$ associant $x+y$ à (x, y) ; ou sur l'ensemble E^E des applications de E vers E l'opération $\circ$, qui associe l'application $g \circ f$ au couple d'applications (g, f). Pour des opérations abstraites, le symbole opératoire $*$ a été à la mode, je l'utiliserai occasionnellement surtout au début, mais me contenterai rapidement de la notation multiplicative ab pour l'élément obtenu en appliquant l'opération à (a, b).

Un peu de vocabulaire au sujet des opérations :

Définition 12-1-93 : Soit $*$ une opération sur un ensemble E. On dit que $*$ est **commutative** lorsque pour tous éléments a, b de E, $a * b = b * a$.

Définition 12-1-94 : Soit $*$ une opération sur un ensemble E. On dit que $*$ est **associative** lorsque pour tous éléments a, b, c de E, $(a * b) * c = a * (b * c)$.

Définition 12-1-95 : Soit $*$ une opération sur un ensemble E. On dit qu'un élément e de E est **élément neutre** pour $*$ lorsque pour tout élément a de E, $a * e = e * a = a$.

La cohérence de ce qui suit nécessite d'énoncer tout de suite la simplissime :

Proposition 12-1-59 : Une opération possède au plus un élément neutre.

Démonstration : Soit e_1 et e_2 deux éléments neutres pour une opération $*$. Comme e_2 est neutre, $e_1 * e_2 = e_1$ et comme e_1 est neutre, $e_1 * e_2 = e_2$. Donc $e_1 = e_2$.

•

On pourra donc parler de **l**'élément neutre, lorsqu'il en existe.

Définition 12-1-96 : Soit $*$ une opération sur un ensemble E admettant un élément neutre noté e et a un élément de E. On dit qu'un élément b de E est **symétrique** (ou **inverse**) de a lorsque $a * b = b * a = e$.

Là encore, glissons sans tarder une évidence :

Proposition 12-1-60 : Pour une opération associative possédant un élément neutre, chaque élément possède au plus un symétrique.

Démonstration : Soit $*$ une telle opération et e son neutre ; soit a un élément de E et soit b_1 et b_2 deux symétriques de a. Alors d'une part $(b_1 * a) * b_2 = e * b_2 = b_2$ et d'autre part $(b_1 * a) * b_2 = b_1 * (a * b_2) = b_1 * e = b_1$, d'où $b_1 = b_2$.

•

Les opérations nous intéressant étant en pratique associatives, on pourra donc faire plein usage de la

Notation 12-1-45 : Le symétrique d'un élément a sera noté a^{-1}.

Maintenant que nous savons manipuler une opération sur un seul ensemble, apprenons à évoluer d'un ensemble muni d'une opération vers un autre, grâce à la

Définition 12-1-97 : Soit E un ensemble muni d'une opération $*_E$ et F un ensemble muni d'une opération $*_F$. On dit qu'une application $f : E \to F$ est un **morphisme** lorsque pour tous éléments a, b de E, on a l'identité :

$$f(a *_E b) = f(a) *_F f(b).$$

Définition 12-1-98 : Un morphisme bijectif est appelé un **isomorphisme**.

Il me semble plus facile d'expliquer la notion d'isomorphisme que celle de morphisme plus général ; deux opérations sur deux ensembles fourniront des structures isomorphes lorsque ces deux opérations agissent de la même façon, seuls les noms des éléments changeant. Hum, ce n'est pas bien clair, donnons plutôt des **Exemples** :

* Considérons tout d'abord la bijection σ de l'ensemble $E = \{0, 1, 2, 3\}$ définie par $\sigma(0) = 1$, $\sigma(1) = 2$, $\sigma(2) = 3$ et $\sigma(3) = 0$.

Avec à peine un peu de bon sens (penser σ comme "faisant tourner" les quatre éléments de E) on voit sans guère de calculs que $\sigma \circ \sigma$ est la bijection τ de E définie par $\tau(0) = 2$, $\tau(1) = 3$, $\tau(2) = 0$, $\tau(3) = 1$, puis que $\sigma \circ \sigma \circ \sigma$ est la bijection ρ de E définie par $\rho(0) = 3$, $\rho(1) = 0$, $\rho(2) = 1$ et $\rho(3) = 2$, et enfin que $\sigma \circ \sigma \circ \sigma \circ \sigma$ tout simplement l'identité de E.

En utilisant la notation en puissances de σ, on peut alors très facilement calculer tous les produits deux à deux des bijections introduites ici ; par exemple $\rho \circ \tau = \sigma^3 \circ \sigma^2 = \sigma^5 = \sigma^4 \circ \sigma = Id_E \circ \sigma = \sigma$.

On considère alors l'ensemble $S = \{Id_E, \sigma, \tau, \rho\}$ et on voit que $\circ$ est une opération sur ce sous-ensemble de E^E, qui sera agréablement décrite par le tableau suivant :

$\circ$	Id	σ	τ	ρ
Id	Id	σ	τ	ρ
σ	σ	τ	ρ	Id
τ	τ	ρ	Id	σ
ρ	ρ	Id	σ	τ

Considérons maitenant l'ensemble des nombres complexes dont la puissance quatre vaut 1, c'est-à-dire l'ensemble $F = \{1, i, -1, -i\}$. Il est très facile de constater que la multiplication des complexes définit une opération sur F, et que la table de cette opération est donnée par :

$\times$	1	i	-1	$-i$
1	1	i	-1	$-i$
i	i	-1	$-i$	1
-1	-1	$-i$	1	i
$-i$	$-i$	1	i	-1

Visuellement, on retrouve la même table, seuls les noms des éléments ont changé. C'est signe qu'il y a un isomorphisme camouflé. On le détectera facilement ; c'est bien sûr l'application g de E vers F définie par :

$$g(Id) = 1 \qquad g(\sigma) = i \qquad g(\tau) = -1 \qquad g(\rho) = -i.$$

* Soit R l'ensemble des rotations de centre $(0, 0)$ dans le plan, et soit $\mathbf{U}$ le cercle-unité de $\mathbf{C}$ (c'est-à-dire l'ensemble des nombres complexes de module 1). Les opérations respectives envisagées sur R et sur $\mathbf{U}$ sont la composition des applications et la multiplication. On définit $f : R \to \mathbf{U}$ en envoyant la rotation d'angle θ sur le nombre $e^{i\theta}$. Il faut tout d'abord se soucier de vérifier que la définition n'est pas ambiguë, car elle n'est pas loin de l'être ! Une rotation peut en effet être caractérisée par plusieurs angles (tourner d'un quart de tour dans le sens trigonométrique, c'est aussi tourner de trois quarts de tour dans le sens des aiguilles d'une montre), mais deux angles distincts θ_1 et θ_2 correspondant à la même bijection diffèrent d'un multiple entier de 2π ; il existe donc $k \in \mathbf{Z}$ tel que $\theta_2 = \theta_1 + 2k\pi$. Les valeurs $e^{i\theta_1}$ et $e^{i\theta_2} = e^{i\theta_1 + 2ki\pi} = e^{\theta_1}(e^{2i\pi})^k = e^{\theta_1}$ sont donc égales, et l'application f est bien définie. Cette vérification une fois faite, vérifier que f est un

morphisme est sans problème : si ρ_1 est la rotation d'angle θ_1 et ρ_2 la rotation d'angle θ_2, la composée $\rho_2 \circ \rho_1$ est la rotation ρ d'angle $\theta_2 + \theta_1$, et on a donc :

$$f(\rho_2 \circ \rho_1) = f(\rho) = e^{i(\theta_2+\theta_1)} = e^{i\theta_1} e^{i\theta_2} = f(\rho_1) f(\rho_2).$$

Montrer que f est bijective n'est pas difficile ; on en conclut donc que f est un isomorphisme, ou en d'autre termes que l'étude des nombres complexes de module 1 nous instruira sur le fonctionnement des rotations.

* Voici enfin un morphisme qui ne soit pas un isomorphisme —simple variante du précédent— considérons F de $\mathbf{R}$ (muni de l'addition) vers $\mathbf{U}$ (le même qu'à l'exemple précédent, muni de la multiplication) définie par $F(\theta) = e^{i\theta}$. C'est très facilement un morphisme, mais il est sans doute assez peu clair de voir quel lien il fait apparaître entre ses ensembles de départ et d'arrivée.

2 - Groupes

Définition 12-2-99 : Soit G un ensemble muni d'une opération $*$. On dit que G est un **groupe** lorsque les trois conditions suivantes sont réalisées :
(i) $*$ est associative.
(ii) $*$ possède un élément neutre.
(iii) Tout élément de G possède un symétrique pour $*$.
Définition 12-2-100 : Un groupe G est dit **abélien** (ou tout simplement commutatif !) lorsque son opération est commutative.

Avant de donner des exemples, quelques remarques d'ordre purement calculatoires sur les groupes :
Proposition 12-2-61 : Soit G un groupe noté multiplicativement. Alors pour tous éléments a, b, x de G :
(1) Si $ax = bx$, alors $a = b$.
(2) Si $xa = xb$, alors $a = b$.
(3) Le symétrique de ab est $b^{-1}a^{-1}$.
Démonstration :

Ce ne sont que simples vérifications à base d'associativité ; pour (1), si on suppose $ax = bx$, en multipliant à droite par x^{-1} on obtient $(ax)x^{-1} = (bx)x^{-1}$ et donc $a(xx^{-1}) = b(xx^{-1})$, c'est-à-dire $a = b$. On prouve (2) de la même façon en multipliant à gauche par x^{-1}. La preuve du (3) se réduit à un calcul élémentaire :

$$(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = aa^{-1} = e \qquad \text{et} \qquad (b^{-1}a^{-1})(ab) = b^{-1}(a^{-1}a)b = b^{-1}b = e.$$

•

Maintenant que vous savez calculer dans les groupes, il est temps de donner les exemples les plus élémentaires : regardons les opérations que nous connaissons le mieux, dans les ensembles de nombres bien connus.

Additions : elles sont associatives, ont un élément neutre noté 0. Dans $\mathbf{N}$, le symétrique peut faire défaut ; ainsi 2 n'a pas d'opposé. Dans $\mathbf{Z}$ (puis dans les ensembles usuels bien connus) l'opposé existe. Ainsi $\mathbf{Z}$ est un groupe pour l'addition.

Multiplication : 0 n'a jamais d'inverse, donc les ensembles de nombres bien connus ne sont jamais des groupes pour la multiplication. En revanche, si on considère le sous-ensemble formé des éléments non nuls, la multiplication y est bien définie, associative, possède un élément neutre noté 1. Le point à problème est l'existence du symétrique —de l'inverse en notation multiplicative. Dans $\mathbf{Z}^*$, il fait défaut à la plupart des éléments, ainsi 2 n'a pas d'inverse ; $\mathbf{Z}^*$ n'est donc pas un groupe. En revanche, dans $\mathbf{Q}^*$ (ensemble des fractions non nulles), ou $\mathbf{R}^*$, ou $\mathbf{C}^*$, l'existence de l'inverse ne pose pas de problème. Ce sont des groupes multiplicatifs.

Encore quelques propriétés de bon sens, mais qu'il ne coûte rien d'énoncer. Elles paraissent évidentes si on comprend qu'un morphisme est moralement une application qui transporte la structure ; si elle transporte l'opération, elle doit aussi transporter ses caractéristiques, telles l'élément neutre et le symétrique.
Proposition 12-2-62 : Soit f un morphisme d'un groupe G, d'élément neutre e, vers un groupe G', d'élément neutre e'.

Alors $f(e) = e'$ et, pour tout élément a de G, $[f(a)]^{-1} = f(a^{-1})$.

Démonstration : Essentiellement de la simple vérification ; pour le neutre une (petite) astuce : on calcule $f(e)f(e) = f(ee) = f(e) = f(e)e'$ puis on simplifie par $f(e)$. Pour l'inverse, calcul très simple : $f(a^{-1})f(a) = f(a^{-1}a) = f(e) = e'$ et simultanément, $f(a)f(a^{-1}) = f(aa^{-1}) = f(e) = e'$. Ceci montre bien que $f(a^{-1})$ est l'inverse de $f(a)$.

•

3 - L'exemple fondamental

Les groupes les plus directement utilisables sont sans doute ceux qui interviennent en géométrie. Ce sont des groupes de transformations "respectant" telle ou telle propriété ; ainsi les isométries, qui conservent les distances, ou les similitudes, qui conservent les angles.

Tous ces groupes ont le point commun d'avoir pour opération $\circ$, la composition des applications, et d'être formés de bijections.

Fondamentale —quoique très facile— sera donc la :

Proposition 12-3-63 : Soit E un ensemble. L'ensemble des bijections de E forme un groupe pour la composition.

Démonstration : Tout est très simple. Il est très simple de vérifier que la composée de deux bijections est une bijection (par exemple parce que $g^{-1} \circ f^{-1}$ se révèle un inverse de $f \circ g$) ; que la composition est associative ; que Id_E est neutre ; que pour f bijection de E, la bijection réciproque est symétrique de f. On a déjà fini !

•

Notation 12-3-46 : L'ensemble des bijections d'un ensemble E sera noté $\mathcal{S}(E)$.

(Je préfèrerais un S un peu plus gothique, mais je n'en trouve que de plutôt anglais dans ma police ; on s'en contentera, l'essentiel étant que ça ait l'air anglo-saxon).

On utilisera occasionnellement —et je ne serais pas étonné que vous voyiez intervenir en informatique— le cas particulier du groupe des bijections d'un ensemble fini. L'archétype d'un tel ensemble fini étant $\{1, \ldots, n\}$, cela justifie d'introduire une toute spéciale :

Notation 12-3-47 : L'ensemble des bijections de $\{1, \ldots, n\}$ sera noté $\mathcal{S}_n$.

Tentons de découvrir comment fonctionne $\mathcal{S}_n$ pour n pas trop gros ; il vaut mieux le prendre même franchement petit, car $\mathcal{S}_n$ possédant $n\,!$ éléments, on serait vite débordé.

Pour $n = 1$, le groupe n'a qu'un élément ; sa table est vite tracée :

$\circ$	e
e	e

Pour $n = 2$, il y a deux bijections de $\{1, 2\}$: celle qui échange les deux éléments, qu'on notera τ, et l'identité.

La table du groupe est donc :

$\circ$	e	τ
e	e	τ
τ	τ	e

Pour $n = 3$, les calculs complets seraient nettement plus fastidieux. Il est facile d'énumérer les éléments de $\mathcal{S}_3$: outre l'identité, il y en a trois d'apparence identique : l'un, que je noterai t, échange 1 et 2 en laissant 3 fixe ; un autre, que je me garderai astucieusement de noter, échange 2 et 3 en laissant 1 fixe ; le dernier échange 3 et 1 en laissant 2 fixe. Enfin deux autres jouent aussi des rôles voisins : l'un, que je noterai a, fait "tourner" les trois éléments de $\{1, 2, 3\}$ en envoyant 1 sur 2, 2 sur 3, et 3 sur 1 ; l'autre, dont je remarquerai que c'est le carré de a, les fait "tourner" dans l'autre sens.

On va remplir la table du groupe par ajouts successifs d'information. L'information la plus récente sera systématiquement portée en gras.

Au point où nous en sommes, il est facile de commencer en remarquant que $a^3 = e$ tandis que a^2, comme on l'a déjà dit, est distinct de a. En outre les trois autres éléments ont un carré égal à e.

$\circ$	$\mathbf{e}$	$\mathbf{a}$	$\mathbf{a^2}$	$\mathbf{t}$		
$\mathbf{e}$	$\mathbf{e}$	$\mathbf{a}$	$\mathbf{a^2}$	$\mathbf{t}$		
$\mathbf{a}$	$\mathbf{a}$	$\mathbf{a^2}$	$\mathbf{e}$			
$\mathbf{a^2}$	$\mathbf{a^2}$	$\mathbf{e}$	$\mathbf{a}$			
$\mathbf{t}$	$\mathbf{t}$			$\mathbf{e}$		
					$\mathbf{e}$	
						$\mathbf{e}$

Le produit at ne peut être présent deux fois dans la colonne a, ni deux fois dans la ligne t. Il est donc distinct des éléments qui y figurent déjà, c'est-à-dire de e, de a, de a^2 et de t. C'est donc un cinquième élément, qu'on peut alors faire figurer dans la cinquième ligne et la cinquième colonne du tableau. On calcule au passage sans mal $(a^2)(at) = (a^3)t = et = t$, et $(at)t = a(t^2) = ae = a$.

$\circ$	e	a	a^2	t	$\mathbf{at}$	
e	e	a	a^2	t	$\mathbf{at}$	
a	a	a^2	e	$\mathbf{at}$		
a^2	a^2	e	a		$\mathbf{t}$	
t	t			e		
$\mathbf{at}$	$\mathbf{at}$			$\mathbf{a}$	e	
						e

Puis à son tour, a^2t ne peut déjà figurer dans la ligne a^2 ni dans la colonne t : c'est donc le sixième élément. On peut l'ajouter au tableau en complétant par quelques calculs évidents.

$\circ$	e	a	a^2	t	at	$\mathbf{a^2t}$
e	e	a	a^2	t	at	$\mathbf{a^2t}$
a	a	a^2	e	at	$\mathbf{a^2t}$	$\mathbf{t}$
a^2	a^2	e	a	$\mathbf{a^2t}$	t	$\mathbf{at}$
t	t			e		
at	$\mathbf{at}$			a	e	
$\mathbf{a^2t}$	$\mathbf{a^2t}$			$\mathbf{a^2}$		e

En utilisant toujours l'astuce "il ne peut y avoir deux fois la même valeur dans une ligne" (ou une colonne), on arrive à calculer $(at)(a^2t)$ et $(a^2t)(at)$ par simple élimination de cinq valeurs impossibles :

$\circ$	e	a	a^2	t	at	a^2t
e	e	a	a^2	t	at	a^2t
a	a	a^2	e	at	a^2t	t
a^2	a^2	e	a	a^2t	t	at
t	t			e		
at	at			a	e	$\mathbf{a^2}$
a^2t	a^2t			a^2	$\mathbf{a}$	e

Surprise ! On vient de montrer avec une étonnante économie de calculs que le groupe n'est pas commutatif ; en effet $(at)(a^2t) \neq (a^2t)(at)$.

Le même truc des répétitions interdites permet de compléter le coin inférieur droit du tableau :

$\circ$	e	a	a^2	t	at	a^2t
e	e	a	a^2	t	at	a^2t
a	a	a^2	e	at	a^2t	t
a^2	a^2	e	a	a^2t	t	at
t	t			e	$\mathbf{a^2}$	$\mathbf{a}$
at	at			a	e	a^2
a^2t	a^2t			a^2	a	e

Dernier obstacle inattendu, alors que nous avions presque fini, avec la méthode maintenant bien rodée de remplir les cases par élimination, cette méthode est insuffisante pour remplir les six misérables cases laissées blanches ! Il faut une nouvelle astuce pour passer cet obstacle. Concentrons nous sur la case correspondant au produit ta. Pour calculer ce produit, bidouillons un peu : $ta = tae = ta(t^2) = [t(at)]t = a^2t$. Une nouvelle case est remplie :

$\circ$	e	a	a^2	t	at	a^2t
e	e	a	a^2	t	at	a^2t
a	a	a^2	e	at	a^2t	t
a^2	a^2	e	a	a^2t	t	at
t	t	$\mathbf{a^2t}$		e	a^2	a
at	at			a	e	a^2
a^2t	a^2t			a^2	a	e

Cette étape franchie, il est désormais très facile de finir de remplir la table en utilisant l'idée simple : pas plus d'une apparition par ligne ou par colonne :

$\circ$	e	a	a^2	t	at	a^2t
e	e	a	a^2	t	at	a^2t
a	a	a^2	e	at	a^2t	t
a^2	a^2	e	a	a^2t	t	at
t	t	a^2t	at	e	a^2	a
at	at	t	a^2t	a	e	a^2
a^2t	a^2t	at	t	a^2	a	e

4 - Sous-groupes

Maintenant que nous connaissons ce que j'ai pompeusement appelé "l'exemple fondamental" il reste à apprendre à tirer de cet exemple trop fondamental pour être vraiment utile des exemples plus concrets.

Pour cela, posons la :

Définition 12-4-101 : Soit G un groupe. On dit qu'un sous-ensemble H de G est un **sous-groupe** de G lorsque les trois conditions suivantes sont vérifiées :

(i) H n'est pas vide.

(ii) Pour tous a, b de H, le produit ab est aussi dans H.

(iii) Pour tout a de H, l'inverse a^{-1} de a est aussi dans H.

Avant de commenter ce que ça veut dire, je donne tout de suite une proposition très simple, et utile en pratique pour vérifier qu'un sous-ensemble d'un groupe est un sous-groupe.

Proposition 12-4-64 : Soit G un groupe. Un sous-ensemble H de G est un sous-groupe de G si et seulement si les deux conditions suivantes sont vérifiées :
(*1*) H n'est pas vide.
(*2*) Pour tous a, b de H, le produit ab^{-1} est aussi dans H.
Démonstration :
Supposons que H est un sous-groupe de G, c'est-à-dire qu'il vérifie (i), (ii) et (iii). Il est alors clair que (*1*) —qui coïncide avec (i) !— est vérifiée.

Montrons que H vérifie (*2*). Soit a, b deux éléments de H. En appliquant (iii) à b, on constate que b^{-1} est aussi dans H, puis en appliquant (ii) à a et b^{-1} que le produit ab^{-1} aussi. C'est déjà fini !

Supposons maintenant que H vérifie (*1*) et (*2*). Vérifier (i) est bien sûr sans problème.

Montrons préalablement que $e \in H$, où e désigne l'élément neutre de G. En effet H n'étant pas vide, on peut prendre un élément c dans H, puis appliquer l'hypothèse (*2*) à c et c pour conclure que $cc^{-1} = e \in H$.

Montrons maintenant que H vérifie (iii). Soit a un élément de H. Puisqu'on sait maintenant que e aussi est dans H, on peut appliquer (*2*) à e et a pour obtenir $ea^{-1} \in H$, c'est-à-dire $a^{-1} \in H$.

Montrons enfin que H vérifie (ii). Soit a, b deux éléments de H. Par la propriété (iii) appliquée à b, $b^{-1} \in H$, puis par la propriété (*2*) appliquée à a et b^{-1}, $a(b^{-1})^{-1} \in H$, c'est-à-dire $ab \in H$.

•

Bien que le résultat qui suive soit très simple à démontrer, son importance lui fait mériter à mes yeux l'appelation de :
Théorème 12-4-18 : Soit G un groupe et H un sous-groupe de G. La restriction à H de l'opération sur G fait de H un groupe.
Démonstration :
* Il ne faut pas manquer de vérifier la possibilité de restreindre l'opération initiale, application de $G \times G$ vers G à une opération sur H, c'est-à-dire une application de $H \times H$ vers H. Comme on veut restreindre non seulement l'ensemble de départ mais aussi l'ensemble d'arrivée, on est dans la situation où il faut spécialement prendre garde. Mais la propriété (ii) de la définition des "sous-groupes" assure précisément que l'opération de G envoie l'ensemble $H \times H$ dans H et que la restriction a donc bien un sens.
* L'associativité de cette restriction est alors évidente.
* Dans la preuve de la proposition précédente, on a montré au passage que le neutre de G était élément de H. Il est alors évidemment neutre pour l'opération restreinte à H.
* Enfin la propriété (iii) garantit l'existence d'un symétrique pour chaque élément de H.

•

Voyons maintenant comment ce théorème permet de fabriquer plein de groupes nouveaux et intéressants.
Exemple : Soit G le groupe des bijections strictement croissantes de $\mathbf{R}$ vers $\mathbf{R}$, muni de la composition. Montrer que G est un groupe.

(On rappellera, au cas où ce serait nécessaire, qu'une application f est dite strictement croissante lorsque pour tous x, y, $x < y$ entraîne $f(x) < f(y)$).

La bonne idée est de montrer que G est un sous-groupe du groupe $\mathcal{S}(\mathbf{R})$. Lançons nous. La vérification de (i) est évidente : il est clair que l'application identique est une bijection strictement croissante de $\mathbf{R}$ sur $\mathbf{R}$. Passons à (ii). Soit f et g deux bijections strictement croissantes de $\mathbf{R}$ sur $\mathbf{R}$. On sait déjà que $g \circ f$ est une bijection ; montrons qu'elle est strictement croissante. Soit x, y deux réels avec $x < y$; alors $f(x) < f(y)$ (croissance de f) puis $g[f(x)] < g[f(y)]$ (croissance de g). Ceci montre bien que $g \circ f$ est strictement croissante. Vérifions enfin (iii). Soit f une bijection strictement croissante de $\mathbf{R}$ vers $\mathbf{R}$. Il est bien clair que f^{-1} est bijective ; vérifions qu'elle est strictement croissante. Soit x, y deux réels avec $x < y$. On ne peut avoir $f^{-1}(x) = f^{-1}(y)$, car f^{-1} est injective ; on ne peut avoir $f^{-1}(y) < f^{-1}(x)$, car f étant strictement croissante on en déduirait $f[f^{-1}(y)] < f[f^{-1}(x)]$, ce qui est faux. Par élimination on a donc bien $f(x) < f(y)$.
Exemple : Soit A un sous-ensemble de $\mathbf{R}^2$ et G l'ensemble des isométries f de $\mathbf{R}^2$ sur $\mathbf{R}^2$ telles que $f(A) = A$. On montrerait par le même genre de méthode que G est un groupe parce que c'est un sous-groupe de $\mathcal{S}(\mathbf{R}^2)$. Pour A trop patatoïdal, G se réduira à $\{Id_{\mathbf{R}^2}\}$ et sera donc peu intéressant, mais si A possède des symétries raisonnables — par exemple si A est un pentagone régulier— le groupe G méritera notre attention (dans le cas du pentagone régulier, il a dix éléments, sauriez-vous les identifier ?)

5 - Un théorème de Lagrange

Il s'agit d'un résultat simple et élégant, surtout là pour le plaisir de faire une démonstration agréable.
Théorème 12-5-19 : Soit G un groupe fini et H un sous-groupe de G. Alors le nombre d'éléments de H divise le nombre d'éléments de G.
Démonstration : Elle repose sur l'introduction de la relation $\sim$ définie pour tous éléments a, b de G par :

$a \sim b \qquad$ lorsque $ab^{-1} \in H$.

Le plan de la preuve est le suivant :
(1) On vérifie que $\sim$, comme son nom le laisse penser, est une relation d'équivalence.
(2) On vérifie que toutes les classes d'équivalence pour la relation $\sim$ ont le même nombre d'éléments, à savoir le nombre d'éléments de H.
(3) On conclut en quelques mots.

Exécution...

(1) Vérifions successivement les trois propriétés requises des relations d'équivalence.

Soit a un élément de G. Comme $aa^{-1} = e \in H$, $a \sim a$. La relation $\sim$ est donc réflexive.

Soit a, b deux éléments de G, avec $a \sim b$. On a donc $ab^{-1} \in H$, donc, en prenant l'inverse, $\left(ab^{-1}\right)^{-1} \in H$, c'est-à-dire $ba^{-1} \in H$, soit $b \sim a$: la relation $\sim$ est donc symétrique.

Soit a, b, c trois éléments de G, avec $a \sim b$ et $b \sim c$. On a donc $ab^{-1} \in H$ et $bc^{-1} \in H$. En multipliant entre eux ces deux éléments de H, on obtient $(ab^{-1})(bc^{-1}) \in H$, c'est-à-dire $ac^{-1} \in H$, soit $a \sim c$. La relation $\sim$ est donc transitive.

La relation $\sim$ est donc une relation d'équivalence.

(2) Soit a un élément fixé de G. L'objectif est de montrer que sa classe d'équivalence $\dot{a}$ possède le même nombre d'éléments que H. Pour ce faire, une bonne idée serait de montrer qu'il existe une bijection entre $\dot{a}$ et H. Et pour montrer qu'une bijection existe, une bonne idée pourrait être de la sortir de sa manche, et voir qu'elle marche !
Introduisons donc une application $f : H \to \dot{a}$ définie par : pour tout h de H, $f(h) = ha$.

Vérifions tout d'abord que f est bien une application. La difficulté vient ici de ce que la formule ha a certes un sens, mais qu'il y a un doute *a priori* quant à savoir si ha appartient bien à $\dot{a}$. Heureusement, la question est plus facile à résoudre qu'à poser ! C'est en effet une simple vérification : $a(ha)^{-1} = aa^{-1}h^{-1} = h^{-1} \in H$; donc $a \sim ha$; en d'autres termes $ha \in \dot{a}$.

Vérifions que f est une bijection. Soit un $b \in \dot{a}$. Cherchons les antécédents de b. Un élément h de H est antécédent de b par f si et seulement si $b = ah$, c'est-à-dire si et seulement si $h = ba^{-1}$. Il y a donc au plus un antécédent, à savoir ba^{-1}, et comme en outre $b \sim a$, l'élément ba^{-1} est dans H et il y a exactement un antécédent.

f est donc une bijection, et $\dot{a}$ compte donc exactement autant d'éléments que H.

(3) Il ne reste plus qu'à conclure... On dispose d'une relation d'équivalence $\sim$, donc d'un ensemble-quotient $G/\sim$, qui constitue une partition de G. Chacune des parties de G figurant dans cette partition possède exactement $\operatorname{Card} H$ éléments ; le nombre total d'éléments de G est donc égal au produit de $\operatorname{Card} H$ par le nombre de parties de G figurant dans la partition $G/\sim$ et est en particulier un multiple de $\operatorname{Card} H$.

•

6 - Noyaux

Une petite définition, à l'usage pratique pour prouver des injectivités... Une section courte sans guère de commentaires.
Définition 12-6-102 : Soit f un morphisme de groupes, allant d'un groupe G vers un groupe G', dont l'élément neutre est noté e'. Le **noyau** de f est par définition l'ensemble des éléments x de G tels que $f(x) = e'$.
Notation 12-6-48 : Le noyau de f est noté $\operatorname{Ker} f$ (abréviation de l'allemand "Kern").

Le fait suivant est presque évident, mais je ne peux m'interdire de le souligner :

Proposition 12-6-65 : Le noyau d'un morphisme est un sous-groupe du groupe de départ.
Démonstration : Soit f un morphisme d'un groupe noté G de neutre noté e vers un groupe noté G' de neutre noté e'.

On sait que $f(e) = e'$ donc $e \in \operatorname{Ker} f$, qui n'est donc pas vide.
Soit a, b deux éléments de $\operatorname{Ker} f$. On a alors $f(ab^{-1}) = f(a)[f(b)]^{-1} = e'e' = e'$, donc $ab^{-1} \in \operatorname{Ker} f$. •

Proposition 12-6-66 : Soit f un morphisme de groupes, le neutre du groupe de départ étant noté e. L'application f est injective si et seulement si $\operatorname{Ker} f = \{e\}$.
Démonstration : Sans surprise, vérifions successivement les deux implications. On notera e' le neutre du groupe d'arrivée.
Preuve de $\Rightarrow$.
Supposons f injective.
On sait déjà que $f(e) = e'$, et donc que $\{e\} \subset \operatorname{Ker} f$. Réciproquement, si $a \in \operatorname{Ker} f$, $f(a) = f(e) = e'$, et comme f est injective, $a = e$. D'où l'égalité $\{e\} = \operatorname{Ker} f$.
Preuve de $\Leftarrow$.
Supposons $\operatorname{Ker} f = \{e\}$.
Soit a et b deux éléments du groupe de départ vérifiant $f(a) = f(b)$. On a alors $f(ab^{-1}) = f(a)[f(b)]^{-1} = e'e' = e'$, donc $ab^{-1} \in \operatorname{Ker} f$, donc $ab^{-1} = e$, donc $a = b$. f est bien injective. •

7 - Puissances et ordre d'un élément d'un groupe

Définition 12-7-103 : Soit a un élément d'un groupe et n un entier relatif. On appelle **puissance n-ème** de a l'élément a^n défini comme valant $\underbrace{aa\ldots aa}_{n \text{ fois}}$ si $n \geq 1$, comme valant l'inverse de a^{-n} si $n \leq -1$ et comme valant l'élément neutre si $n = 0$.
Notation 12-7-49 : L'ensemble des puissances de a sera noté $<a>$.
Proposition 12-7-67 : Soit a un élément d'un groupe et n, m deux entiers ; on a alors $a^{m+n} = a^m a^n$ et $(a^m)^n = a^{mn}$.
Démonstration : C'est très simple à voir avec des points de suspension, en n'oubliant pas de distinguer plein de cas selon les signes des divers entiers des formules —la définition dépendant de ce signe. Comme c'est à la fois très facile et très fastidieux, j'oublierai discrètement de le faire. •

On en déduit aussitôt la très élémentaire
Proposition 12-7-68 : Soit G un groupe, et a un élément de G. L'ensemble $<a>$ est un sous-groupe de G.
Démonstration : $<a>$ n'est pas vide, puisqu'il contient $e = a^0$. Si x et y sont deux éléments de $<a>$, on peut trouver deux entiers (relatifs) m et n permettant d'écrire $x = a^m$ et $y = a^n$. Dès lors $xy^{-1} = a^{m-n}$ et donc $xy^{-1} \in <a>$.
Définition 12-7-104 : Soit a un élément d'un groupe, dont le neutre est noté e. Si pour tout $n \geq 1$, $a^n \neq e$ on dit que a est **d'ordre infini**. Sinon on appelle **ordre** de a le plus petit entier $n \geq 1$ tel que $a^n = e$.
Afin de tenter de prévenir les confusions, introduisons un autre sens du mot "ordre", pas du tout synonyme du précédent et un peu superflu :
Définition 12-7-105 : Soit G un groupe fini. L'**ordre** de G est son nombre d'éléments.
Histoire d'appliquer rétroactivement la division euclidienne qui sera correctement définie dans quelques pages, démontrons le
Théorème 12-7-20 : Soit a un élément d'un groupe. L'ordre de a est égal au nombre d'éléments de $<a>$.
Démonstration : La preuve étant plus longue que la moyenne, essayons de dégager des étapes intermédiaires avec des énoncés précis, qui nous permettront de souffler quand ils seront atteints. On notera e l'élément neutre du groupe considéré.
Étape intermédiaire 1 : si l'ordre de a est fini, noté n, $<a> = \{e\ (= a^0), a, a^2, \ldots, a^{n-1}\}$
Preuve de l'étape 1. Soit b un élément de $<a>$, c'est-à-dire une puissance de a. On peut donc mettre b sous forme a^k pour un entier relatif k. Effectuons la division euclidienne de k par n, ainsi $k = nq + r$, avec $0 \leq r < n$. On a alors $b = a^k = a^{nq+r} = (a^n)^q a^r = e^q a^r = a^r$, donc $b \in \{e, a, a^2, \ldots, a^{n-1}\}$, ce qui montre l'inclusion $<a> \subset \{e, a, a^2, \ldots, a^{n-1}\}$; l'autre inclusion étant évidente,
l'étape 1 est prouvée.
Étape intermédiaire 2 : si l'ordre de a est fini, le théorème est vrai.
Preuve de l'étape 2. Notons n l'ordre de a. Il découle du résultat de l'étape 1 que dans cette hypothèse l'ensemble $<a>$ possède **au plus** n éléments. L'étudiant distrait croira même qu'on a déjà

prouvé qu'il en possède exactement n et qu'on a donc fini, mais son condisciple plus observateur remarquera que nous ne savons pas encore si dans l'énumération $e, a, a^2, \ldots, a^{n-1}$ figurent bien n éléments **distincts**.

Prouvons donc ce dernier fait ; supposons que dans cette énumération il y ait deux termes a^i et a^j qui représentent le même élément du groupe, avec pourtant $i < j$. On aurait alors $a^{j-i} = e$. Mais par ailleurs, comme $i < j$, on obtient $0 < j - i$ et donc $1 \leq i - j$, et comme $0 \leq i$ et $j < n$, on obtient $j - i < n$. Mais ceci contredit la définition de n comme le **plus petit** entier supérieur ou égal à 1 tel que $a^n = e$. L'hypothèse était donc absurde, et l'énumération décrivant $<a>$ à l'étape 1 est une énumération sans répétition.

Le nombre d'éléments de $<a>$ est donc bien égal à n, et
l'étape 1 est prouvée.

Étape intermédiaire 3 : si l'ordre de a est infini, le théorème est vrai.

Preuve de l'étape 3. Dans ce cas, tout le travail consiste à prouver que $<a>$ est un ensemble infini. La vérification est du même esprit qu'à l'étape 2, en plus simple : on va prouver que pour $i < j$, les éléments a^i et a^j de $<a>$ sont distincts. Pour ce faire, supposons que deux d'entre eux soient égaux ; on aurait alors $a^{j-i} = e$, avec pourtant $1 \leq j - i$ et a ne serait pas d'ordre infini. Ainsi
l'étape 3 est prouvée.

•

Histoire d'utiliser un peu la notion d'ordre, donnons un énoncé qui peut servir pour gagner du temps dans tel ou tel exercice très concret.

Proposition 12-7-69 : Soit G un groupe **fini** et H un sous-ensemble de G. Alors H est un sous-groupe de G si et seulement si :

(i) H n'est pas vide.

(ii) Pour tous a, b de H, le produit ab est aussi dans H.

En d'autres termes, dans le cas particulier d'un sous-ensemble d'un groupe **fini** (et seulement dans ce cas !) on peut faire des économies et éviter de travailler sur les ennuyeux symétriques pour examiner un potentiel sous-groupe.

Démonstration : La seule difficulté est évidemment de vérifier la propriété (iii) de la définition des "sous-groupes". Prenons donc un élément a de H. On commence par traiter à part le cas stupide où $a = e$, et où il est clair qu'on a aussi $a^{-1} = e \in H$. Pour le cas sérieux où $a \neq e$, considérons le sous-groupe $<a>$ de G. Ce sous-groupe est fini, puisqu'inclus dans G. On déduit donc du théorème précédent (en fait de sa partie la plus facile, l'étape 3 de sa preuve) que a est d'ordre fini. Notons n l'ordre de a ; comme $a \neq e$, on a l'inégalité $n \geq 2$ et donc $n - 1 \geq 1$; écrivons l'identité $a^{n-1} = a^n a^{-1} = a^{-1}$, et revenons dans cette formule à la définition de a^{n-1} : on obtient $a^{-1} = \underbrace{aa\ldots aa}_{n-1 \text{ fois}}$ comme produit d'un nombre positif d'exemplaires de a ; par la propriété (ii), on en déduit que $a^{-1} \in H$.

•

Chapitre 13 - Autres structures usuelles

Il s'agit ici simplement de rajouter un peu de vocabulaire pour pouvoir décrire les propriétés que possèdent les ensembles de nombres usuels. Le chapitre se limitera donc à quelques définitions.

1 - Anneaux

Définition 13-1-106 : Soit A un ensemble muni de deux opérations, notées $+$ et $\times$. On dit que A est un **anneau** lorsque :
(i) Pour l'addition, A est un groupe commutatif.
(ii) La multiplication est associative.
(iii) La multiplication possède un élément neutre.
(iv) Pour tous a, b, c de A, $(a+b)c = ac+bc$ et $c(a+b) = ca+cb$.

L'archétype de l'anneau est l'ensemble $\mathbf{Z}$ des entiers relatifs ; dans un anneau quelconque on peut calculer "comme" dans $\mathbf{Z}$. Méfiance sur un seul point toutefois : la définition n'exigeant pas que la mutiplication soit commutative, certaines formules peuvent être un peu plus perverses ; par exemple $(a+b)^2$ se développe en $a^2+ba+ab+b^2$, mais ne peut pas dans un anneau trop général être regroupé en $a^2+2ab+b^2$ (puisque ab n'a aucune raison d'être égal à ba).

Voici un autre exemple :

Proposition 13-1-70 : Soit E un espace vectoriel. Pour l'addition et la composition $\mathcal{L}(E)$ est un anneau.

Démonstration : Les propriétés d'"anneau" sont généralement évidentes à vérifier ; la plus intéressante est la distributivité, faisant l'objet de la proposition 8-2-45, qui est liée à la linéarité. Le neutre pour la composition est sans surprise l'application identique.

•

Définition 13-1-107 : Un anneau est dit **commutatif** quand sa multiplication est commutative.

Définition 13-1-108 : Un anneau A est dit **intègre** lorsque :
(i) A possède au moins deux éléments.
(ii) Pour tous a, b non nuls de A, $ab \neq 0$.

On notera que l'anneau des endomorphismes, dès que la dimension est supérieure ou égale à 2, n'est ni commutatif ni intègre.

2 - Corps commutatifs

Définition 13-2-109 : On dit qu'un anneau K est un **corps commutatif** lorsque :
(i) La multiplication est commutative.
(ii) K possède au moins deux éléments.
(iii) Tout élément non nul de K possède un inverse pour la multiplication.

Les archétypes des corps commutatifs étant naturellement $\mathbf{Q}$, ensemble des fractions, et, encore mieux connus des étudiants, $\mathbf{R}$ et $\mathbf{C}$.

Chapitre 14 - Arithmétique

Guère d'introduction tonitruante à faire, sinon pour souligner que ce chapitre a le charme de n'utiliser comme notions admises que celles dont on a parlé jusqu'ici, à savoir les notations de la théorie des ensembles naïve et les connaissances évidentes sur les entiers, et présente donc l'agrément de donner une image de démonstrations (que j'espère) totalement crédibles.

1 - Vocabulaire de base

Définition 14-1-110 : On dit qu'un entier a ($\in \mathbf{Z}$) est un **multiple** d'un entier b ($\in \mathbf{Z}$), ou que b est un **diviseur** de a lorsqu'il existe un entier k ($\in \mathbf{Z}$) tel que $a = kb$.

2 - Nombres premiers

Définition 14-2-111 : On dit qu'un entier $p \geq 2$ est premier lorsqu'il possède pour seuls diviseurs positifs 1 et lui-même.

On notera au passage qu'au hasard des définitions, on parle parfois d'entiers de $\mathbf{Z}$ et parfois d'entiers positifs. Ce n'est qu'exceptionnellement très significatif ; la principale fonction est d'être cohérent avec le reste du monde. Ainsi, comme partout ailleurs, dans ce cours 3 est un nombre premier alors que -3 n'en est pas un. En revanche, les nombres négatifs étant autorisés dans la définition de "diviseurs", l'entier 3 possède en tout et pour tout quatre diviseurs (à savoir -3, -1,1 et 3).

Et tout de suite un joli théorème, qui semble dû à Euclide, autant que je sache :

Théorème 14-2-21 : Il y a une infinité de nombres premiers.

Démonstration : Soit A l'ensemble des nombres premiers. A est une partie de $\mathbf{N}$, et est non vide (2 est premier). On va supposer A finie et aboutir à une absurdité.

Supposons donc A finie. Dès lors que A est une partie finie de $\mathbf{N}$, évidemment non vide (2 est premier), il possède un plus grand élément. Notons P ce plus grand élément, le mystérieux "plus grand nombre premier".

Considérons alors l'entier $N = P\,! + 1$. Pour tout entier k tel que $2 \leq k \leq P$, comme k divise $P\,!$ et ne divise pas 1, k ne peut diviser N. Tout diviseur de N (et en particulier tout diviseur premier de N) est donc strictement supérieur à P.

Souvenons nous alors d'avoir démontré la proposition 2-1-4, qui assure que N possède au moins un diviseur premier. Mais alors, chacun de ces diviseurs premiers contredit la maximalité de P. Absurdité ! •

3 - Division euclidienne

Il s'agit de formaliser avec précision la bonne division euclidienne forcément déjà connue.

Théorème 14-3-22 : Soit a un entier (relatif) et $b \geq 1$ un entier strictement positif.

Alors il existe un couple (q, r) unique (d'entiers) vérifiant la double condition :

$$a = bq + r \qquad \text{et} \qquad 0 \leq r < b.$$

Démonstration : On prouvera successivement l'existence et l'unicité de (q, r).

* Existence de (q, r) : la démonstration se prête bien à discuter selon le signe de a. Le cas où $a \geq 0$ est le cas contenant l'essentiel de la démonstration ; lorsque $a < 0$, on ne peut utiliser mot à mot la même preuve, mais on se ramène alors sans mal au cas intéressant déjà traité.

• Premier cas (le cas significatif) : si $a \geq 0$.

L'idée de la preuve est de dire que le quotient de a par b est le plus grand entier Q tel que bQ soit encore plus petit que a.

Introduisons donc l'ensemble $A = \{c \in \mathbf{N} \mid bc \leq a\}$. L'ensemble A est un ensemble d'entiers naturels ; il est non vide, car il est clair que $0 \in A$. Il est fini : en effet soit d un entier tel que $d \geq a + 1$; on a alors $bd \geq b(a + 1) \geq a + 1 > a$, donc $d \notin A$ et ainsi A ne contient que des entiers inférieurs ou égaux à a.

L'ensemble A possède donc un plus grand élément q. Posons $r = a - bq$. La première condition sur (q,r) est alors évidemment vérifiée, c'est la seconde qui nécessite une vérification.
Comme $q \in A$, par définition de A, on a $bq \leq a$. Donc $r = a - bq \geq 0$.
Comme q est maximal parmi les éléments de A, $q+1 \notin A$. Donc $b(q+1) > a$, donc $r = a - bq < b$.
L'existence est prouvée dans ce cas.
• Second cas (preuve sans imagination) : si $a < 0$.
Posons $a' = a(1-b)$. Comme $a < 0$ et $1 - b \leq 0$, on obtient $a' \geq 0$.
On peut donc, en appliquant le premier cas, faire la division euclidienne de a' par b ; notons (q',r) le couple ainsi obtenu : on a alors $a' = bq' + r$, avec en outre $0 \leq r < b$. En réinjectant la définition de a', on écrit alors $a - ba = bq' + r$, donc $a = b(q' + a) + r$. Si on pose $q = q' + a$, on constate qu'on a réussi la division euclidienne de a par b.

* Unicité de (q,r) : soit (q_1, r_1) et (q_2, r_2) deux couples vérifiant tous deux les deux conditions exigées dans l'énoncé du théorème.
On déduit de $q = bq_1 + r_1 = bq_2 + r_2$ que $b(q_1 - q_2) = r_1 - r_2$. Ainsi, $r_1 - r_2$ est un multiple de b.
Des conditions $0 \leq r_1$ et $r_2 < b$, on déduit que $-b < r_1 - r_2$.
Des conditions $r_1 < b$ et $0 \leq r_2$, on déduit que $r_1 - r_2 < b$.
Ainsi $r_1 - r_2$ est un multiple de b compris strictement entre $-b$ et b. La seule possibilité est que $r_1 - r_2$ soit nul. On en déduit $r_1 = r_2$, puis, en allant reprendre l'égalité $b(q_1 - q_2) = r_1 - r_2$, que $q_1 = q_2$.

•

4 - PGCD et PPCM

Les deux théorèmes qui se suivent sont agréablement parallèles ; il est donc amusant de constater que leurs preuves sont plus différentes qu'on ne pourrait s'y attendre. Il est possible de les déduire l'un de l'autre, mais il est instructif de les prouver très séparément. Vous verrez donc plusieurs preuves de l'un comme de l'autre.

Théorème 14-4-23 : Soit $a \geq 1$ et $b \geq 1$ deux entiers. Alors il existe un unique entier $M \geq 1$ tel que pour tout $m \geq 1$

$$m \text{ multiple de } a \text{ et } b \iff m \text{ multiple de } M.$$

Théorème 14-4-24 : Soit $a \geq 1$ et $b \geq 1$ deux entiers. Alors il existe un unique entier $D \geq 1$ tel que pour tout $d \geq 1$

$$d \text{ divise } a \text{ et } b \iff d \text{ divise } D.$$

Ces théorèmes sont vendus avec deux compléments, le premier occasionnellement utile, le second totalement fondamental.

Complément 1 : $MD = ab$.

Complément 2 (identité de Bézout) : il existe deux entiers (relatifs) s et t tels que $D = sa + tb$.

Et tant qu'on y est avant de passer aux démonstrations :

Définition 14-4-112 : Le **plus petit multiple commun** de deux entiers a et b est l'entier M apparaissant dans l'énoncé du théorème 14-4-23.

Notation 14-4-50 : Le plus petit multiple commun de a et b sera noté $\mathrm{PPCM}(a,b)$.

Définition 14-4-113 : Le **plus grand commun diviseur** de deux entiers a et b est l'entier M apparaissant dans l'énoncé du théorème 14-4-24.

Notation 14-4-51 : Le plus grand commun diviseur de a et b sera noté $\mathrm{PGCD}(a,b)$.

Première démonstration du théorème 14-4-23 : Cette démonstration est la plus élémentaire ; elle consiste à choisir pour M le multiple commun de a et b le plus "petit" (au sens de la relation habituelle $\leq$), puis vérifier qu'il marche. La preuve est en deux parties : d'abord l'existence de M (partie significative) puis son unicité (partie très facile).

* Existence de M.
Introduisons l'ensemble A formé des entiers strictement positifs simultanément multiples de a et de b. L'ensemble A n'est pas vide, puisqu'il contient l'entier ab. Il admet donc un plus petit élément M. On va vérifier que ce M convient.

Pour faire cette vérification, soit un $m \geq 1$; nous avons désormais à montrer une équivalence, distinguons méthodiquement les deux sens.

• Preuve de $\Rightarrow$: Supposons donc que m est un multiple commun de a et b, et montrons que c'est un multiple de M. Pour ce faire, effectuons la division euclidienne de m par M, soit $m = Mq + r$, avec $0 \leq r < M$. Comme m et M sont des multiples de a, $r = m - Mq$ aussi ; de même avec b. Ainsi r est un multiple commun de a et b. Si r était un entier strictement positif, vu l'inégalité $r < M$ il contredirait la minimalité de M. C'est donc que $r = 0$ et donc que m est un multiple de M.

• Preuve de $\Leftarrow$: Supposons ici que m est un multiple de M. Comme M est lui-même multiple de a, m est à son tour multiple de a ; de même avec b. C'est réglé.

* Unicité de M.

Soit M et M' vérifiant les hypothèses du théorème. Comme M est multiple de M, c'est un multiple commun de a et b, donc un multiple de M'. De même, M' est un multiple de M. Comme ils sont tous deux strictement positifs, ils sont forcément égaux.

•

Voici maintenant une première démonstration de l'existence (et l'unicité) du PGCD, qui l'obtient à partir du PPCM. Cette démonstration a le confort d'être dépourvue d'idée subtile, l'avantage de prouver le "complément 1". Elle a l'inconvénient de ne pas prouver le "complément 2", et de ne pas fournir une méthode rapide de calcul du PGCD.

Première démonstration du théorème 14-4-24 :

* Existence de D.

On note M le PPCM de a et b, et on pose $D = ab/M$. Remarquons que ce D et bien un entier : en effet ab étant un multiple commun évident de a et b, c'est un multiple de leur PPCM. Reste à prouver qu'il marche...
Pour faire cette vérification, soit un $d \geq 1$; nous avons désormais à montrer une équivalence, distinguons méthodiquement les deux sens.

• Preuve de $\Rightarrow$: Supposons donc que d est un diviseur commun de a et b. On peut donc introduire deux entiers k et l tels que $a = kd$ et $b = ld$. Pour travailler sur ce sur quoi nous avons des informations, à savoir les multiples de a et b, introduisons le nombre $m = ab/d$. Ce nombre m vaut aussi $(a/d)b = kb$ et $(b/d)a = la$. C'est donc un entier, et même un multiple commun de a et b. C'est donc un multiple de M. Il existe donc un entier c tel que $m = cM$, soit $ab/d = c\,ab/D$, donc $D = cd$. On a bien prouvé que d divise D.

• Preuve de $\Leftarrow$: puisque $a = D\,(M/b)$ où M/b est un entier, D divise a ; symétriquement puisque $b = D\,(M/a)$, D divise b. Supposons maintenant que d divise D. On voit alors aussitôt que d divise a et b.

* Unicité de D.

C'est exactement le même principe que pour le PPCM, je le laisse en exercice (très) facile.

* Preuve du "complément 1".

Il tombe immédiatement vu la formule donnant D à partir de M.

•

Comme promis, voici maintenant une deuxième démonstration du théorème 14-4-24, très différente dans son esprit, et qui permet pour guère plus cher de montrer simultanément le "complément 2".

Deuxième démonstration du théorème 14-4-24 :

La démonstration est une récurrence sur b ; techniquement, on gagne sérieusement en confort si on autorise b à être nul, ce que je n'ai pas fait volontairement en énonçant le théorème dans l'espoir qu'il soit plus clair. On montrera donc légèrement mieux que l'énoncé de la page précédente, puisqu'on prouvera le résultat sous l'hypothèse "$a \geq 1$ et $b \geq 0$".

Avant de se lancer dans la récurrence proprement dite, je vais donner un "résumé de la preuve" sous forme de programme informatique récursif :

* PGCD$(a, 0) = a$.

* En notant r le reste de la division euclidienne de a par b, les diviseurs communs de a et b sont les diviseurs communs de b et r, d'où :

PGCD$(a, b) =$ PGCD(b, r).

Ce résumé de démonstration convaincra peut-être les esprits les plus agiles, mais à notre niveau d'entraînement, il est plus prudent de faire ce qui est derrière les formulations récursives : une bonne vieille récurrence. On va démontrer par "récurrence forte" sur $b \geq 0$ l'hypothèse (H_b) suivante :

Pour tout $a \geq 1$, il existe deux entiers (relatifs) s et t tels que, pour tout $d \geq 1$, d divise a et $b \iff d$ divise $sa + tb$.

* Vérifions (H_0).
Soit a un entier avec $a \geq 1$; tout entier $d \geq 1$ qui divise a divise aussi $b = 0$ puisque $0d = 0$. Pour tout $d \geq 1$, on a donc : d divise a et $0 \iff d$ divise a. Prenons alors $s = 1$ et $t = 0$: on a donc bien pour tout $d \geq 1$: d divise a et $0 \iff d$ divise $sa + t \times 0$.

* Soit b un entier fixé, avec $b \geq 1$. Supposons la propriété (H_c) vraie pour tout c avec $0 \leq c < b$ et montrons (H_b).
Soit a un entier avec $a \geq 1$. Notons $a = bq + r$ la division euclidienne de a par b (qu'on peut réaliser puisque $b \geq 1$).
Vérifions l'affirmation intermédiaire suivante : pour tout $d \geq 1$, d est un diviseur commun de a et $b \iff d$ est un diviseur commun de b et r. (Avec des mots peut-être plus lisibles : "les diviseurs communs de a et b sont les mêmes que ceux de b et r").

Soit d un diviseur commun de a et b, alors d divise aussi $r = a - bq$; réciproquement soit d un diviseur commun de b et r, alors d divise aussi $a = bq + r$.

L'affirmation intermédiaire est donc démontrée.
On peut alors appliquer l'hypothèse de récurrence (H_r) (puisque précisément $0 \leq r < b$) sur l'entier $b \geq 1$.
On en déduit qu'il existe deux entiers relatifs σ et τ tels que pour tout $d \geq 1$, d divise b et $r \iff d$ divise $\sigma b + \tau r$.
Remarquons enfin que $\sigma b + \tau r = \sigma b + \tau(a - bq) = \tau a + (\sigma - q)b$, et qu'ainsi, si on pose $s = \tau$ et $t = \sigma - q$ on a bien prouvé que, pour tout $d \geq 1$, d divise a et $b \iff d$ divise $sa + tb$.
(H_b) est donc démontrée.

* On a donc bien prouvé (H_b) pour tout $b \geq 0$, — donc *a fortiori* pour tout $b \geq 1$, ce qui prouve le théorème 14-4-24 et son complément 2.

* En fait, il reste à prouver l'unicité de D, pour laquelle je renvoie à la démonstration précédente (où je disais que je la laissais en exercice).

•

Un petit exemple sur des vrais nombres concrets, pour nous soulager l'esprit après tant de lettres :

Calcul du PGCD de 137 et 24 :

On fait des divisions euclidiennes successives :

(1)	$137 =$	$5 \times$	$24 +$	17	PGCD(137,24) = PGCD(24,17)
(2)	$24 =$	$1 \times$	$17 +$	7	PGCD(24,17) = PGCD(17,7)
(3)	$17 =$	$2 \times$	$7 +$	3	PGCD(17,7) = PGCD(7,3)
(4)	$7 =$	$2 \times$	$3 +$	1	PGCD(7,3) = PGCD(3,1)
(5)	$3 =$	$3 \times$	$1 +$	0	PGCD(3,1) = PGCD(1,0) = 1

Ces calculs permettent ensuite sans mal de reconstituer une identité de Bézout :

On écrit :	$1 =$	$1 \times$	$1 +$	0×0		
On va repêcher le				0	dans (5) :	$0 = 3 - (3 \times 1)$
On reporte	$1 =$	$1 \times$	$1 +$	$0 \times [3 - (3 \times 1)]$		
On regroupe :	$1 =$	$0 \times$	$3 +$	1×1		
On va repêcher le				1	dans (4) :	$1 = 7 - (2 \times 3)$
On reporte	$1 =$	$0 \times$	$3 +$	$1 \times [7 - (2 \times 3)]$		
On regroupe :	$1 =$	$1 \times$	$7 +$	$(-2) \times 3$		
On va repêcher le				3	dans (3) :	$1 = 17 - (2 \times 7)$
On reporte	$1 =$	$1 \times$	$7 +$	$(-2) \times [17 - (2 \times 7)]$		
On regroupe :	$1 =$	$(-2) \times$	$17 +$	5×7		
On va repêcher le				7	dans (2) :	$7 = 24 - (1 \times 17)$
On reporte	$1 =$	$(-2) \times$	$17 +$	$5 \times [24 - (1 \times 17)]$		
On regroupe :	$1 =$	$5 \times$	$24 +$	$(-7) \times 17$		
On va repêcher le				17	dans (1) :	$17 = 137 - (5 \times 24)$
On reporte	$1 =$	$5 \times$	$24 +$	$(-7) \times [137 - (5 \times 24)]$		
On termine en regroupant :	$1 =$	$(-7) \times$	$137 +$	40×24		

Donnons, avant de quitter les PGCDs une dernière

Définition 14-4-114 : On dit que deux entiers $a \geq 1$ et $b \geq 1$ **premiers entre eux** lorsque leur seul diviseur commun positif est 1.

Bien évidemment, on veillera à ne pas confondre cette notion avec celle de "nombre premier" qui n'a (en forçant un peu la note) strictement rien à voir.

5 - Lemme de Gauss et décomposition en facteurs premiers

Le lemme de Gauss est mis en relief par certains ; dans la mesure où le théorème qu'il permet de démontrer —l'unicité de la décomposition en facteurs premiers— me semble beaucoup plus facile d'usage pour un utilisateur peu expérimenté, je l'énonce sans commentaires, ou plus exactement sans autre commentaire que ce commentaire négatif.

Lemme 14-5-5 : Soit a, b, c trois entiers de $\mathbf{N}^*$. Si a divise bc et est premier avec c, alors a divise b.

Démonstration : Puisque a est premier avec c, le PGCD de a et c est 1, donc il existe des entiers relatifs s et t tels que $sa + tc = 1$. Multiplions cette identité par b : on obtient $b = asb + tbc$. Mais dans cette écriture, asb est évidemment multiple de a tandis que tbc l'est parce que bc est multiple de a. On en déduit que b, somme des deux multiples de a que sont asb et tbc, est lui-même un multiple de a.

•

Théorème 14-5-25 : (énoncé approximatif) Tout entier $n \geq 2$ peut être écrit de façon unique comme produit de facteurs premiers.

J'ai qualifié l'énoncé d'"approximatif" car il n'est pas si clair de savoir ce que veut dire le "unique" : on peut écrire $6 = 2 \times 3 = 3 \times 2$ mais il faut évidemment considérer que c'est la même chose... Pour pouvoir comprendre voire utiliser le théorème, cet énoncé suffira bien ; mais pour le démontrer, il faut être plus précis.

Répétons donc le

Théorème 14-5-25 : (énoncé précis) Tout entier $n \geq 2$ peut être écrit comme produit de facteurs premiers. De plus, si on dispose de deux écritures :

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k} \qquad \text{et} \qquad n = q_1^{\beta_1} q_2^{\beta_2} \cdots q_l^{\beta_l}$$

dans lesquelles $k \geq 1$, $l \geq 1$, les entiers $p_1 < p_2 < \ldots < p_k$ et $q_1 < q_2 < \ldots < q_l$ sont tous premiers et rangés dans l'ordre croissant, les exposants $\alpha_1, \alpha_2, \ldots, \alpha_k, \beta_1, \beta_2, \ldots, \beta_l$ sont tous des entiers strictement positifs, alors ces deux écritures sont les mêmes (au sens précis suivant : $k = l$ et pour tout i avec $1 \leq i \leq k = l$, $p_i = q_i$ et $\alpha_i = \beta_i$).

Démonstration : À énoncé indigeste, démonstration indigeste...

L'existence a déjà été prouvée (c'était la proposition 2-1-4). Il faut passer à l'unicité, qu'on va prouver par récurrence.

On va donc montrer par récurrence ("forte") sur n le résultat d'unicité (H_n) énoncé ci-dessus.

* Démonstration de (H_2) (et en fait même de (H_p) pour tout nombre premier p).

Supposons $n = p$ premier écrit sous forme de produit $p = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$. Chaque p_i est un diviseur positif de p non égal à 1, donc chaque p_i est égal à p. Ceci entraîne aussitôt que $k = 1$ et que $\alpha_k = 1$ (sans cela le produit serait supérieur ou égal à p^2 donc distinct de p). L'écriture $p = p$ est donc la seule possible pour p.

* Soit n un entier fixé, non premier, avec $n > 2$, et supposons l'hypothèse d'unicité (H_m) prouvée pour tout entier m avec $2 \leq m < n$.

Première étape

• Montrons dans un premier temps que $p_k = q_l$. Supposons tout d'abord que l'on ait $p_k > q_l$ et montrons que l'on aboutit à une absurdité.

Puisque les q_j sont supposés rangés dans l'ordre croissant, p_k est alors forcément distinct de tous les q_j ; p_k et chaque q_j étant premiers, on en conclut que leur seul diviseur commun positif est 1 : p_k et q_j sont donc premiers entre eux.

Fixons un j entre 1 et l et montrons par récurrence sur $b \geq 0$ l'énoncé fort intuitif suivant : (H'_b) : p_k est premier avec q_j^b.

* (H'_0) est évident.

* Soit $b \geq 0$ un entier fixé, supposons (H'_b) vrai et montrons (H'_{b+1}).

Si (H'_{b+1}) était faux, le PGCD de p_k et q_j^{b+1} ne serait pas 1 ; comme c'est un diviseur positif de p_k, ce serait p_k qui diviserait donc q_j^{b+1}. On peut alors appliquer le lemme de Gauss : comme p_k divise $q_j^{b+1} = q_j^b q_j$ et que p_k est premier avec q_j, p_k divise q_j^b. Mais ceci contredit l'hypothèse (H'_b). L'hypothèse (H'_{b+1}) est donc vraie.

On a donc bien montré que pour tout $b \geq 0$, p_k est premier avec q_j^b. En particulier, p_k est premier avec $q_j^{\beta_j}$. Comme on a prouvé cette affirmation pour un j quelconque, on a prouvé que pour tout j entre 1 et l, p_k est premier avec $q_j^{\beta_j}$. Ce qu'on a fait avec les puissances de chaque q_j, on va maintenant le recommencer avec le produit de ces puissances. Précisément, on va montrer par récurrence sur l'entier j que pour tout j avec $1 \leq j \leq l$, on a l'énoncé (H''_j) : p_k est premier avec $q_1^{\beta_1} q_2^{\beta_2} \cdots q_j^{\beta_j}$. Les lecteurs encore éveillés (s'il en reste) comprendront que la preuve est à peu près la même que celle des (H'_b), pour les autres, la voilà :

* Pour $j = 1$, on doit prouver que p_k est premier avec $q_1^{\beta_1}$. C'est déjà fait.
* Fixons un j avec $1 \leq j < l$ et supposons l'hypothèse (H''_j) vraie.

Si (H''_{j+1}) était fausse, le PGCD de p_k et $q_1^{\beta_1} q_2^{\beta_2} \cdots q_j^{\beta_j} q_{j+1}^{\beta_{j+1}}$ ne serait pas 1 ; comme c'est un diviseur positif de p_k, ce serait p_k qui diviserait donc $q_1^{\beta_1} q_2^{\beta_2} \cdots q_j^{\beta_j} q_{j+1}^{\beta_{j+1}}$. On peut alors appliquer le lemme de Gauss : comme p_k divise $q_1^{\beta_1} q_2^{\beta_2} \cdots q_j^{\beta_j} q_{j+1}^{\beta_{j+1}} = \left(q_1^{\beta_1} q_2^{\beta_2} \cdots q_j^{\beta_j}\right) q_{j+1}^{\beta_{j+1}}$ et que p_k est premier avec $q_{j+1}^{\beta_{j+1}}$, p_k divise $q_1^{\beta_1} q_2^{\beta_2} \cdots q_j^{\beta_j}$. Mais ceci contredit l'hypothèse (H''_j). L'hypothèse (H''_{j+1}) est donc vraie.

On a donc montré (H''_j) pour tout j entre 1 et l ; en particulier on a montré (H''_l), à savoir que p_k est premier avec $q_1^{\beta_1} q_2^{\beta_2} \cdots q_l^{\beta_l} = n$. Mais pourtant p_k figure dans l'autre décomposition en facteurs premiers de n (ce n'est pas une illusion d'optique, puisqu'on a pris soin de supposer $\alpha_k \geq 1$), donc p_k divise n. D'où contradiction. Ouf !

On ne peut donc avoir $p_k > q_l$.En échangeant les rôles de p et des q, on voit qu'on ne peut pas non plus avoir $q_l > p_k$. On en déduit donc que $q_l = p_k$.

Fin de la première étape

Deuxième étape

• On va alors profiter de ce tout petit morceau d'égalité pour arriver à utiliser l'hypothèse de récurrence et faire tomber toutes les autres égalités requises en cascade.

Notons $m = n/p_k = n/q_l$, on a ainsi :

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k - 1} \qquad \text{et} \qquad m = q_1^{\beta_1} q_2^{\beta_2} \cdots q_l^{\beta_l - 1}.$$

De plus m est strictement inférieur à n, et m est strictement plus grand que 1 car on a fort opportunément supposé n non premier. On va donc appliquer l'hypothèse de récurrence (H_m) à ces deux écritures de m en facteurs premiers. Si on n'est pas méticuleux, on oubliera de s'assurer que tous les exposants sont strictements positifs, et on aura fini tout de suite ; ce sera faux, mais de si peu... Hélas, je ne puis me le permettre et dois donc veiller à ce petit détail, qui me force à distinguer deux sous-cas :

* Si $\alpha_k = 1$. Dans ce cas, la première écriture de m se lit en réalité, après effacement du p_k^0 qui l'encombre :

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_{k-1}^{\alpha_{k-1}}.$$

Ainsi m possède une décomposition en facteurs premiers dans laquelle p_k ne figure pas. Comme sa décomposition est unique, p_k ne peut non plus figurer dans l'autre décomposition, et comme $q_l = p_k$, la seule possibilité est que l'exposant $\beta_l - 1$ soit nul ; ainsi $\beta_l = \alpha_k$ =1, et les deux représentations

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_{k-1}^{\alpha_{k-1}} = m = q_1^{\beta_1} q_2^{\beta_2} \cdots q_{l-1}^{\beta_{l-1}}$$

sont deux décompositions de m en facteurs premiers. On en déduit que $k-1 = l-1$ —et donc $k = l$— puis l'égalité de tous les facteurs premiers et exposants encore en attente d'élucidation.

* Si $\alpha_k > 1$. C'est la même chanson. On remarque tout d'abord qu'on a aussi $\beta_l > 1$ (sans cela, en échangeant les rôles des p_i et des q_j et en utilisant le premier cas, on montrerait que $\alpha_k = 1$) ; donc les deux décompositions

$$m = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k - 1} \qquad \text{et} \qquad m = q_1^{\beta_1} q_2^{\beta_2} \cdots q_l^{\beta_l - 1}$$

vérifient bien les hypothèses du théorème. Elles sont égales, donc $k = l$ et chaque p_i est égal au q_i correspondant, avec le même exposant.

Fin de la deuxième étape

(H_n) est donc prouvée.

La récurrence est donc terminée, et avec elle la démonstration.

•

6 - Sous-groupes de Z

Notation 14-6-52 : Soit $b \geq 0$ un entier positif. On note $b\mathbf{Z}$ l'ensemble des multiples de b.

L'objet de la section est un théorème d'énoncé très simple, et assez pratique :

Théorème 14-6-26 : Les sous-groupes de $\mathbf{Z}$ (pour l'addition) sont exactement les $b\mathbf{Z}$, $b \geq 0$.

Démonstration : Il y a deux choses à démontrer : que les $b\mathbf{Z}$ sont des sous-groupes, et que tout sous-groupe est un $b\mathbf{Z}$.

* Commençons donc par vérifier (c'est très facile) que pour $b \geq 0$ fixé, $b\mathbf{Z}$ est un sous-groupe de $\mathbf{Z}$.

- 0 est multiple de b, donc $b\mathbf{Z}$ n'est pas vide.
- Soit x et y deux éléments de $b\mathbf{Z}$, c'est-à-dire deux multiples de b. Il est clair que $x - y$ est aussi un multiple de b, donc appartient à $b\mathbf{Z}$.

C'est fait. Pour les amateurs d'abstraction, on pouvait remarquer que $b\mathbf{Z} = \langle b \rangle$ ce qui est camouflé par la notation additive de l'opération.

* Soit maintenant H un sous-groupe de $\mathbf{Z}$, montrons qu'il existe un $b \geq 0$ tel que $H = b\mathbf{Z}$. On distinguera deux cas.

- Si $H = \{0\}$, on remarque que $H = 0\mathbf{Z}$ et on a fini.
- Si $H \neq \{0\}$, H possède au moins un élément non nul x, donc au moins un élément strictement positif y (on prendra $y = x$ ou $y = -x$ selon le signe de x). Si on introduit l'ensemble $B = H \cap \mathbf{N}^*$, B est donc un ensemble d'entiers positifs non vide. Il possède un plus petit élément b. On va montrer que b convient.

  * Il me semble raisonnablement clair que $b\mathbf{Z} \subset H$ (hum, est-ce si clair ou est-ce un petit moment de paresse ?).

  * Réciproquement soit a un élément de H. Si on fait la division euclidienne de a par b, soit $a = qb + r$, on en déduit que $r = a - bq \in H$. Comme $r < b$, $r \notin B$, et comme $r \in H \cap \mathbf{N}$ la seule possibilité est que $r = 0$. On en déduit donc que $a = bq \in b\mathbf{Z}$. Ceci prouve l'inclusion $H \subset b\mathbf{Z}$.

  On a donc montré que $H = b\mathbf{Z}$.

On a donc montré, dans les deux cas, que H est de la forme $b\mathbf{Z}$.

•

En application de ce théorème, donnons de nouvelles et élégantes démonstrations des théorèmes 14-4-23 et 14-4-24 ; l'outil à la base reste la division euclidienne, mais il aura été utilisé une seule fois, dans la preuve du théorème qui précède, et on ne fait plus que d'assez simples manipulations ensemblistes.

Deuxième démonstration du théorème 14-4-23 :

Introduisons les sous-groupes de Z que sont $H = a\mathbf{Z}$ et $K = b\mathbf{Z}$. Pour tout $m \geq 1$, m est un diviseur commun de a et b si et seulement si m est dans $H \cap K$. Or $H \cap K$, comme intersection de deux sous-groupes de $\mathbf{Z}$, est lui-même un sous-groupe de $\mathbf{Z}$ (bon, d'accord, je ne l'ai pas mentionné dans le cours sur les sous-groupes, mais j'aurais dû, et de toutes façons c'est très facile). Il existe donc un $M \geq 0$ tel que $H \cap K = M\mathbf{Z}$ (et il est clair que $M > 0$, car $H \cap K$ contient d'autres entiers que 0, par exemple ab). On a alors pour tout $m \geq 1$:

$$m \text{ est un diviseur commun de } a \text{ et } b \iff m \in H \cap K \iff m \in M\mathbf{Z} \iff m \text{ est multiple de } M.$$

L'unicité reste à prouver comme dans la preuve initiale.

•

Troisième démonstration du théorème 14-4-24 :

Introduisons l'ensemble $L \subset \mathbf{Z}$ défini par $L = \{sa + tb \mid s \in \mathbf{Z}, t \in \mathbf{Z}\}$.

On vérifie sans mal que L est un sous-groupe de $\mathbf{Z}$. C'est si facile, que je le laisse au lecteur.

Il existe donc un $D \geq 0$ tel que $L = D\mathbf{Z}$. De plus L n'est manifestement pas réduit à $\{0\}$ (il contient par exemple $a = 1a + 0b$, ou $b = 0a + 1b$), donc $D > 0$. Montrons que D convient.

On a remarqué que a et b sont dans $L = D\mathbf{Z}$. En d'autres termes, ils sont tous deux multiples de D, ou, pour dire cela encore autrement, D est un diviseur commun de a et b. Il est donc clair que tout diviseur de D est à son tour un diviseur commun de a et b.
Par ailleurs, D est dans L, donc peut être mis sous forme $sa + tb$ pour des entiers relatifs s et t. Si on part d'un diviseur commun $d \geq 1$ de a et b, sa et tb sont à leur tour des multiples de d, donc aussi D. d est donc bien un diviseur de D.
Là aussi, je renvoie à la preuve initiale pour l'unicité.

•

7 - Congruences

Juste quelques notations pratiques. La section se réduit à quasiment rien.
Définition 14-7-115 : Soit a, b des entiers relatifs et $n \geq 1$ un entier strictement positif. On dit que a est **congru** à b modulo n lorsque $b - a$ est un multiple de n.

Il est tellement évident de vérifier que, pour n fixé, la relation "est congru à" est une relation d'équivalence sur $\mathbf{Z}$ que cet énoncé n'aura pas même l'honneur d'être qualifié de proposition.
Notation 14-7-53 : Lorsque a est congru à b modulo n, on note :

$$a \equiv b \quad [n].$$

L'intérêt des congruences est d'être "compatibles" avec l'addition et la multiplication, au sens suivant :
Proposition 14-7-71 : Soit $n \geq 1$ fixé et soit a, b, c trois entiers relatifs. Alors :

$$\text{si} \quad a \equiv b \quad [n] \quad \text{alors} \quad a + c \equiv b + c \quad [n] \quad \text{et} \quad ac \equiv bc \quad [n].$$

Démonstration : C'est vraiment trop facile.

•

Exemples : * Quel est le reste de la division par 9 de 123456 ?

$$\begin{aligned} 123456 &= 10^5 + 2 \cdot 10^4 + 3 \cdot 10^3 + 4 \cdot 10^2 + 5 \cdot 10 + 6 \equiv 1^5 + 2 \cdot 1^4 + 3 \cdot 1^3 + 4 \cdot 1^2 + 5 \cdot 1 + 6 \quad [9] \\ &= 1 + 2 + 3 + 4 + 5 + 6 = 21 = 2 \cdot 10 + 1 \qquad \equiv 2 \cdot 1 + 1 = 2 + 1 = 3 \quad [9]. \end{aligned}$$

La réponse est 3.
* Et par 11 ?

$$\begin{aligned} 123456 &= 10^5 + 2 \cdot 10^4 + 3 \cdot 10^3 + 4 \cdot 10^2 + 5 \cdot 10 + 6 \\ &\equiv (-1)^5 + 2 \cdot (-1)^4 + 3 \cdot (-1)^3 + 4 \cdot (-1)^2 + 5 \cdot (-1) + 6 = 3 \quad [11]. \end{aligned}$$

La réponse est aussi 3.

8 - $\mathbf{Z}/n\mathbf{Z}$

En apparence, cette section contient du formalisme très gratuit : désormais, au lieu d'écrire :

$$a \equiv b \quad [n],$$

on apprendra à écrire :

$$\dot{a} = \dot{b} \quad \text{dans } \mathbf{Z}/n\mathbf{Z}.$$

Maigre progrès en apparence ! Toutefois, comme des exemples judicieusement choisis le montreront en fin de section, on a fait plus qu'un simple changement de notations : on a su construire un pont entre ce chapitre et le chapitre précédent, pont par lequel on pourra rapatrier des résultats connus sur les groupes pour effectivement affiner notre connaissance des entiers.
Définition 14-8-116 : Soit $n \geq 1$ fixé. On appelle $\mathbf{Z}/n\mathbf{Z}$ l' ensemble-quotient de $\mathbf{Z}$ par la relation d'équivalence "est congru à" (modulo n).

Exemple : Pour $n = 2$, soit a un entier. Si a est pair, la classe d'équivalence $\dot{a}$ pour la relation de congruence modulo 2 est l'ensemble P de tous les nombres pairs ; si a est impair, $\dot{a}$ est l'ensemble I de tous les nombres impairs, et finalement $\mathbf{Z}/2\mathbf{Z} = \{I, P\}$.
Proposition 14-8-72 : Pour tout $n \geq 1$, $\mathbf{Z}/n\mathbf{Z}$ possède exactement n éléments.

Démonstration : Montrons tout d'abord que $\mathbf{Z}/n\mathbf{Z} = \{\dot{0}, \dot{1}, \ldots, \overset{\cdot}{\overbrace{n-1}}\}$, d'où on déduit aussitôt que $\mathbf{Z}/n\mathbf{Z}$ possède au plus n éléments.

Soit x un élément de $\mathbf{Z}/n\mathbf{Z}$; il existe alors $a \in \mathbf{Z}$ tel que $x = \dot{a}$. Effectuons la division euclidienne de a par n, soit $a = nq + r$; on voit alors que $a \equiv r \quad [n]$ ou encore que $x = \dot{a} = \dot{r}$. Mais $0 \leq r < n$, donc on a bien prouvé que x était dans l'ensemble proposé.

Montrons maintenant que ces n éléments sont deux à deux distincts, prouvant ainsi que $\mathbf{Z}/n\mathbf{Z}$ possède au moins n éléments.

Soit a et b deux entiers distincts avec $0 \leq a < n$ et $0 \leq b < n$. Des inégalités $0 \leq a$ et $b < n$ on déduit que $-n < b - a$; des inégalités $a < n$ et $0 \leq b$ on déduit que $b - a < n$ et de l'hypothèse $a \neq b$ on déduit que $b - a \neq 0$. On en conclut que $a \not\equiv b \quad [n]$, c'est-à-dire que $\dot{a}$ et $\dot{b}$ sont deux éléments distincts de $\mathbf{Z}/n\mathbf{Z}$.

On a donc bien prouvé que $\mathbf{Z}/n\mathbf{Z}$ possède exactement n éléments.

Définition 14-8-117 : Soit $\dot{a}$ et $\dot{b}$ deux éléments de $\mathbf{Z}/n\mathbf{Z}$. On définit la **somme** de $\dot{a}$ et $\dot{b}$ par $\dot{a} + \dot{b} = \overset{\bullet}{\overset{\cdot}{\overbrace{a+b}}}$ et le **produit** de $\dot{a}$ et $\dot{b}$ par $\dot{a} + \dot{b} = \overset{\cdot}{\overbrace{ab}}$.
Prudence ! : Cette définition est aussi innocente en apparence que les 116 qui l'ont précédée. Et pourtant, elle pourrait n'avoir aucun sens ! En effet, la définition de la somme de deux éléments x et y de $\mathbf{Z}/n\mathbf{Z}$ nécessite implicitement de les mettre préalablement sous forme $x = \dot{a}$ et $y = \dot{b}$. Mais il y a plusieurs façons de les mettre sous cette forme ! Il faut donc vérifier que la définition ne dépend pas du choix fait dans cette phase préparatoire. Pour montrer à quel point c'est indispensable, donnons une : Fausse définition (buggée) : Soit $\dot{a}$ et $\dot{b}$ deux éléments de $\mathbf{Z}/n\mathbf{Z}$. On dira que $\dot{a} \leq \dot{b}$ lorsque $a \leq b$. Il est facile de comprendre pourquoi cette "définition" est bonne pour la corbeille à papier : dans $\mathbf{Z}/3\mathbf{Z}$, prenons $x = \dot{0}$ et $y = \dot{2}$. En les écrivant ainsi, la "définition" nous donne : $x \leq y$. Mais on peut aussi écrire $x = \dot{6}$ et $y = \dot{5}$. En s'y prenant ainsi, x n'est pas inférieur ou égal à y. La "définition" n'a en fait aucun sens.
Faisons donc cette indispensable vérification. Soit $x = \dot{a} = \dot{\alpha}$ et $y = \dot{b} = \dot{\beta}$ deux éléments de $\mathbf{Z}/n\mathbf{Z}$. La cohérence de la définition exige de prouver que $\overset{\cdot}{\overbrace{a+b}} = \overset{\cdot}{\overbrace{\alpha+\beta}}$. La vérification est alors évidente $(\alpha + \beta) - (a + b) = (\alpha - a) + (\beta - b)$ étant un multiple de n parce que $\alpha - a$ et $\beta - b$ le sont tous les deux. De même pour $\alpha\beta - ab = \alpha\beta - \alpha b + \alpha b - ab = \alpha(\beta - b) + b(\alpha - a)$.

Ainsi au point où nous en sommes, $\mathbf{Z}/n\mathbf{Z}$ est muni d'une addition et d'une multiplication. Traçons un exemple de tables, pour voir quelle tête ils ont, (et pour rentabiliser le travail qu'a été d'apprendre à taper de belles tables). Ce sera l'exemple de $\mathbf{Z}/5\mathbf{Z}$:

$+$	$\dot{0}$	$\dot{1}$	$\dot{2}$	$\dot{3}$	$\dot{4}$
$\dot{0}$	$\dot{0}$	$\dot{1}$	$\dot{2}$	$\dot{3}$	$\dot{4}$
$\dot{1}$	$\dot{1}$	$\dot{2}$	$\dot{3}$	$\dot{4}$	$\dot{0}$
$\dot{2}$	$\dot{2}$	$\dot{3}$	$\dot{4}$	$\dot{0}$	$\dot{1}$
$\dot{3}$	$\dot{3}$	$\dot{4}$	$\dot{0}$	$\dot{1}$	$\dot{2}$
$\dot{4}$	$\dot{4}$	$\dot{0}$	$\dot{1}$	$\dot{2}$	$\dot{3}$

$\times$	$\dot{0}$	$\dot{1}$	$\dot{2}$	$\dot{3}$	$\dot{4}$
$\dot{0}$	$\dot{0}$	$\dot{0}$	$\dot{0}$	$\dot{0}$	$\dot{0}$
$\dot{1}$	$\dot{0}$	$\dot{1}$	$\dot{2}$	$\dot{3}$	$\dot{4}$
$\dot{2}$	$\dot{0}$	$\dot{2}$	$\dot{4}$	$\dot{1}$	$\dot{3}$
$\dot{3}$	$\dot{0}$	$\dot{3}$	$\dot{1}$	$\dot{4}$	$\dot{2}$
$\dot{4}$	$\dot{0}$	$\dot{4}$	$\dot{3}$	$\dot{2}$	$\dot{1}$

Après la présentation de l'objet, un peu de théorie à son sujet :
Proposition 14-8-73 : Pour tout $n \geq 1$, $\mathbf{Z}/n\mathbf{Z}$ est un anneau commutatif.
Démonstration : Elle est d'un ennui mortel, et ne présente aucune difficulté. Pour en faire un tout petit bout, montrons que l'addition est associative : soit x, y, z trois éléments de $\mathbf{Z}/n\mathbf{Z}$. On peut les écrire sous forme $x = \dot{a}$, $y = \dot{b}$, $z = \dot{c}$. Vu la définition de l'addition dans $\mathbf{Z}/n\mathbf{Z}$, on a alors $(x + y) + z = (\dot{a} + \dot{b}) + \dot{c} = \overbrace{a + b}^{\cdot} + \dot{c} = \overbrace{(a + b) + c}^{\cdot} = \overbrace{a + (b + c)}^{\cdot} = \dot{a} + \overbrace{b + c}^{\cdot} = \dot{a} + (\dot{b} + \dot{c}) = x + (y + z)$.
Et toutes les vérifications seraient de ce genre... Décidons donc de les laisser au lecteur.

•

Plus intéressant et légèrement plus subtil est le
Théorème 14-8-27 : Pour tout $n \geq 1$, $\mathbf{Z}/n\mathbf{Z}$ est un corps commutatif si et seulement si n est un nombre premier.
Démonstration : Montrons tour à tour les deux sens de l'équivalence.

Preuve de $\Rightarrow$. On va montrer cette implication par contraposition. Supposons donc que n n'est pas premier, et montrons que $\mathbf{Z}/n\mathbf{Z}$ n'est pas un corps commutatif (on verra même en passant que ce n'est même pas un anneau intègre).

* Traitons à part le cas "stupide" où n vaudrait 1. Dans ce cas, $\mathbf{Z}/\mathbf{Z}$ ne possède qu'un élément, donc n'est pas un corps commutatif.
* Examinons le cas significatif, où n n'est pas premier, mais n'est pas non plus égal à 1. Dans ce cas, on peut écrire $n = ab$, où $1 < a < n$ et $1 < b < n$. Dans l'anneau $\mathbf{Z}/n\mathbf{Z}$, on obtient alors la relation $\dot{n} = \dot{a}\dot{b}$, soit $\dot{a}\dot{b} = \dot{0}$. Pourtant, vu les inégalités vérifiées par a et b, ni $\dot{a}$ ni $\dot{b}$ n'est nul. $\mathbf{Z}/n\mathbf{Z}$ n'est donc pas intègre, et *a fortiori* n'est pas un corps commutatif.

On a bien prouvé dans les deux cas que $\mathbf{Z}/n\mathbf{Z}$ n'est pas un corps commutatif.

Preuve de $\Leftarrow$. Supposons n premier, et montrons que $\mathbf{Z}/n\mathbf{Z}$ est alors un corps commutatif.

* Nous savons déjà que la multiplication sur $\mathbf{Z}/n\mathbf{Z}$ est commutative.
* Comme $\mathbf{Z}/n\mathbf{Z}$ possède n éléments, il en possède au moins deux.
* Soit x un élément non nul de $\mathbf{Z}/n\mathbf{Z}$. On peut écrire $x = \dot{a}$ pour un $a \in \{1, \ldots, n-1\}$. Puisque n est premier, a ne possède d'autre diviseur positif commun avec n que 1 et donc a et n sont premiers entre eux. Il existe donc deux entiers $s, t \in \mathbf{Z}$ tels que $1 = sa + tn$. En passant aux classes d'équivalence, on obtient : $\dot{1} = \dot{s}\dot{a} + \dot{t}\dot{n}$, soit $\dot{1} = \dot{s}x + \dot{t}\dot{0} = \dot{s}x$.

  On a donc trouvé un inverse de x, à savoir $\dot{s}$.

$\mathbf{Z}/n\mathbf{Z}$ est donc bien un corps commutatif.

•

Remarque : On retiendra de cette démonstration la technique pratique de calcul de l'inverse d'un élément non nul de $\mathbf{Z}/n\mathbf{Z}$: écrire une identité de Bézout entre un représentant de cet élément et n, et redescendre aux classes d'équivalence.

Et voilà, on sait tout... Reste à donner quelques illustrations afin de convaincre de l'utilité de l'introduction de cette notion abstraite.

Exemple : Résoudre dans $\mathbf{Z}$ l'équation suivante, d'inconnue x :

$$24x + 5 \equiv 0 \quad [137].$$

On peut traiter cet exemple avec ou sans usage de $\mathbf{Z}/137\mathbf{Z}$. Faisons les deux successivement ; on constatera que les énoncés simples sur les propriétés algébriques de $\mathbf{Z}/137\mathbf{Z}$ remplacent avantageusement les techniques, il est vrai elles aussi simples, d'arithmétique classique.

Première résolution : (sans utiliser $\mathbf{Z}/137\mathbf{Z}$).

Remarquons que 137 est premier, et donc que 137 et 24 sont premiers entre eux ; cherchons à écrire une identité de Bézout entre 137 et 24 ; en utilisant l'algorithme décrit plus haut, on découvre que :

$$1 = 40 \times 24 - 7 \times 137,$$

d'où on déduit (par une simple multiplication par 5) que :

$$5 = 200 \times 24 - 35 \times 137.$$

Reportons cette identité dans l'équation, qui devient donc :

$$\begin{array}{lcl} & 24x + 200 \times 24 - 35 \times 137 \equiv 0 \quad [137] & \\ \iff & 24(x+200) \equiv 0 \quad [137] & \\ \iff & 137 \text{ divise } 24(x+200) & \\ \iff & 137 \text{ divise } x+200 & \text{(en utilisant le lemme de Gauss)} \\ \iff & x + 200 \equiv 0 \quad [137] & \\ \iff & x \equiv -200 \quad [137] & \\ \iff & x \equiv 74 \quad [137]. & \end{array}$$

Deuxième résolution : (avec $\mathbf{Z}/137\mathbf{Z}$).

Remarquons que 137 est premier, et donc que $\mathbf{Z}/137\mathbf{Z}$ est un corps commutatif. Faisons tous les calculs dans ce corps.

L'équation proposée se réécrit :

$$\begin{array}{lc} & \dot{24}\dot{x} + \dot{5} = \dot{0} \\ \iff & \dot{24}\dot{x} = -\dot{5} \\ \iff & \dot{x} = -\dot{5}(\dot{24})^{-1}. \end{array}$$

Calculons donc $(\dot{24})^{-1}$; pour cela nous connaissons la bonne méthode : écrire une identité de Bézout entre 24 et 137, à savoir :

$$1 = 40 \times 24 - 7 \times 137,$$

puis redescendre aux classes d'équivalence dans $\mathbf{Z}/137\mathbf{Z}$:

$$\dot{1} = \dot{40} \cdot \dot{24},$$

soit : $(\dot{24})^{-1} = \dot{40}$.

On en conclut que l'équation proposée équivaut à :

$$\dot{x} = -\dot{5}(\dot{24})^{-1} = -\dot{5} \times \dot{40} = -\dot{200} = \dot{74}.$$

Exemple : Résoudre dans $\mathbf{Z}$ l'équation suivante, d'inconnue x :

$$x^4 \equiv 81 \quad [73].$$

Là aussi, écrire deux solutions serait possible, mais celle utilisant $\mathbf{Z}/73\mathbf{Z}$ est tellement plus agréable à écrire que je m'en contenterai.

L'équation s'écrit, dans $\mathbf{Z}/73\mathbf{Z}$:

$$\begin{array}{lc} & \dot{x}^4 - \dot{81} = \dot{0} \\ \iff & (\dot{x}^2)^2 - \dot{9}^2 = \dot{0} \\ \iff & (\dot{x}^2 - \dot{9})(\dot{x}^2 + \dot{9}) = \dot{0} \end{array}$$

$$\begin{array}{ll}
\iff & (\dot{x}^2 - \dot{9})(\dot{x}^2 - \dot{64}) = \dot{0} \\
\iff & (\dot{x} - \dot{3})(\dot{x} + \dot{3})(\dot{x} - \dot{8})(\dot{x} + \dot{8}) = \dot{0} \\
\iff & (\dot{x} - \dot{3})(\dot{x} - \dot{70})(\dot{x} - \dot{8})(\dot{x} - \dot{65}) = \dot{0} \\
\iff & \dot{x} = \dot{3} \text{ ou } \dot{x} = \dot{8} \text{ ou } \dot{x} = \dot{65} \text{ ou } \dot{x} = \dot{70} \qquad (\mathbf{Z}/73\mathbf{Z} \text{ étant un corps commutatif, donc intègre})
\end{array}$$

Exemple : Résoudre dans $\mathbf{Z}$ l'équation suivante, d'inconnue x :

$$x^{17} \equiv 3 \quad [19].$$

Là encore, on ne saurait trop recommander le passage dans $\mathbf{Z}/19\mathbf{Z}$. L'équation s'écrit dès lors : $\dot{x}^{17} = \dot{3}$. Notons a l'inconnue auxiliaire $a = \dot{x}$ et remarquons que $\dot{0}^{17} \neq \dot{3}$. On peut donc ne chercher à résoudre $a^{17} = \dot{3}$ que dans $(\mathbf{Z}/19\mathbf{Z}) \setminus \{\dot{0}\}$.

Mais pour des $a \neq \dot{0}$, $a^{17} = \dot{3} \iff a^{18} = \dot{3}a$. Maintenant, pour tout a dans le groupe multiplicatif $(\mathbf{Z}/19\mathbf{Z}) \setminus \{\dot{0}\}$, on sait que l'ordre de a, qui est le nombre d'éléments du groupe $<a>$, divise le nombre d'éléments de $(\mathbf{Z}/19\mathbf{Z}) \setminus \{\dot{0}\}$, c'est-à-dire 18.

Ainsi, pour tout a de $(\mathbf{Z}/19\mathbf{Z}) \setminus \{\dot{0}\}$, $a^{18} = \dot{1}$. L'équation étudiée se simplifie donc grandement en $1 = \dot{3}a \iff a = (\dot{3})^{-1}$. Sa résolution se ramène donc à la recherche de l'inverse de $\dot{3}$ dans $\mathbf{Z}/19\mathbf{Z}$; on écrit alors une relation de Bézout : $13 \times 3 - 2 \times 19 = 1$ et on en déduit que $(\dot{3})^{-1} = \dot{13}$.

Finalement les solutions de l'équation initiale sont donc les x congrus à 13 modulo 19.

Exemple : Résoudre dans $\mathbf{Z}$ l'équation suivante, d'inconnue x :

$$x^{14} \equiv 1 \quad [19].$$

Ce sont les mêmes idées que dans l'exemple précédent qui font marcher cet exercice, en un peu plus astucieux encore.

Comme dans l'exemple prédédent, on commence par passer dans $\mathbf{Z}/19\mathbf{Z}$, où l'équation s'écrit dès lors : $\dot{x}^{14} = \dot{1}$. On note $a = \dot{x}$, on remarque que $\dot{0}$ n'est pas solution, et on décide donc de résoudre $a^{14} = 1$ dans $(\mathbf{Z}/19\mathbf{Z}) \setminus \{\dot{0}\}$.

Maintenant, on remarque que pour tout a de $(\mathbf{Z}/19\mathbf{Z}) \setminus \{\dot{0}\}$, dire que $a^{14} = \dot{1}$ équivaut à dire que l'ordre de a divise 14. Par ailleurs, comme dans l'exemple précédent, pour tout a de $(\mathbf{Z}/19\mathbf{Z}) \setminus \{\dot{0}\}$, l'ordre de a divise 18. Ainsi, l'ordre de a divise 14 si et seulement s'il divise 14 et 18, donc si et seulement s'il divise $\mathrm{PGCD}(14, 18) = 2$.

On a donc montré que pour tout a de $(\mathbf{Z}/19\mathbf{Z}) \setminus \{\dot{0}\}$, $a^{14} = \dot{1} \iff a^2 = \dot{1}$.

Cette nouvelle équation est alors très facile à résoudre : $a^2 = \dot{1} \iff (a + \dot{1})(a - \dot{1}) = \dot{0} \iff a = \dot{1}$ ou $a = -\dot{1} = \dot{18}$.

Finalement, les solutions de l'équation initiale sont donc les x congrus à 1 ou à 18 modulo 19.

Chapitre 15 - Espaces vectoriels généraux

On va généraliser ce qu'on sait sur $\mathbf{R}^n$ et ses sous-espaces vectoriels. La généralisation rencontrera deux types de difficulté : utiliser d'autres nombres que des réels ; manipuler des espaces pouvant être de dimension infinie (ou, ce qui sera souvent indispensable, tout au moins manipuler des systèmes infinis dans des espaces de dimension finie ou non).

Dans tout le chapitre, $\mathbf{K}$ désigne un corps commutatif.

1 - Définition des espaces vectoriels

Définition 15-1-118 : Soit E un ensemble ; on dispose sur cet ensemble d'une opération (notée additivement) et on dispose par ailleurs d'une application $g : \mathbf{K} \times E \to E$ (notée multiplicativement : on note λx au lieu de $g(\lambda, x)$).
On dit que E est un **espace vectoriel** lorsque :
* E est un groupe commutatif (pour l'addition).
* Pour tout vecteur e de E, $1e = e$ (1 désignant le neutre de la multiplication de $\mathbf{K}$).
* Pour tous scalaires λ, μ de $\mathbf{K}$ et tout vecteur e de E, $(\lambda\mu)e = \lambda(\mu e)$.
* Pour tous scalaires λ, μ de $\mathbf{K}$ et tout vecteur e de E, $(\lambda + \mu)e = \lambda e + \mu e$.
* Pour tout scalaire λ de $\mathbf{K}$ et tous vecteurs e, f de E, $\lambda(e + f) = \lambda e + \lambda f$.

Exemple : $\mathbf{K}$ est un espace vectoriel sur lui-même (en utilisant comme multiplication g de la définition d'espace vectoriel l'opération multiplication du corps commutatif $\mathbf{K}$).

Cet exemple mérite qu'on s'y arrête quelques lignes, dans le cas particulier important et encore accessible où $\mathbf{K} = \mathbf{C}$ –en anticipant légèrement sur les définitions qui suivront. Le même ensemble $E = \mathbf{C}$ peut être considéré comme espace vectoriel sur $\mathbf{C}$ ou sur $\mathbf{R}$ (avec la même addition dans les deux cas). Si on le voit comme espace vectoriel sur $\mathbf{R}$ (c'est le plus facile, car l'algèbre linéaire sur $\mathbf{R}$ se conforme à notre intuition géométrique au quotidien), une base de $\mathbf{C}$ comme $\mathbf{R}$-espace vectoriel est $(1, i)$. Mais si $\mathbf{C}$ est pensé comme espace vectoriel sur $\mathbf{C}$, $(1, i)$ n'est plus une famille libre, puisque en prenant $\lambda = i$ et $\mu = -1$, $\lambda 1 + \mu i = 0$. Ce qui est désormais une base de $\mathbf{C}$, c'est le système (1) formé d'un seul vecteur, et le $\mathbf{C}$-espace vectoriel $\mathbf{C}$ est devenu une droite !

On pourra faire le rapprochement entre cette subtilité conceptuelle et l'erreur "classique" des étudiants qui, placés devant une identité telle que $\dfrac{1}{2} + i\dfrac{\sqrt{3}}{2} = x + iy$ en déduisent que $\dfrac{1}{2} = x$ et $\dfrac{\sqrt{3}}{2} = y$ sans vérifier préalablement que x et y sont **réels**. Bien évidemment, s'ils ne le sont pas, les étudiants sont alors précipités dans un gouffre : ils avaient confondu liberté sur $\mathbf{R}$ et liberté sur $\mathbf{C}$.

L' "exemple fondamental"

De même que pour les groupes me paraissait fondamental le groupe des bijections d'un ensemble – dont tant de groupes importants sont des sous-groupes– pour les espaces vectoriels me paraît essentiel l'espace des applications vers un espace –dont tant d'espaces seront des sous-espaces.

Voici les opérations sur cet "espace fondamental".

Définition 15-1-119 : Soit A un ensemble et E un $\mathbf{K}$-espace vectoriel ; soit f et g deux applications de A vers E. La **somme** de f et g est l'application $f + g$ définie pour tout $a \in A$ par $(f + g)(a) = f(a) + g(a)$. Soit λ un scalaire ; le **produit** de f par λ est l'application λf définie par $(\lambda f)(a) = \lambda f(a)$.

Les étudiants observateurs remarqueront qu'à force de généralisation, cette définition rend obsolètes –comme cas particuliers– les définitions 8-2-82, 8-2-83, 5-1-54 (en tant qu'elle parle d'addition), et même, si on est observateur, 3-1-37 et 3-1-38.

Proposition 15-1-74 : Muni de ces addition et multiplication externe, l'ensemble E^A est un espace vectoriel sur $\mathbf{K}$.

Démonstration : D'une facilité absolue et d'un ennui profond.

•

Un cas particulier déjà connu, (du moins pour $\mathbf{K} = \mathbf{R}$) est l'espace $\mathbf{K}^n$ des applications de $\{1, \ldots, n\}$ vers $\mathbf{K}$. La définition de la base canonique de $\mathbf{R}^n$ se transposera à l'identique dans $\mathbf{K}^n$.

2 - Ce qui se conserve sans rien changer du cours sur $\mathbf{R}^n$

Personne, je l'espère, ne lira attentivement ce paragraphe, qui n'est qu'un outil de référence :

* Dans le chapitre 6, la section 0 ne se rapporte qu'à $\mathbf{R}^n$; dans la section 1, il faut préciser que le mot "vecteur" sera désormais utilisé pour les éléments d'un espace vectoriel et le mot scalaire pour les éléments du corps commutatif $\mathbf{K}$. ; que la notation $e_1 + \cdots + e_n$ sera utilisée dans tous les espaces vectoriels. La section 2 reste en attente. La section 3 (sous-espaces) est récupérable en remplaçant $\mathbf{R}^n$ par un espace vectoriel E_0 (mais pas pour l'instant la proposition qui la termine, qui parle d'espace engendré). La section 4 (système générateur ou libre) est laissée pour un peu plus loin, et donc aussi les section 5 et 6 qui vont avec. J'avais déjà dit que la section 7 (base canonique) se généralisait à $\mathbf{K}^n$ mais pas à n'importe quel espace vectoriel. Enfin la section 8 (intersection et somme de sous-espaces) se généralise sans bug.

* Je referai le point sur le chapitre 7 (dimension) plus loin.

* Dans le chapitre 11 (applications linéaires) on récupère à l'identique tout ce qui ne parle pas de "dimension finie" en modifiant dans le chapeau du chapitre –que j'avais habilement déclaré provisoire– la définition de "soit E un espace vectoriel" qui voudra désormais dire "soit E un espace vectoriel".

3 - Le concept de famille

Pour travailler sur des espaces qui pourront ne plus être de dimension finie, on va être amené à ne plus utiliser des k-uplets mais des systèmes plus généraux susceptibles de contenir une infinité de vecteurs –dans les applications pratiques en TD ce seront généralement des suites, mais ça ne coûte pas plus cher de définir des familles infinies générales.

Définition 15-3-120 : Soit I et X deux ensembles. Une **famille** d'éléments de X indexée par I est une application de I vers X.

Il ne s'agit donc de rien de plus que d'une bête application ! Simplement la problématique n'est plus la même et pour des raisons peut-être plus historiques que profondes on utilise des notations bien différentes.

Ainsi dans le contexte des applications on note f une application et $f(i)$ sa valeur en un élément $i \in I$. Dans le contexte des familles, on note $(x_i)_{i\in I}$ (ou (x_i) quand il n'y a pas de risque de confusion quant à l'ensemble de départ) la famille et x_i sa valeur en un indice $i \in I$.

L'archétype de la famille est évidemment la suite, que vous connaissez déjà : c'est une famille indexée par $\mathbf{N}$.

4 - Familles et opérations

La principale idée que je cherche à faire passer dans cette section est une mise en garde : on ne peut additionner qu'un nombre **fini** de vecteurs. Des techniques pour faire des additions infinies apparaîtront en analyse (l'intégration, en un sens, en est une...) mais ne s'appliqueront que dans des contextes bien spécifiques et en aucun cas dans des cadres d'espaces vectoriels "généraux".

Il faut donc se préparer à bien comprendre quand on peut ajouter et quand on ne le peut pas.

Définition 15-4-121 : Soit E un $\mathbf{K}$-espace vectoriel (un groupe commutatif suffirait) et soit $(e_i)_{i\in I}$ une famille **finie** de vecteurs de E. On définit la **somme** des e_i par les formules :

* $\displaystyle\sum_{i\in\emptyset} e_i = 0$.

* Pour $a \in I$, $\displaystyle\sum_{i\in I} e_i = \sum_{i\in I\setminus\{a\}} e_i + e_a$.

Il faut alors justifier qu'on a bien donné une définition cohérente... cette définition ayant une allure récursive, ce n'est guère qu'une explication, la "bonne" façon d'écrire la définition serait de la présenter comme une définition par récurrence sur le cardinal de I. La seconde mise au point nécessaire, une fois la somme finie de n éléments définie, est de s'assurer que la définition qu'on a donnée pour la somme de $n+1$ éléments est cohérente. Cela nécessite de vérifier que la somme qu'on a définie ne dépend pas du choix qu'on a fait du a qu'on a particularisé. En d'autres termes, il faut vérifier que lorsqu'on prend deux éléments distincts a et b de I, $\displaystyle\sum_{i\in I\setminus\{a\}} e_i + e_a = \sum_{i\in I\setminus\{b\}} e_i + e_b$, ce qui découle facilement de la commutativité et de l'associativité de l'addition.

Cette définition a une conséquence immédiate et utilisée implicitement tout au long des démonstrations (elle se démontrerait par récurrence sur le cardinal de J) : si I et J sont finis disjoints, et $(e_i)_{i\in I\cup J}$ est une famille indexée par $I\cup J$, $\displaystyle\sum_{i\in I\cup J} e_i = \sum_{i\in I} e_i + \sum_{i\in J} e_i$.

Une fois cette définition posée, on peut donner des versions "faciles" des définitions de combinaison linéaire d'une famille, de famille génératrice, de famille libre.

Définition 15-4-122 : Soit $(e_i)_{i\in I}$ une famille de vecteurs d'un espace E, et soit e un vecteur de E. On dit que e est une **combinaison linéaire** de (e_i) s'il existe un sous-ensemble **fini** I_1 de I et une famille de scalaires $(\lambda_i)_{i\in I_1}$ indexée par I_1 telle que $e = \displaystyle\sum_{i\in I_1} \lambda_i e_i$.

Définition 15-4-123 : Soit $(e_i)_{i\in I}$ une famille de vecteurs d'un espace E, on dit que (e_i) **engendre** E lorsque tout e de E est combinaison linéaire des (e_i), c'est-à-dire explicitement si pour tout e de E il existe un sous-ensemble **fini** I_1 de I et une famille de scalaires $(\lambda_i)_{i\in I_1}$ indexée par I_1 telle que $e = \displaystyle\sum_{i\in I_1} \lambda_i e_i$.

Définition 15-4-124 : Soit $(e_i)_{i\in I}$ une famille de vecteurs d'un espace E, on dit que (e_i) est **libre** lorsque pour tout sous-ensemble **fini** I_1 de I et toute famille de scalaires $(\lambda_i)_{i\in I_1}$ indexée par I_1 telle que $\displaystyle\sum_{i\in I_1} \lambda_i e_i = 0$, tous les λ_i sont nuls.

Une autre opération utile occasionnellement concernera des familles d'**ensembles**. Il s'agit de pratiquer des intersections ou des réunions d'ensembles en nombre infini –c'est un peu hors sujet ici mais il faut bien les définir quelque part.

Définition 15-4-125 : Soit $(A_i)_{i\in I}$ une famille d'ensembles. On définit la **réunion** de (A_i) comme l'ensemble des x pour lequels il existe un indice $i \in I$ tel que $x \in A_i$.

Notation 15-4-54 : Cette réunion est notée $\displaystyle\bigcup_{i\in I} A_i$.

Définition 15-4-126 : Soit $(A_i)_{i\in I}$ une famille d'ensembles (avec $I \neq \emptyset$). On définit l'**intersection** de (A_i) comme l'ensemble des x tels que pour tout indice $i \in I$, $x \in A_i$. [Si tous les ensembles intervenant dans un problème sont des parties d'un même gros ensemble Ω, on pourra sans danger compléter cette définition en considérant que l'intersection d'une famille vide est Ω.]

Notation 15-4-55 : Cette intersection est notée $\displaystyle\bigcap_{i\in I} A_i$.

Pas d'inquiétude, ces notions ont le même sens intuitif que celles qui sont déjà bien connues, et les diverses formules qui peuvent avoir l'air vraisemblables les concernant sont vraies.

Refaisons une petite visite –à ne pas lire, sauf pour les masochistes– du cours du premier semestre concernant ces questions : dans le chapitre 6, la section 2 est désormais remplacée (on se convaincra bien sûr que dans le cas particulier où $I = \{1, \ldots, n\}$ les définitions gardent le même sens) ; la dernière proposition de la section 3 est désormais généralisée et redémontrée. La section 4 est généralisée et revue (les mots "lié" et "base" se récupérant de façon évidente). Pour la section 5, la proposition 3-5-12 reste vraie –vérifiez que vous savez la prouver–, les deux propositions suivantes aussi si on précise bien ce que veut dire ajouter ou retrancher un vecteur à une famille. La fin de la section, qui ne concernait que des systèmes, se recopie sans modification autre que remplacer $\mathbf{R}^n$ par E quelconque. La section 6 aussi se récupère dans tout espace possédant des bases finies.

5 - Un petit complément : espace engendré par une partie

Sans que la notion ne soit très importante, il peut être confortable de parler d'ensemble générateur plutôt que de famille génératrice dans divers contextes. C'est ici un prétexte pour présenter une de ces petites astuces sans profondeur et pourtant terriblement simplificatrices.

Une première méthode qui vient à l'esprit pour introduire ce concept serait de reprendre à zéro toutes les définitions qu'on a données et, comme on a défini la somme d'une famille finie de vecteurs, définir la somme d'un ensemble fini de vecteurs par récurrence sur le nombre de ceux-ci, et ainsi de suite... Il y aurait d'ailleurs des chausse-trapes sur notre route.

Plus ingénieux est de ramener l'étude des ensembles de vecteurs à l'étude des familles de vecteurs en introduisant la stupide

Définition 15-5-127 : Soit A un ensemble. On appelle **famille auto-indexée** par A la famille $(x)_{x\in A}$ (soit plus formellement encore l'application identique de A).

L'idée est d'étiqueter chaque élément de A en écrivant sur l'étiquette non un numéro d'identification (je ne suis pas un numéro, je ne suis pas un matricule) mais son propre nom.

Avec cette définition, on peut étendre tout ce qui est défini pour des familles à des ensembles : il suffit d'appliquer la définition telle qu'elle a été écrite pour une famille à la famille auto-indexée correspondant à l'ensemble.

L'intérêt ici de cette gymnastique est de pouvoir énoncer la

Proposition 15-5-75 : Le sous-espace engendré par une partie A d'un espace vectoriel E est le plus petit sous-espace de E contenant A.

(N.B. : la définition formelle de "plus petit" sera écrite dans une section de compléments relatifs aux relations d'ordre qui viendra prochainement).

Démonstration : On sait déjà que le sous-espace engendré par A est un sous-espace ; il est totalement évident qu'il contient A. Reste à montrer qu'il est inclus dans tout sous-espace contenant A.

Soit donc G un tel sous-espace ; notons F le sous-espace engendré par A : nous devons donc montrer que $F\subset G$. Pour ce faire, prenons un vecteur y de F, c'est-à-dire une combinaison linéaire de $(x)_{x\in A}$. Par définition des "combinaisons linéaires", il existe donc un nombre fini de vecteurs $x_1,\dots,x_n$ de A et un nombre fini de scalaires $\lambda_1,\dots,\lambda_n$ tels que $y=\lambda_1x_1+\cdots+\lambda_nx_n$. G étant un sous-espace vectoriel, et tous les x_i étant dans A donc dans G, y est aussi dans G.

•

Une agréable application (toutefois assez peu utile en pratique) sera une occasion de rappeler qu'il ne faut pas confondre somme et réunion !

Proposition 15-5-76 : Soit F et G deux sous-espaces d'un $\mathbf{K}$-espace vectoriel E. Alors $F+G$ est le sous-espace engendré par $F\cup G$.

Démonstration : Nous pouvons utiliser la caractérisation qui précède : il est clair que $F+G$ est un sous-espace contenant $F\cup G$. Il nous reste à montrer que c'est le plus petit, c'est-à-dire qu'il est inclus dans tout sous-espace H de E contenant $F\cup G$. Soit donc H un tel sous-espace, et soit x un vecteur de $F+G$. On peut écrire $x=y+z$, où $y\in F$, donc $y\in H$, et $z\in G$, donc $z\in H$. On en déduit que $x=y+z\in H$.

•

6 - Nouvelle visite à la dimension

Il reste à comprendre comment le chapitre 7 –celui concernant la dimension– s'adapte à un cadre plus général que celui de $\mathbf{R}^n$. Et là les choses cessent d'être vraiment simples. L'espace $\mathbf{R}^n$ possède en effet une propriété fort pratique : une base nous saute aux yeux (la base canonique). Au moment de travailler dans un espace abstrait, cette première base va nous manquer, et il faudra quelques précautions conceptuelles ; en fait, dès que les espaces ne possèdent pas de base finie –et cela arrive– les choses sont significativement plus techniques que ce que nous avons rencontré jusqu'à présent, et donc hors d'atteinte dans un cours de DEUG.

Reprenons pas à pas le cours sur la dimension. Sur la première section ("le nœud des démonstrations"), nous ne trébuchons pas : elle se généralise très bien à n'importe quel espace sur n'importe quel corps commutatif. Nous disposons donc du lemme d'échange, et surtout du principe (énoncé seulement pour des familles finies) qui assure qu'une famille libre est plus courte au sens large qu'une famille génératrice.

La section 2 se généralise aussi : si un espace possède deux bases finies, elles ont le même nombre de vecteurs. On notera qu'on peut aussi prouver pour pas plus cher qu'un espace ne peut posséder simultanément une base finie et une base infinie.

La section 3 reste vraie –avec les mêmes preuves– en tant qu'elle concerne des familles libres ou génératrices finies. Elle ne serait pas très difficile à adapter même pour des familles infinies, mais je n'en ai pas besoin pour continuer.

En revanche, avant d'aller plus loin, il va être indispensable d'introduire un nouveau concept.

Définition 15-6-128 : Soit E un $\mathbf{K}$-espace vectoriel. On dit que E est **de type fini** (ou **de dimension finie**) lorsqu'il possède une famille génératrice finie.

Remarque : Le vocabulaire le plus courant est le second ("de dimension finie") mais il est un peu dangereux, car nous ne savons pas encore qu'on peut associer à ces espaces un concept de dimension. Pour cette bonne raison, j'utiliserai le terme “de type fini” jusqu'à la fin de ce chapitre sur les dimensions, puis reprendrai l'habitude de parler de “dimension finie” quand la dimension n'aura plus de secrets pour nous.

Ce concept étant posé, nous pouvons énoncer le théorème de la base incomplète dans un cadre le généralisant :

Théorème 15-6-28 : Soit E un $\mathbf{K}$-espace vectoriel de type fini. Toute famille libre de E est finie, et peut être prolongée en une base de E par l'adjonction de nouveaux vecteurs de E.

Démonstration : C'est la même que pour les sous-espaces de $\mathbf{R}^n$; simplement l'arrêt du processus d'adjonction de vecteurs n'est plus ici garanti par l'existence de la base canonique de $\mathbf{R}^n$ mais par celle de la famille génératrice finie de E dont on a supposé l'existence.

•

Il est amusant de constater que le théorème “symétrique” qui était si facile à prouver quand les seules familles connues étaient finies dérape discrètement dès lors que nous connaissons le concept de famille génératrice infinie. Même dans un espace de type fini, la famille génératrice dont nous voudrions extraire une base peut être infinie, et la technique d'enlever des vecteurs un à un ne fonctionnera plus. Le résultat reste toutefois vrai, et de démonstration pas trop difficile, mais il faut la reprendre.

Théorème 15-6-29 : Soit E un $\mathbf{K}$-espace vectoriel de type fini. De toute famille génératrice (même infinie !) de E, on peut extraire une base de E par suppression de certains vecteurs. (Plus formellement : soit $(x_i)_{i\in I}$ une famille génératrice de E. Il existe un sous-ensemble (forcément fini) I_1 de I tel que $(x_i)_{i\in I_1}$ soit une base de E).

Démonstration : Elle est un peu technique, quoique pas vraiment novatrice par rapport à ce que nous avons fait jusqu'à présent. Sa lecture ne me semble instructive que pour les étudiants vraiment motivés !

Cette précaution oratoire étant prise, introduisons un concept nouveau mais pas bien compliqué quoique lourd : pour un système libre $(f_1,\dots,f_p)$ extrait de (x_i), on dira que $(f_1,\dots,f_p)$ est libre maximal relativement à (x_i) lorsque pour tout vecteur v figurant dans la famille (x_i), $(f_1,\dots,f_p,v)$ n'est plus libre.

On peut alors avec ce nouveau concept reprendre la proposition 4-3-52, avec cette fois l'énoncé suivant : un système libre maximal relativement à (x_i) est une base de E. La démonstration reposera sur les mêmes techniques : soit un tel système $(f_1,\dots,f_p)$ libre maximal relativement à (x_i). Alors pour tout vecteur v figurant dans la famille (x_i), la famille $(f_1,\dots,f_p,v)$ n'étant plus libre, v est une combinaison linéaire de $(f_1,\dots,f_p)$. Mais tous les vecteurs de E étant eux-mêmes combinaisons linéaires d'un nombre fini de vecteurs figurant dans (x_i), ils sont donc combinaisons linéaires de $(f_1,\dots,f_p)$, qui est donc génératrice.

La démonstration se termine comme celle du théorème de la base incomplète : on part du système () et s'il n'est pas maximal relativement à (x_i) on lui ajoute un des vecteurs de (x_i) et ainsi de suite ; on ne peut continuer indéfiniment car on est limité dans la croissance de ces systèmes libres par le nombre fini d'éléments d'un système générateur de E. Quand on s'arrête, on a bien la base annoncée. La surprise étant qu'on l'a obtenue par adjonctions successives et non, comme au semestre précédent, par ablations successives.

•

Plus de problème pour le reste... Comme pour les sous-espaces de $\mathbf{R}^n$, on montrera pour les espaces vectoriels généraux le

Théorème 15-6-30 : Tout espace vectoriel de type fini possède des bases.

La théorie de la dimension marche finalement avec les mêmes énoncés et les mêmes preuves ; la formule de Grassman reste vraie ; le paragraphe sur les sommes directes aussi –on notera simplement que le théorème liant sommes directes et dimensions ne concernera que des espaces de dimension finie. La proposition qui en a été déduite (caractérisation pour deux sous-espaces seulement) est donc prouvée pour des espaces de dimension finie. Elle reste vraie pour des espaces de dimension quelconque (c'est un exercice facile...)

7 - Un premier exemple d'espace abstrait : espaces d'applications linéaires

Très brièvement, on constate en relisant ce qu'on sait déjà sur les applications linéaires (stabilité par addition, ou par multiplication par un scalaire) que l'ensemble $\mathcal{L}(E,F)$ des applications linéaires d'un $\mathbf{K}$-espace vectoriel vers un $\mathbf{K}$-espace vectoriel F est un sous-espace vectoriel du $\mathbf{K}$-espace vectoriel F^E. Nous verrons au chapitre suivant qu'il est de dimension finie quand E et F sont eux-même de dimension finie.

On notera que même quand E et F sont de bons sous-espaces vectoriels de $\mathbf{R}^n$ (voire tous deux égaux à $\mathbf{R}^n$) l'espace $\mathcal{L}(E,F)$ ne se laisse pas si facilement interpréter comme un sous-espace d'un $\mathbf{R}^k$. Le concept d'espace abstrait (même de dimension finie) montre donc son intérêt pour affiner nos connaissances sur $\mathcal{L}(E,F)$ –nous connaitrons bientôt sa dimension.

Chapitre 16 - Matrices

Les matrices sont des tableaux rectangulaires de réels (ou plus généralement d'éléments d'un corps commutatif $\mathbf{K}$). Annonçons donc tout de suite : dans tout le chapitre, $\mathbf{K}$ est un corps commutatif fixé, et tous les espaces vectoriels considérés sont des espaces vectoriels sur $\mathbf{K}$; toutes les matrices considérées seront aussi, sauf précision contraire, à coefficients dans $\mathbf{K}$).

Une idée intéressante à retenir est que les matrices font des apparitions dans des contextes externes à l'algèbre linéaire : analyse de données, mécanique, etc... Il y aura donc des allers-retours tout à fait plaisants : les matrices nous serviront à démontrer des résultats d'algèbre linéaire pure (par exemple à déterminer la dimension de $\mathcal{L}(E,F)$), mais réciproquement les théorèmes prouvés en algèbre linéaire pourront être appliqués à des matrices dans des contextes mathématiques tout autres –voire des contextes non mathématiques.

1 - Définitions et notations

Définition 16-1-129 : (Version informelle) Une **matrice** (m,n) à coefficients dans $\mathbf{K}$ est un tableau rectangulaire d'éléments de $\mathbf{K}$ avec m lignes et n colonnes. (Et version formelle : une matrice (m,n) à coefficients dans $\mathbf{K}$ est une famille indexée par l'ensemble $\{1,\ldots,m\}\times\{1,\ldots,n\}$ où m et n sont deux entiers naturels fixés ; quand ce sera pratique, on pourra aussi autoriser des indexations commençant à 0, par exemple pour traiter des polynômes dont le degré commence à 0.).

Convention de notation Quand je parlerai d'une matrice notée par une lettre majuscule, disons A, je sous-entendrai que le terme en i-ème ligne et j-ème colonne de A est noté a_{ij}, avec la minuscule correspondante. Cette règle a une exception un peu bizarre, remontant à Kronecker : le i,j-ème coefficient de la matrice I est traditionnellement noté δ_{ij}.

Notation 16-1-56 : L'ensemble des matrices à m lignes et n colonnes à coefficients dans $\mathbf{K}$ est noté $\mathcal{M}_{mn}(\mathbf{K})$. Lorsque $m=n$, c'est-à-dire lorsque les matrices sont carrées, on abrègera cette notation en $\mathcal{M}_n(\mathbf{K})$.

Définition 16-1-130 : La **somme** de deux matrices (m,n) A et B est la matrice (m,n) C définie pour tout (i,j) avec $1\leq i\leq m$, $1\leq j\leq n$ par : $c_{ij}=a_{ij}+b_{ij}$; le **produit** d'une matrice (m,n) A par un scalaire λ est la matrice (m,n) D définie pour tout (i,j) avec $1\leq i\leq m$, $1\leq j\leq n$ par : $d_{ij}=\lambda a_{ij}$.

Les étudiants les plus observateurs remarqueront que je me répète encore, et que ces définitions font double emploi avec la définition 15-1-119.

Définition 16-1-131 : Le **produit** d'une matrice (m,n) A par une matrice (n,p) B est la matrice (m,p) C définie pour tout (i,k) avec $1\leq i\leq m$, $1\leq k\leq p$ par :

$$c_{ik}=\sum_{j=1}^{n}a_{ij}b_{jk}.$$

On remarquera tout de suite sur des exemples très simples que cette multiplication n'est pas commutative. D'ailleurs, si $m\neq p$, le produit BA n'existe même pas, alors que le produit AB existe, et si $m=p$ mais $m\neq n$, le produit BA est une matrice (n,n) alors que le produit AB est une matrice (m,m) !

En revanche, cette multiplication est “associative” (je mets des guillemets car ce n'est pas exactement une opération : elle s'applique entre matrices de taille variable.) Précisément , on a l'énoncé :

Proposition 16-1-77 : Pour tous entiers m,n,p,q et toutes matrices A de taille (m,n), B de taille (n,p) et C de taille (p,q), $(AB)C=A(BC)$.

Démonstration : Notons $D=AB$, $E=BC$, $F=(AB)C$ et $G=A(BC)$.

Alors pour tous i,l avec $1\leq i\leq m$, $1\leq l\leq q$, on a :

$$f_{il}=\sum_{k=1}^{p}d_{ik}c_{kl}=\sum_{k=1}^{p}(\sum_{j=1}^{n}a_{ij}b_{jk})c_{kl}=\sum_{\substack{1\leq j\leq n\\1\leq k\leq p}}a_{ij}b_{jk}c_{kl}=\sum_{j=1}^{n}a_{ij}(\sum_{k=1}^{p}b_{jk}c_{kl})=\sum_{j=1}^{n}a_{ij}e_{jl}=g_{il}$$

donc $F=G$.

•

Définition 16-1-132 : La **matrice identité** (n,n) est la matrice I_n définie pour tout (i,j) avec $1 \leq i \leq n$, $1 \leq j \leq n$: $\delta_{ij} = 0$ pour $i \neq j$ et $\delta_{ii} = 1$.

On remarque sans mal que pour toutes matrices A ou B pour lesquelles le produit a un sens, $AI = A$ et $IB = B$.

Proposition 16-1-78 : Avec ces opérations et cet élément unité, $\mathcal{M}_n(\mathbf{K})$ est un anneau ; $\mathcal{M}_{mn}(\mathbf{K})$ est un $\mathbf{K}$-espace vectoriel.

Démonstration : Fastidieuse et facile, l'associativité a été faite proprement, ne m'en demandez pas plus. •

Définition 16-1-133 : Pour m et n fixés, on appelle **matrices élémentaires** les matrices E_{ab} $(1 \leq a \leq m$, $1 \leq b \leq n)$ dont les coefficients sont nuls, sauf le coefficient de la a-ème ligne, b-ème colonne qui vaut 1.

Proposition 16-1-79 : La famille $(E_{ij})_{\substack{1\leq i\leq m \\ 1\leq j\leq n}}$ est une base de $\mathcal{M}_{mn}(\mathbf{K})$ (qu'on appellera "base canonique").

Démonstration : Remarquons que pour toute matrice (m,n) Λ, $\sum_{\substack{1\leq i\leq m \\ 1\leq j\leq n}} \lambda_{ij}E_{ij} = \Lambda$. Dès lors pour toute matrice A et toute famille de scalaires Λ, $A = \sum_{\substack{1\leq i\leq m \\ 1\leq j\leq n}} \lambda_{ij}E_{ij} \iff A = \Lambda$. De ce fait, A s'écrit de façon unique dans la famille (E_{ij}) qui est donc une base de $\mathcal{M}_{mn}(\mathbf{K})$. •

Corollaire 16-1-1 : La dimension de $\mathcal{M}_{mn}(\mathbf{K})$ est mn.

Démonstration : La famille (E_{ij}) est formée de mn matrices. •

On aura parfois besoin de ne regarder qu'un morceau d'une matrice :

Définition 16-1-134 : Soit A une matrice (m,n). On appelle **sous-matrice** de A toute matrice obtenue en éliminant certaines (ou aucune) lignes de A et certaines (ou aucune) colonne de A. En termes plus formels (et sans utilité), une matrice B de taille (m',n') sera dite sous-matrice de A s'il existe une application croissante φ_1 de $\{1,\ldots,m'\}$ vers $\{1,\ldots,m\}$ et une application croissante φ_2 de $\{1,\ldots,n'\}$ vers $\{1,\ldots,n\}$ telle que pour tous i,j avec $1 \leq i \leq m'$, $1 \leq j \leq n'$, $b_{ij} = a_{\varphi_1(i)\varphi_2(j)}$.

On notera enfin que je recule devant la frappe de l'explication du produit par blocs, mais que j'en aurai parlé en amphi...

Il reste à terminer cette section par une notion fort simple : l'échange des lignes et des colonnes, appelé transposition.

Définition 16-1-135 : Soit A une matrice (m,n). La **transposée** de A est la matrice (n,m) B définie pour tous i,j avec $1 \leq i \leq n$, $1 \leq j \leq m$ par : $b_{ij} = a_{ij}$.

Notation 16-1-57 : La transposée de A est notée ${}^tA$.

Proposition 16-1-80 : Pour toutes matrices A de taille (m,n), B de taille (n,p), ${}^t(AB) = {}^tB{}^tA$.

Démonstration : Notons $C = AB$, $D = {}^t(AB)$, $E = {}^tA$, $F = {}^tB$ et $G = {}^tB{}^tA$ Alors par définitions de la transposition et du produit, pour tout k avec $1 \leq k \leq p$ et tout i avec $1 \leq i \leq m$,

$$d_{ki} = c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk} = \sum_{j=1}^{n} e_{ji}f_{kj} = g_{ki}.$$

Donc $D = G$. •

Définition 16-1-136 : Une matrice est dite **symétrique** lorsqu'elle est égale à sa transposée.

En clair, une matrice symétrique est une matrice symétrique par rapport à sa diagonale nord-ouest/sud-est.

2 - Matrices et applications linéaires

Définition 16-2-137 : Soit E et F deux espaces vectoriels de dimension finie, $(e_1,\ldots,e_n)$ et $(f_1,\ldots,f_m)$ des bases respectives de E et F et u une application linéaire de E vers F. La **matrice de** u dans les bases $\underline{e}$ et $\underline{f}$ est la matrice (m,n) dont pour tout j avec $1 \leq j \leq n$ la j-ème colonne est la matrice-colonne des coordonnées de $u(e_j)$ dans la base $\underline{f}$.

Notation 16-2-58 : Cette matrice sera notée $\mathrm{mat}_{\underline{e},\underline{f}}(u)$. On osera éventuellement omettre les indices rappelant les bases, mais avec prudence et lorsqu'aucune confusion n'est possible, et surtout sans perdre de vue qu'une même application linéaire peut avoir des matrices d'aspects fort différents selon les bases que l'on choisit pour la représenter.

Lorsque les bases sont fixées, chaque matrice représente exactement une application linéaire. C'est exactement ce que dit en termes plus abstraits (je choisis même consciemment de donner une version peut-être trop abstraite, pour inviter à réfléchir...) la

Proposition 16-2-81 : Soit E et F deux espaces vectoriels de dimension finie, $(e_1, \ldots, e_n)$ et $(f_1, \ldots, f_m)$ des bases respectives de E et F, alors l'application $\mathrm{mat}_{\underline{e},\underline{f}}$ est un isomorphisme d'espaces vectoriels de $\mathcal{L}(E, F)$ vers $\mathcal{M}_{m,n}(\mathbf{K})$.

Démonstration : Lourde à écrire mais sans intérêt ni difficulté. •

On va enfin faire le lien avec la notion de "matrice d'un vecteur" introduite au premier semestre.

Proposition 16-2-82 : Soit E et F deux espaces vectoriels de dimension finie, $(e_1, \ldots, e_n)$ et $(f_1, \ldots, f_m)$ des bases respectives de E et F, x un vecteur de E et u une application linéaire de E vers F. Alors

$$\mathrm{mat}_{\underline{f}}[u(x)] = \mathrm{mat}_{\underline{e},\underline{f}}(u) \times \mathrm{mat}_{\underline{e}}(x).$$

Démonstration : C'est une simple vérification. Soit $x_1, \ldots, x_n$ les coordonnées de x dans $\underline{e}$, c'est à dire les coefficients de $\mathrm{mat}_{\underline{e}}(x)$. On a alors, en notant A la matrice de u dans les bases considérées :

$$u(x) = \sum_{i=1}^{n} x_i u(e_i) = \sum_{i=1}^{n} x_i (\sum_{j=1}^{m} a_{ji} f_j) = \sum_{j=1}^{m} (\sum_{i=1}^{n} a_{ji} x_i) f_j.$$

La j-ème coordonnée de $u(x)$ dans $\underline{f}$ est donc $\sum_{i=1}^{n} a_{ji} x_i$; en d'autres termes, le j-ème coefficient de $\mathrm{mat}_{\underline{f}}[u(x)]$ est bien le j-ème coefficient de $A \times \mathrm{mat}_{\underline{e}}(x)$. •

Il reste à apprendre à composer les applications linéaires. Pour ce faire, on démontre préalablement le

Lemme 16-2-6 : Soit A une matrice (m, p). On suppose que pour toute matrice-colonne $(p, 1)$ X, $AX = 0$. Alors $A = 0$.

Démonstration : Il suffit de remarquer que si on prend pour X la matrice-colonne $(p, 1)$ dont tous les coefficients sont nuls sauf le k-ème qui vaut 1, le produit AX est alors égal à la k-ème colonne de A. Dès lors toutes les colonnes de A sont nulles, donc A est nulle. •

Proposition 16-2-83 : Soit E, F et G trois espaces vectoriels de dimension finie, $(e_1, \ldots, e_p)$, $(f_1, \ldots, f_n)$ et $(g_1, \ldots, g_m)$ des bases respectives de E, F et G, u une application linéaire de E vers F et v une application linéaire de F vers G. Alors :

$$\mathrm{mat}_{\underline{e},\underline{g}}(v \circ u) = \mathrm{mat}_{\underline{f},\underline{g}}(v) \times \mathrm{mat}_{\underline{e},\underline{f}}(u)$$

Démonstration :

Notons A la matrice de u, B la matrice de v et C la matrice de $v \circ u$ dans les bases introduites. Soit x un vecteur de E, de matrice X dans $\underline{e}$. La matrice de $[v \circ u](x)$ dans la base $\underline{g}$ est CX. Par ailleurs, la matrice de $u(x)$ dans $\underline{f}$ est AX, puis celle de $v[u(x)]$ dans $\underline{g}$ est donc $B(AX) = (BA)X$.

Toutes les matrices colonnes $(p, 1)$ sont matrices d'un vecteur de E, donc on a prouvé que pour toute matrice colonne $(p, 1)$ X, on a $CX = (BA)X$ ou encore $(C - BA)X = 0$. On applique alors le lemme qui précède et on conclut que $C = BA$. •

3 - Matrices inversibles

Définir les matrices inversibles ferait double emploi avec la définition générale d'élément inversible dans un ensemble muni d'une opération avec un neutre. Précisons simplement que ce terme s'applique aux matrices carrées P et qu'il s'agit d'inverse multiplicatif (c'est-à-dire d'une autre matrice carrée Q telle que $PQ = QP = I_n$, n étant le côté de P.)

Il n'y a pas grand chose à ajouter aux généralités qu'on a déjà fait observer sur les éléments inversibles dans un contexte plus général et abstrait – on peut toujours rappeler que si P_1 et P_2 sont inversibles de même côté, $P_1 P_2$ l'est aussi, et que $(P_1 P_2)^{-1} = P_2^{-1} P_1^{-1}$.

Il y a toutefois un énoncé qui mérite d'être connu, car il est propre à ce contexte d'algèbre linéaire –nous le connaissons déjà pour des endomorphismes, et il se reporte tel quel pour des matrices carrées.

Proposition 16-3-84 : Soit P et Q deux matrices carrées (n,n). Si $PQ = I_n$, alors P et Q sont inversibles, et mutuellement inverses. En d'autres termes, un inverse à droite est automatiquement un inverse ; un inverse à gauche est automatiquement un inverse.

Démonstration : Introduisons un espace vectoriel E de dimension n et $(e_1,\ldots,e_n)$ une base de E (par exemple $E = \mathbf{K}^n$ et $\underline{e}$ sa base canonique, mais tout choix plus loufoque serait aussi satisfaisant). Soit u l'endomorphisme de E dont la matrice est P dans $\underline{e}$ et v celui dont la matrice est Q. On a alors $u \circ v = Id_E$. En appliquant la proposition 8-2-47, on conclut que u et v sont bijectifs et mutuellement réciproques, c'est-à-dire que $v \circ u = Id_E$ et donc $QP = I_n$.

•

4 - Changements de base

J'ai déjà eu une ou deux fois dans ce qui précède l'occasion de mettre en garde : un vecteur n'a pas dans l'absolu des coordonnées, il en a relativement à une base donnée ; de même une application linéaire n'a pas une matrice, elle en a tout plein : une par choix de bases sur les espaces de départ et d'arrivée.

Une question très raisonnable est donc de se demander comment reconstituer la matrice dans une base connaissant celle dans une autre.

Pour cela, le concept opératoire est celui de matrice de passage. Par exception aux bonnes habitudes à avoir en général, je conseille (sauf aux plus ambitieux) de ne **pas** connaître la définition de matrice de passage, que je crains féconde en confusions mentales et de se borner à la lecture de l'explication qui la précède.

Explication : Soit E un espace vectoriel de dimension finie, $(e_1,\ldots,e_n)$ et $(e'_1,\ldots,e'_n)$ deux bases de E. La **matrice de passage** f de $\underline{e}$ à $\underline{e'}$ est fabriquée de la façon suivante : dans sa première colonne, on dispose les coordonnées de e'_1 (le premier "nouveau" vecteur) dans $\underline{e}$ (l'"ancienne" base), dans sa deuxième colonne les coordonnées de e'_2, etc... C'est donc une matrice (n,n).

Cette "explication" est agréable pour écrire concrètement des matrices de passage, elle a le défaut de ne pas être agréable du tout pour démontrer des formules. Je la doublerai donc par une définition –que je conseille de ne pas connaître ! Les plus volontaires d'entre vous pourront toutefois se convaincre que l'explication et la définition recouvrent le même concept.

Définition 16-4-138 : Soit E un espace vectoriel de dimension finie, $(e_1,\ldots,e_n)$ et $(e'_1,\ldots,e'_n)$ deux bases de E. La **matrice de passage** f de $\underline{e}$ à $\underline{e'}$ est par définition la matrice $\mathrm{mat}_{\underline{e'},\underline{e}}\, Id_E$.

Avec cette définition obscure, les formules qui suivent –et qui sont à connaître– auront des démonstrations courtes (mais obscures).

Proposition 16-4-85 : Soit E un espace vectoriel de dimension finie, $(e_1,\ldots,e_n)$ et $(e'_1,\ldots,e'_n)$ deux bases de E. La matrice de passage de $\underline{e}$ à $\underline{e'}$ est inversible, et son inverse est la matrice de passage de $\underline{e'}$ à $\underline{e}$.

Démonstration :

$$\mathrm{mat}_{\underline{e'},\underline{e}}\, Id_E \times \mathrm{mat}_{\underline{e},\underline{e'}}\, Id_E = \mathrm{mat}_{\underline{e},\underline{e}}\, Id_E = I_n$$

et ce n'est pas la peine de vérifier la multiplication dans l'autre sens, s'agissant de matrices carrées.

•

Proposition 16-4-86 : Soit E un espace vectoriel de dimension finie, $(e_1,\ldots,e_n)$ et $(e'_1,\ldots,e'_n)$ deux bases de E. Notons P la matrice de passage de $\underline{e}$ à $\underline{e'}$. Soit x un vecteur de E, notons X sa matrice dans $\underline{e}$ et Y sa matrice dans $\underline{e'}$. Alors

$$X = PY.$$

Démonstration :

$$X = \mathrm{mat}_{\underline{e}}(x) = \mathrm{mat}_{\underline{e}}[Id(x)] = \mathrm{mat}_{\underline{e'},\underline{e}}(Id) \times \mathrm{mat}_{\underline{e'}}(x) = PY.$$

•

Remarque : Ce que dit cette proposition, c'est que quand on sait écrire une "nouvelle base" en fonction d'une "ancienne" il est facile de reconstituer les "anciennes" coordonnées d'un vecteur en fonction des "nouvelles", mais passer dans l'autre sens (ce qui est généralement ce qu'on a besoin de faire !) est plus lourd, puisque la formule à appliquer serait $Y = P^{-1}X$, qui nécessite de calculer P^{-1} –ou du moins de résoudre un système.

Proposition 16-4-87 : Soit E et F deux espaces vectoriels de dimension finie et u une application linéaire de E vers F. Soit $(e_1,\ldots,e_n)$ et $(e'_1,\ldots,e'_n)$ deux bases de E ; soit $(f_1,\ldots,f_m)$ et $(f'_1,\ldots,f'_m)$ deux bases de F. Notons P la matrice de passage de $\underline{e}$ à $\underline{e'}$ et Q la matrice de passage de $\underline{f}$ à $\underline{f'}$. Soit A la matrice de u dans $\underline{e}$ et $\underline{f}$ et B la matrice de u dans $\underline{e'}$ et $\underline{f'}$. Alors

$$B = Q^{-1}AP.$$

Démonstration :

$$B = \mathrm{mat}_{\underline{e'},\underline{f'}}(u) = \mathrm{mat}_{\underline{e'},\underline{f'}}(Id_F \circ u \circ Id_E) = \mathrm{mat}_{\underline{f},\underline{f'}}(Id_F) \times \mathrm{mat}_{\underline{e},\underline{f}}(u) \times \mathrm{mat}_{\underline{e'},\underline{e}}(Id_E) = Q^{-1}AP.$$

•

Remarque : Je ne soulignerai jamais trop à quel point ces formules sont pénibles à mémoriser, tant il est facile de confondre ce qui est "ancien" et ce qui est "nouveau" dans chacune. Un effort de concentration sera donc nécessaire ici.

5 - Matrices équivalentes et matrices semblables

Définition 16-5-139 : Soit A et B deux matrices (m,n). On dit que A et B sont **équivalentes** lorsqu'il existe deux matrices inversibles Q de côté m et P de côté n telles que $B = Q^{-1}AP$.
Remarques : * Il va de soi que j'aurais pu aussi bien donner la définition sous la forme "il existe deux matrices inversibles Q de côté m et P de côté n telles que $B = QAP$" (il suffit d'appeler Q ce qu'on avait appelé Q^{-1}). Je l'écris avec ce $^{-1}$ superflu pour une simple question de mémorisation, pour ne pas accumuler des formules qui se ressemblent un peu mais sont toutefois un peu différentes.

* Au vu de la proposition qui clôt la section précédente, on comprend ce que peut signifier cette notion d'équivalence : dire que deux matrices sont équivalentes, c'est dire qu'elles sont susceptibles de représenter la même application linéaire dans des bases différentes.

* Comme il est très facile de le vérifier, la relation "est équivalente à" est une relation d'équivalence sur $\mathcal{M}_{mn}(\mathbf{K})$.

Une autre relation d'équivalence entre matrices ne concerne que les matrices **carrées**. Il faut évidemment ne pas la confondre avec la précédente !
Définition 16-5-140 : Soit A et B deux matrices carrées (n,n). On dit que A et B sont **semblables** lorsqu'il existe une matrice inversible carrée P (n,n) telle que $B = P^{-1}AP$.
Remarques : * Il est aussi très facile de vérifier qu'il s'agit là d'une relation d'équivalence sur $\mathcal{M}_n(\mathbf{K})$. Elle est plus exigeante que la relation "est équivalente à" : je veux dire par là que si des matrices sont semblables, elles sont équivalentes, mais que la réciproque est fausse. La conséquence est que l'ensemble-quotient de $\mathcal{M}_n(\mathbf{K})$ par la similitude aura beaucoup plus d'éléments que celui par l'équivalence (concentrez-vous pour savoir ce que je veux dire exactement par là) ; d'ailleurs nous saurons dans quelques pages décrire le second alors que le premier est significativement plus compliqué (quoiqu'élucidable).

* Cette relation s'interprète aussi en termes de changement de matrices, mais cette fois pour un même endomorphisme. L'idée qu'il faut avoir pour cette interprétation est qu'il est stupide de considérer un endomorphisme en utilisant des bases différentes au départ et à l'arrivée : si je bouge mon repère en même temps que je bouge, il sera bien difficile de savoir de combien j'ai bougé. (Certes, concèderai-je, j'ai fait cette stupidité pour définir les matrices de passage, mais en vous invitant à ne pas me lire). Dès lors qu'on exige de ne considérer que des matrices de la forme $\mathrm{mat}_{\underline{e},\underline{e}}$ pour un endomorphisme u, on pourra de nouveau dire que des matrices sont semblables si et seulement si elles décrivent un même endomorphisme de E dans des bases différentes de E.

6 - Rang et équivalence

Le mot "rang" a plusieurs sens très voisins selon le contexte.
Définition 16-6-141 : Soit E un espace vectoriel et $(e_1,\ldots,e_k)$ un système de vecteurs de E. Le **rang** de $(e_1,\ldots,e_k)$ est la dimension du sous-espace vectoriel de E qu'ils engendrent.
Définition 16-6-142 : Soit A une matrice (m,n). Pour $1 \le j \le n$, notons C_j la j-ème colonne de A (qui est une matrice-colonne $(m,1)$). Le **rang** de A est le rang du système $(C_1,\ldots,C_n)$.

Définition 16-6-143 : Soit u une application linéaire. Lorsque $\operatorname{Im} u$ est de dimension finie, sa dimension est appelée le **rang** de u.

Les deux premiers concepts sont visiblement liés, le troisième l'est aussi, par la très élémentaire

Proposition 16-6-88 : Soit E et F deux espaces vectoriels de dimension finie et u une application linéaire de E vers F.Soit $(e_1, \dots, e_n)$ une base de E et $(f_1, \dots, f_m)$ une base de F ; soit A la matrice de u dans $\underline{e}$ et $\underline{f}$. Le rang de A est égal au rang de u.

Démonstration : Le rang de u est la dimension de $\operatorname{Im} u$, dont un système générateur évident est le système $(u(e_1), \dots, u(e_n))$. L'application qui à un vecteur associe sa matrice dans $\underline{f}$ est maintenant un isomorphisme d'espaces vectoriels entre F et $\mathcal{M}_{m1}(\mathbf{K})$, et cet isomorphisme envoie $u(e_j)$ sur le vecteur colonne C_j, j-ème colonne de la matrice A. Le sous-espace vectoriel de F engendré par $(u(e_1), \dots, u(e_n))$ a donc même dimension que le sous-espace vectoriel de $\mathcal{M}_{m1}(\mathbf{K})$ engendré par $(C_1, \dots, C_n)$, c'est-à-dire $\operatorname{rg} A$.

•

On peut s'étonner d'avoir défini une notion sur les colonnes des matrices sans avoir sur son élan donné une définition analogue sur les lignes. C'est que la dimension de l'espace engendré par les lignes d'une matrice est **lui aussi** égal au rang. Ce qui n'a rien d'évident et que nous pouvons considérer comme une motivation pour avancer plus loin dans l'étude du rang.

Le résultat suivant m'a paru valoir la peine d'être mis en relief, tant il est utilisable pour prouver un peu tout sur le rang, surtout si on se donne la peine de maîtriser également les matrices rencontrées dans sa démonstration.

Théorème 16-6-31 : Deux matrices (m, n) sont équivalentes si et seulement si elles ont même rang.

Démonstration : Soit A et B deux matrices (m, n).

Nous sommes mieux outillés pour travailler sur le rang des applications linéaires que sur celui des matrices. Nous allons donc construire des applications linéaires auxiliaires sur lesquelles portera l'essentiel de la démonstration. Pour cela, introduisons E un espace vectoriel de dimension n et $(e_1, \dots, e_n)$ une base de E, F un espace vectoriel de dimension m et $(f_1, \dots, f_m)$ une base de F (on peut prendre $\mathbf{K}^n$ et $\mathbf{K}^m$ avec leurs bases canoniques respectives, mais rien ne nous y oblige). Soit u l'application linéaire de E vers F dont la matrice dans les bases $\underline{e}$ et $\underline{f}$ est A, et v pour B.

* Preuve de $\Rightarrow$. Supposons donc A et B équivalentes, et soit Q et P des matrices inversibles telles que $B = Q^{-1}AP$.

Soit ψ l'endomorphisme de F dont la matrice dans la base $\underline{f}$ est Q^{-1}, et φ l'endomorphisme de E dont la matrice dans $\underline{e}$ est P. Puisque Q et P sont inversibles, ψ et φ sont bijectifs. La relation $B = Q^{-1}AP$ se réécrit $v = \psi \circ u \circ \varphi$, et nous devons prouver que $\operatorname{rg} u = \operatorname{rg} v$.

Les noyaux étant très légèrement plus faciles à manipuler que les images, remarquons tout de suite que puisque $\operatorname{rg} u = n - \dim \operatorname{Ker} u$ et $\operatorname{rg} v = n - \dim \operatorname{Ker} v$, il suffit de prouver l'égalité des dimensions des noyaux. Pour ce faire, on va prouver l'égalité ensembliste $\operatorname{Ker} v = \varphi^{-1}(\operatorname{Ker} u)$.

Soit $x \in E$. Alors $x \in \operatorname{Ker} v \iff v(x) = 0 \iff \psi(u[\varphi(x)]) = 0 \iff u[\varphi(x)] = 0$ (on utilise ici le fait que ψ est bijectif), donc $x \in \operatorname{Ker} v \iff \varphi(x) \in \operatorname{Ker} u \iff x \in \varphi^{-1}(\operatorname{Ker} u)$.

Comme φ est bijective, le sous-espace $\varphi^{-1}(\operatorname{Ker} u)$ est l'image du sous-espace $\operatorname{Ker} u$ par la bijection φ^{-1} et a donc la même dimension que lui. On a bien prouvé que $\dim \operatorname{Ker} u = \dim \operatorname{Ker} v$ et donc que $\operatorname{rg} A = \operatorname{rg} B$.

* Preuve de $\Leftarrow$. C'est le sens sérieux, et c'est pour celui-ci qu'une nouvelle idée va nous servir. Supposons donc que $\operatorname{rg} A = \operatorname{rg} B$ et notons r leur valeur commmune.

Introduisons la matrice J_r dont les coefficients a_{ij} sont définis par $a_{ii} = 1$ pour $1 \leq i \leq r$ et $a_{ij} = 0$ pour tous les autres coefficients. Pour prouver que A et B sont équivalentes, on va montrer qu'elles sont toutes les deux équivalentes à J_r.

Pour ce faire, on va construire de nouvelles bases de E et F bien adaptées à l'endomorphisme u de sorte que sa matrice soit la plus simple possible –ce sera la matrice J_r. On recommencera avec v.

Prenons une base $(e'_{r+1}, \dots, e'_n)$ de $\operatorname{Ker} u$ – la façon de la numéroter peut paraître bizarre, mais elle est cohérente puisque la dimension de $\operatorname{Ker} u$ est $n - r$. Par le théorème de la base incomplète appliqué à cette famille libre dans E, on peut prolonger ce système en une base $(e'_1, \dots, e'_n)$ de E. Posons alors pour $1 \leq j \leq r$, $f'_j = u(e'_j)$.

J'affirme que le système $(f'_1, \dots, f'_r)$ est libre. Soit en effet des scalaires $\lambda_1, \dots, \lambda_r$ tels que $\lambda_1 f'_1 + \cdots + \lambda_r f'_r = 0$. On a donc $0 = \lambda_1 u(e'_1) + \cdots + \lambda_r u(e'_r) = u(\lambda_1 e'_1 + \cdots + \lambda_r e'_r)$, donc $\lambda_1 e'_1 + \cdots + \lambda_r e'_r \in \operatorname{Ker} u = \mathbf{K}e'_{r+1} + \cdots + \mathbf{K}e'_n$. Le système $(e'_1, \dots, e'_n)$ étant une base de E, ceci n'est possible que si $\lambda_1 e'_1 + \cdots + \lambda_r e'_r = 0$ et donc si tous les λ_j sont nuls. Ceci prouve bien la liberté affirmée.

Une fois ce point prouvé, le théorème de la base incomplète appliqué à ce système permet de le prolonger en une base $(f'_1, \dots, f'_m)$ de F. Penchons nous sur la matrice de u dans les bases $\underline{e'}$ et $\underline{f'}$.

Cette matrice se construit colonne par colonne : la première colonne s'obtient en écrivant les coordonnées de $u(e'_1) = f'_1$ dans $\underline{f'}$, donc en écrivant un 1 puis tout plein de zéros. Pour la colonne suivante, on écrit les coordonnées de $u(e'_2) = f'_2$ dans $\underline{f'}$, c'est-à-dire un zéro, puis un 1, puis plein de zéros. Et ainsi de suite jusqu'à la r-ème colonne. Quand vient le tour de la $r+1$-ème, l'image par u de e'_{r+1} est nulle, donc on la remplit de zéros. Puis on continue à aligner zéro sur zéro jusqu'à plus soif.

C'est bien la matrice J_r que nous contemplons, le pensum terminé.

Ainsi la matrice J_r est la matrice de u dans d'autres bases que celles considérées initialement. Comme on l'a vu à la section précédente, elle est donc équivalente à la matrice A.

On recommence tout avec l'application linéaire v, prouvant que J_r est équivalente à B. A est donc équivalente à B.

•

Voyons maintenant tout ce qu'on peut désormais montrer à l'aide de ce théorème.

Théorème 16-6-32 : Soit A une matrice. Son rang est égal à celui de sa transposée. (En d'autres termes : le rang peut être calculé sur les lignes aussi bien que sur les colonnes).

Démonstration : Notons r le rang de A. Avec la notation de la preuve du théorème précédent, considérons la matrice J_r de même largeur et de même hauteur que A. Cette matrice J_r est de façon évidente de rang r ; elle est donc équivalente à A (en fait on a prouvé ce point au cours de la démonstration du théorème précédent). Soit donc Q et P inversibles telles que $A = Q^{-1}J_rP$.

Transposons le tout : ${}^tA = {}^tP^tJ_r{}^tQ^{-1}$. Mais la matrice J_r a le bon goût d'être symétrique (et le transposé de l'inverse le bon goût d'être l'inverse de la transposée, économisant des parenthèses). La formule se simplifie donc en ${}^tA = {}^tPJ_r{}^tQ^{-1}$. Ceci prouve que ${}^tA$ est équivalente à J_r et a donc elle aussi le rang r.

•

Proposition 16-6-89 : Soit $\mathbf{K}_1$ un corps contenant $\mathbf{K}$, et A une matrice à coefficients dans $\mathbf{K}$. Le rang de A est le même qu'on considère A comme une matrice à coefficients dans $\mathbf{K}$ ou comme une matrice à coefficients dans $\mathbf{K}_1$.

Démonstration : Soit r le rang de A, vue comme matrice à coefficients dans $\mathbf{K}$; avec les mêmes notations que précédemment, A est équivalente à J_r : il existe donc des matrices P et Q à coefficients dans $\mathbf{K}$, ayant des inverses à coefficients dans $\mathbf{K}$ telles que $A = Q^{-1}J_rP$. Toutes les matrices dans cette égalité peuvent être vues comme des matrices à coefficients dans $\mathbf{K}_1$, donc A est encore équivalente à J_r quand on la pense comme une matrice à coefficients dans $\mathbf{K}_1$. Son rang est donc toujours r.

•

Proposition 16-6-90 : Le rang d'une matrice A est égal au côté de la sous-matrice inversible de A de côté maximal.

Démonstration : Notons r le rang de A et c le maximum des côtés des sous-matrices inversibles de A. On va montrer que $r = c$ en prouvant la double inégalité entre ces entiers. On notera m et n les entiers tels que A soit une matrice (m, n).

* Montrons que $c \leq r$. Pour ce faire, considérons une sous-matrice carrée (c, c) inversible M de A, où on a conservé les lignes de A portant les numéros $i_1, \dots, i_c$ et les colonnes de A portant les numéros $j_1, \dots, j_c$. Les c colonnes de C forment donc un système libre dans $\mathcal{M}_{c1}(\mathbf{K})$. Chacune est un morceau de la colonne de même numéro de A ; il est facile de se convaincre que ces colonnes de A forment donc elles aussi un système libre, cette fois dans $\mathcal{M}_{m1}(\mathbf{K})$ (si je dis qu'"il est facile" plutôt que de le faire, c'est précisément que c'est un peu pénible ; c'est aussi parce que j'évite l'usage de la définition formelle de sous-matrice, et que je suis donc réduit à agiter les mains pour tenter de convaincre). Le sous-espace de $\mathcal{M}_{m1}(\mathbf{K})$ engendré par toutes les colonnes de A est donc de dimension supérieure ou égale à c : on a bien prouvé que $c \leq r$.

* Montrons que $r \leq c$. Pour ce faire, considérons le système $(C_1, \dots, C_n)$ formé des n colonnes de A. Ce système engendre un espace de dimension r ; on peut donc en extraire une base $(C_{i_1}, \dots, C_{i_r})$. Portons notre regard sur la matrice B formée de ces seules colonnes, qui est une sous-matrice (m, r) de A : cette matrice B a des colonnes linéairement indépendantes, donc est de rang r. Nous savons désormais que le rang de B peut être calculé sur ses lignes. Cela nous permet de faire subir aux lignes de B les mêmes outrages que subirent les colonnes de A. On en extrait donc r lignes $(L_{j_1}, \dots, L_{j_r})$ formant un système libre. La sous-matrice carrée de B formée de ces r lignes est alors encore de rang r, et elle est (r, r) : c'est donc une sous-matrice carrée inversible de A. Nous en déduisons que $r \leq c$.

•

Chapitre 17 - Complément sur les relations d'ordre

Ce chapitre, qui semble faire double emploi avec le cours d'informatique vise à donner un peu de vocabulaire indispensable pour comprendre vraiment l'analyse sur $\mathbf{R}$. Il s'agit donc essentiellement de définitions, presque rien n'a besoin de démonstrations dans ce court chapitre.

Définition 17-0-144 : Une relation d'ordre $\leq$ sur un ensemble E est dite **d'ordre total** lorsque pour tous éléments x et y de E, on a $x \leq y$ ou $y \leq x$.

L'archétype en est donc la relation d'ordre sur les ensembles de nombres familiers ; le contre-exemple typique la relation de divisibilité sur $\mathbf{N}^*$.

Dans toute la suite, les définitions ne seront données que "d'un côté" (majorée, pas minorée ; plus grand, pas plus petit). Tout lecteur muni d'un cerveau reconstituera les définitions manquantes.

Définition 17-0-145 : Soit une relation d'ordre $\leq$ sur un ensemble E, et A une partie de E. On dit que A est **majorée** par un élément M de E lorsque pour tout $x \in A$, $x \leq M$. Lorsqu'une partie de E possède au moins un majorant, on dit qu'elle est **majorée**. Lorsqu'une partie de E est majorée et minorée, on dit qu'elle est **bornée**.

Définition 17-0-146 : Soit une relation d'ordre $\leq$ sur un ensemble E, et A une partie de E. On dit qu'un élément M de A est **plus grand élément** de A lorsque M majore A.

Remarques : * Il est à peu près évident qu'on peut dire **le** plus grand élément : s'il y en a deux M et N, comme $M \in A$ et N majore A, $M \leq N$, et réciproquement $N \leq M$. On conclut par l'antisymétrie.

* Il est facile de prouver (et sera régulièrement utilisé) qu'un ensemble **fini** totalement ordonné (et non vide) possède forcément un plus grand élément.

Notation 17-0-59 : Le plus grand élément d'une partie A –s'il existe !– sera noté $\operatorname{Max} A$.

Définition 17-0-147 : Soit une relation d'ordre $\leq$ sur un ensemble E, et A une partie de E. On dit qu'un élément M de E est **borne supérieure** de A lorsque c'est le plus petit élément de l'ensemble des majorants de A.

Notation 17-0-60 : La borne supérieure d'une partie A –si elle existe !– sera notée $\operatorname{Sup} A$.

Définition 17-0-148 : Soit f une application d'un ensemble X vers un ensemble E muni d'une relation d'ordre $\leq$. On dira que f est **majorée** lorsque la partie $f(X) \subset E$ l'est. De même on définira f bornée, la notation $\operatorname{Max} f$, la notation $\operatorname{Sup} f$ (ou plus lourdement $\operatorname*{Sup}_{x \in X} f(x)$).

Reste à vérifier qu'on a bien compris toutes ces notions. Les implications suivantes sont vraies et même évidentes. Vous le paraissent-elles bien ?

Pour A partie d'un ensemble ordonné :

$$\operatorname{Max} A \text{ existe} \quad \Rightarrow \quad \operatorname{Sup} A \text{ existe} \quad \Rightarrow \quad A \text{ est majorée.}$$

Dans l'autre sens, les implications seraient fausses. Un exemple d'ensemble possédant un Sup mais pas de Max devrait sauter aux yeux : l'intervalle $[0,1[$ de $\mathbf{R}$ a cette propriété. Trouver un exemple d'ensemble majoré mais n'ayant pas de Sup n'est pas du tout évident en revanche. Un exemple simple quoique camouflé est $\emptyset$ comme partie de $\mathbf{R}$, mais sa considération n'aide en rien à comprendre les notions. Comme on le verra au chapitre suivant, la recherche d'un tel exemple dans $\mathbf{R}$ est vouée à l'échec. L'exemple que je soumettrai donc à votre méditation doit être pêché dans un ensemble moins familier, l'ensemble $\mathbf{Q}$ des fractions. Si dans $\mathbf{Q}$ on considère l'ensemble $A = \{\frac{p}{q} \in \mathbf{Q} \mid \left(\frac{p}{q}\right)^2 < 2\}$, on peut montrer (c'est un exercice abordable mais difficile) que A n'admet pas de borne supérieure. Pourtant il est évidemment majoré, par $3/2$ par exemple. L'idée autant qu'on puisse l'expliquer informellement est que c'est $\sqrt{2}$ qui aimerait être la borne supérieure de A, mais qu'il fait malheureusement défaut à $\mathbf{Q}$ laissant A orphelin.

Pour finir de remplir la page, j'inviterai à observer les théorèmes d'existence du PGCD et du PPCM dans $\mathbf{N}^*$. Êtes-vous convaincus que ces théorèmes ne sont que des théorèmes d'existence du Sup et de l'Inf pour les parties à deux éléments (puis par une récurrence facile pour les parties finies) de $\mathbf{N}^*$ et la relation de divisibilité ?

Chapitre 18 - Nombres réels

Sans tenter de définir les nombres réels –ce serait faisable et ennuyeux– ni perdre du temps à démontrer tout ce qui est totalement évident les concernant, une mise au point sur leur secret (la propriété de la borne supérieure) et ses conséquences. Comme je l'avais fait pour les entiers, je ne définis pas les réels sans même m'offrir le luxe d'énoncer une "non-définition" les concernant.

1 - Les propriétés admises

Fait : l'ensemble des nombres réels est un corps ; il est muni d'une relation d'ordre total notée $\leq$ qui vérifie les propriétés suivantes :

* Pour tous a, b, c réels, si $a \leq b$, $a + c \leq b + c$.
* Pour tous a, b, c réels, si $a \leq b$ et $0 \leq c$, $ac \leq bc$.
* Toute partie non vide majorée de $\mathbf{R}$ admet une borne supérieure.

Nous ne pouvons démontrer ces propriétés –puisque je vous ai caché ce que sont les réels– mais il n'y aura plus besoin d'en ajouter d'autres discrètement : à partir de ces faits, nous pourrons tout prouver.

2 - Les propriétés les plus idiotes des réels

Ce sont en gros les propriétés pour la preuve desquelles je n'ai pas besoin de la dernière propriété, la plus ésotérique. Comme l'ensemble $\mathbf{Q}$ des nombres rationnels (les fractions) vérifie aussi les autres propriétés des réels, ce que nous verrons ici serait aussi vrai dans $\mathbf{Q}$ et est d'une façon générale très facile à prouver. Si facile que je ne chercherai pas à être exhaustif –des tas de livres le sont– et je considèrerai comme bien connues les propriétés élémentaires dont j'aurai besoin de ci de là.

Faisons en une toutefois juste pour montrer que nous en sommes capable.

Proposition 18-2-91 : Le carré de tout réel est positif (au sens large).

Démonstration : Soit x un réel. Supposons dans un premier temps x positif. On a alors $0 \leq x$ et $0 \leq x$. Appliquons la propriété de compatibilité de l'ordre et la multiplication à 0, x et x : on obtient $0 \leq x^2$. Examinons maintenant le cas où x ne serait pas positif. L'ordre $\leq$ étant total, x est donc négatif, soit $x \leq 0$. Ajoutons $-x$ aux deux côtés de cette inégalité. On en conclut que $-x$ est positif. On peut alors appliquer le premier temps à $-x$ et conclure que $x^2 = (-x)^2$ est positif.

•

3 - La fonction valeur absolue

C'est encore des choses bien connues, et qui seraient vraies dans $\mathbf{Q}$. Vu leur utilité technique dans tous les calculs d'analyse, ce ne peut faire de mal de les mettre en relief.

Définition 18-3-149 : La **valeur absolue** d'un réel x est par définition égale à x si x est positif, à $-x$ sinon.

Lemme 18-3-7 : Pour tous réels x et M,

$$-M \leq x \leq M \iff |x| \leq M.$$

Démonstration : Traitons dans un premier temps le cas où x est positif. Pour de tels x, $\Rightarrow$ est évident (c'est effacer l'inégalité de gauche), et $\Leftarrow$ ne l'est guère moins (puisque M est plus grand que $|x|$, il est positif, donc $-M$ est négatif, donc plus petit que x).

Si x est négatif, remarquons que $-M \leq x \leq M \iff -M \leq -x \leq M$ (multiplication par -1) et que $|x| \leq M \iff |-x| \leq M$ (puisque $|x| = |-x|$) ; le lemme s'en déduit en appliquant le premier temps à $-x$. •

Remarque : Bien qu'il soit tout à fait évident, j'ai choisi de mettre en relief ce lemme car il est à la source d'une idée simple et très fréquemment utile : pour montrer, par exemple, qu'une fonction f à valeurs réelles est bornée, il est souvent plus confortable de se contenter de majorer $|f|$ par un réel M. On en déduit aussitôt que f est elle-même minorée par $-M$ et majorée par M, donc bornée.

Proposition 18-3-92 : (inégalité triangulaire) Pour tous réels a, b, c,

$$|a+b| \leq |a| + |b|$$

$$|a| - |b| \leq |a+b|.$$

Démonstration : Pour la première inégalité, appliquons le lemme à $x = a$ et $M = |a|$ puis à $x = b$ et $M = |b|$.

On obtient :
$$\begin{array}{ccccc} -|a| & \leq & a & \leq & |a| \\ -|b| & \leq & b & \leq & |b| \end{array}$$

On additionne : $$-(|a|+|b|) \leq a+b \leq |a|+|b|$$

et on réapplique le lemme dans l'autre sens, cette fois à $x = a+b$ et $M = |a|+|b|$.
Pour la seconde, on applique la première à $a+b$ et $-b$.

•

4 - La fonction partie entière

Plus subtil ici ! Tous les énoncés sont encore vrais dans **Q** mais les preuves qui en sont données ici sont propres à **R**.

Lemme 18-4-8 : Dans **R** l'ensemble **N** des entiers positifs n'est pas majoré.

Démonstration : Cela peut paraître évident, mais la preuve demande de bien s'y prendre. On va démontrer le résultat par l'absurde : supposons **N** majoré. Comme il n'est pas vide, il possèderait une borne supérieure M, en utilisant –pour la première fois– la propriété de la borne supérieure. On aurait donc pour tout entier n positif, $n \leq M$, et en particulier pour tout entier m positif, $m+1 \leq M$, donc $m \leq M-1$. Ainsi $M-1$ serait un majorant de **N**. Mais M, en tant que borne supérieure, est le plus petit majorant de **N**. On en déduit donc que $M \leq M-1$ donc que $0 \leq -1$ et c'est absurde.

•

Ce lemme nous permet de conclure à l'existence de la partie entière :

Proposition 18-4-93 : Pour tout réel x, il existe un unique entier (relatif) $n \in \mathbf{Z}$ tel que $n \leq x < n+1$.

Remarque : La preuve va ressembler de façon étonnante à celle de l'existence et l'unicité de la division euclidienne d'un entier par un autre (ce qu'on fait, c'est une sorte de division euclidienne d'un réel par l'entier 1). Le parallèle sera peut-être encore plus frappant si on énonce la proposition de la façon plus lourde suivante (en faisant apparaître la partie entière mais aussi la partie fractionnaire de x) : “il existe un unique entier (relatif) n et un unique réel s vérifiant la double condition : $x = 1n + s$ et $0 \leq s < 1$.”

Démonstration : On prouvera successivement l'existence et l'unicité de n.

* Existence de n : la démonstration se prête bien à discuter selon le signe de x. Le cas où $x \geq 0$ est le cas contenant l'essentiel de la démonstration ; lorsque $x < 0$, on ne peut utiliser mot à mot la même preuve, mais on se ramène alors sans mal au cas intéressant déjà traité.

- Premier cas (le cas significatif) : si $x \geq 0$.
  L'idée de la preuve est de prendre pour n le plus grand entier N tel que N soit plus petit que x.
  Introduisons donc l'ensemble $A = \{c \in \mathbf{N} \mid c \leq x\}$. L'ensemble A est un ensemble d'entiers naturels ; il est non vide car il contient 0. Il est fini : en effet comme **N** n'est pas majoré, x n'est pas un majorant de **N**, donc il existe un entier d tel que $x < d$. L'ensemble A ne contient donc que des entiers inférieurs ou égaux à $d-1$ et est donc fini.
  L'ensemble A possède donc un plus grand élément n, qui vérifie évidemment $n \leq x$. Comme $n+1 \notin A$, on en déduit que $x < n+1$.
  L'existence est prouvée dans ce cas.
- Second cas (preuve sans imagination) : si $x < 0$.
  Si x est entier, on prend $n = x$. Sinon, on applique le premier cas à $-x$ et on obtient un m tel que $m < -x < m+1$. On pose alors $n = -1 - m$.

* Unicité de n : soit n_1 et n_2 deux entiers vérifiant la condition exigée dans l'énoncé de la proposition.
Comme $n_1 \leq x$ et $-n_2 - 1 < -x$, on obtient $n_1 - n_2 < 1$. De même $n_2 - n_1 < 1$, ou si on préfère $-1 < n_1 - n_2$.
Comme $n_1 - n_2$ est entier, il ne peut être que nul, donc $n_1 = n_2$.

•

Définition 18-4-150 : L'unique entier n tel que $n \leq x < n+1$ est appelé la **partie entière** du réel x.

Notation 18-4-61 : La partie entière de x est notée $\mathrm{E}(x)$ ou $\lfloor x \rfloor$.

5 - Intervalles

On peut caractériser les intervalles de **R** d'(au moins) deux façons : par une désagréable énumération de tous les cas possibles, ou par une propriété plus concise, plus économique pour prouver qu'une partie est un intervalle, moins opératoire pour travailler directement sur cette partie. Je donne pour définition cette seconde caractérisation, qui a l'agrément de la concision.

Définition 18-5-151 : Soit I un sous-ensemble de $\mathbf{R}$. On dira que I est un **intervalle** lorsque pour tous x, y, z réels tels que $x \leq y \leq z$, si x et z sont dans I, y aussi est dans I.

Ceux qui connaissent le mot constatent qu'on a défini les intervalles comme convexes de **R**.

Cette définition a l'avantage de se prêter à des démonstrations concises, elle a le défaut d'être peu explicite.

Définissons donc précisément les divers types d'intervalle :

Notation 18-5-62 : La notation $]-\infty, +\infty[$ désigne l'ensemble $\mathbf{R}$. Pour a réel, la notation $[a, +\infty[$ désigne $\{t \in \mathbf{R} \mid a \leq t\}$ et la notation $]a, +\infty[$ désigne $\{t \in \mathbf{R} \mid a < t\}$ (de même avec une borne à droite). Pour a, b réels avec $a \leq b$, la notation $[a, b]$ désigne $\{t \in \mathbf{R} \mid a \leq t \leq b\}$, la notation $[a, b[$ désigne $\{t \in \mathbf{R} \mid a \leq t < b\}$ (de même avec les crochets dans l'autre sens) et la notation $]a, b[$ désigne $\{t \in \mathbf{R} \mid a < t < b\}$. (Par convention, dans ce cours, ce genre de notation n'a aucun sens lorsque $b < a$, cette convention n'ayant rien d'universel...)

Proposition 18-5-94 : Soit I une partie de $\mathbf{R}$. I est un intervalle si et seulement si I est d'un des neuf types suivants :

$$\begin{array}{lll}]-\infty, +\infty[&]a, +\infty[& [a, +\infty[\\ ]-\infty, b[&]a, b[& [a, b[\\ ]-\infty, b] &]a, b] & [a, b]. \end{array}$$

Démonstration : On va renoncer à l'écrire complètement, car elle se subdivise fatalement en un nombre de cas abondant qui la rend ennuyeuse...

$\Leftarrow$ ne contient pas l'ombre d'une astuce, et demande seulement de la patience, vu le nombre de cas...

Pour $\Rightarrow$, partons donc d'un intervalle I. Le cas particulier où I est vide se traite à part (dans ce cas, $I =]0, 0[$) ; on supposera désormais I non vide.

On est amené à diviser la situation en trois cas, eux-mêmes subdivisés en trois sous-cas.

1) Si I n'est pas minoré.
2) Si I est minoré, et admet un plus petit élément.
3) Si I est minoré, et n'admet pas de plus petit élément.

Les trois sous-cas correspondant à la même division, mais du côté droit de I : majoré ou non, et s'il est majoré disposant ou non d'une borne supérieure.

On va traiter ici un seul cas, les huit autres étant "laissés au lecteur".

Supposons donc I non minoré, majoré, mais ne possédant pas de plus grand élément. Comme I est majoré et non vide, il possède une borne supérieure b.

On va montrer que $I =]-\infty, b[$; pour ce faire, la double inclusion s'impose.

- Preuve de $\subset$: il n'y a aucune difficulté. Si t est dans I, comme b est un majorant de I, $t \leq b$; comme I ne possède pas de plus grand élément, b n'est pas un élément de I et donc $t < b$.
- Preuve de $\supset$: soit t un élément de $]-\infty, b[$. Comme $t < b$ et que b est le plus petit majorant de I, t n'est pas un majorant de I, donc il existe un $\beta \in I$ avec $t < \beta$. Comme I n'est pas minoré, t n'est pas un minorant de I, donc il existe un $\alpha \in I$ avec $\alpha < t$. Appliquons la définition de "intervalle" aux trois réels $\alpha \leq t \leq \beta$, où les deux extrêmes α et β sont dans I. On en déduit que $t \in I$. L'inclusion est donc prouvée.

Ceci prouve que $I =]-\infty, b[$.

•

Chapitre 19 - Suites de réels

Au risque d'être un peu trop concis sur les définitions qui ouvrent le chapitre, je récupèrerai ce qu'on sait sur les limites d'une façon générale et qui a fait l'objet d'un chapitre au premier semestre : à l'aide de ces références, il n'y a plus rien de sérieux à démontrer dans les premiers paragraphes, les généralités sur les limites de suites.

À partir de la deuxième section, on va faire usage de la relation d'ordre sur $\mathbf{R}$ et de sa propriété non évidente fondamentale (celle de la borne supérieure) pour démontrer des résultats d'énoncé assez simple, tout à fait fondamentaux, et nouveaux par rapport au premier semestre.

1 - Limites de suites

Pour pouvoir faire le lien avec le chapitre relatif aux limites en général, il faut commencer par mettre côte à côte le résultat visuellement évident du chapitre précédent "$\mathbf{N}$ n'est pas majoré" et la définition (technique et oubliable) de "$+\infty$ est adhérent à un ensemble A". En relisant les deux, on constate qu'on sait désormais que $+\infty$ est adhérent à $\mathbf{N}$.

Dès lors, les suites de réels, fonctions à valeurs réelles définies sur la partie $\mathbf{N} \subset \mathbf{R}$, sont des cas particuliers de fonctions réelles d'une variable réelle sur lesquelles toutes les définitions et propriétés plus ou moins élémentaires ont été visitées au premier semestre. Je pourrais donc clôturer ici cette section. Cela ne peut toutefois faire de mal de mettre en relief la

Remarque : Si on recopie à la lettre la définition de "u_n tend vers l quand $n \to +\infty$ donnée au premier semestre, on obtient la formulation :

$$(1) \qquad \forall \epsilon > 0, \exists A \in \mathbf{R}, \forall n \in \mathbf{N}, ((n \geq A) \Rightarrow (|u_n - l| \leq \epsilon))$$

Ces A réels quelconques sont bien peu judicieux lorsque l'ensemble de définition de la suite ne contient que des entiers, et on retiendra plutôt la formulation très voisine mais tout de même formellement distincte :

$$(2) \qquad \forall \epsilon > 0, \exists N \in \mathbf{N}, \forall n \in \mathbf{N}, ((n \geq N) \Rightarrow (|u_n - l| \leq \epsilon))$$

Il faut se convaincre, si ce n'est déjà fait, que les deux définitions (1) et (2) illustrent le même concept. Si (2) est réalisée, (1) l'est de façon évidente, l'entier N étant en particulier un réel ; réciproquement, si (1) est réalisée, comme $\mathbf{N}$ n'est pas majoré, il existe un entier N vérifiant $A < N$: si on prend cet entier N, (2) est alors vérifiée.

On modifiera de la même façon la définition de "$u_n \to +\infty$ quand $n \to +\infty$" :

$$(2) \qquad \forall B \in \mathbf{R}, \exists N \in \mathbf{N}, \forall n \in \mathbf{N}, ((n \geq N) \Rightarrow (B \leq u_n)).$$

On peut donc réutiliser sur les suites tout ce que l'on sait pour les fonctions d'une variable réelle : possibilité de faire des opérations, principe des gendarmes, et (plus méconnu et il est vrai plus technique) méthode de recollement de restrictions ; la proposition 5-4-29 (ou plus exactement sa variante que je n'ai pas écrite pour des limites en l'infini) permet d'affirmer par exemple que si (u_{2n}) et (u_{2n+1}) tendent toutes deux vers l, (u_n) tend elle-même vers l.

Les deux mots ci-dessous sont certainement déjà connus de vous, mais je ne les ai encore définis nulle part.

Définition 19-1-152 : Une suite de réels est dite **convergente** lorsqu'elle admet une limite (finie) et **divergente** dans le cas contraire.

On ne perdra pas de vue que les suites "divergentes" ne sont pas seulement celles qui tendent vers $+\infty$ ou $-\infty$: la plupart d'entre elles courent dans tous les sens dans le désordre le plus total.

2 - Suites et monotonie

Le résultat suivant est des plus simples, et facile à montrer –à partir de la propriété de la borne supérieure de $\mathbf{R}$. On le généralisera aux fonctions au chapitre suivant.

Théorème 19-2-33 : Soit (u_n) une suite **croissante** de réels.

* Si (u_n) n'est pas majorée, $u_n \to \infty$ quand $n \to \infty$.

* Si (u_n) est majorée, elle est convergente ; sa limite est égale à $\mathrm{Sup}\, u_n$.

Démonstration :

* Cas où (u_n) n'est pas majorée. Soit alors un $B \in \mathbf{R}$. Comme B n'est pas un majorant de (u_n) il existe un $N \in \mathbf{N}$ tel que $B < u_N$. Comme (u_n) est croissante, pour tout $n \geq N$, $u_N \leq u_n$ donc $B \leq u_n$. C'est exactement la définition de “tendre vers l'infini”.

* Cas où (u_n) est majorée. L'ensemble des valeurs prises par cette suite est alors un ensemble de réels non vide et majoré. Il admet donc une borne supérieure, qu'on notera l. Comme l est un majorant de (u_n), on a pour tout entier $n \in \mathbf{N}$ l'inégalité $u_n \leq l$. Soit maintenant un $\epsilon > 0$. Le réel $l - \epsilon$ est alors strictement inférieur au réel l, qui est le plus petit majorant de (u_n). Le réel $l - \epsilon$ n'est donc pas un majorant de (u_n), donc il existe un $N \geq 0$ tel que $l - \epsilon < u_N$. En utilisant comme dans la première partie la croissance de (u_n), on en déduit que pour tout $n \geq N$, $l - \epsilon < u_N \leq u_n$, et donc en synthétisant les deux inégalités prouvées que $l - \epsilon < u_n \leq l$; les réels l et u_n sont donc à moins de ϵ l'un de l'autre, ou en d'autres termes $|u_n - l| < \epsilon$. La convergence vers l est prouvée.

•

On a évidemment un résultat analogue pour les suites décroissantes, minorées ou non.

Le critère qui suit, conséquence facile du résultat qui précède, couvre un cas particulier assez anecdotique vu de haut, mais étonnamment utile dans les vraies situations pratiques, celles qu'on rencontre en TD ou dans les sujets d'examen.

Proposition 19-2-95 : (“critère des suites adjacentes”) Soit (u_n) et (v_n) deux suites de réels. On suppose :

* que (u_n) croît tandis que (v_n) décroît ;

* que $v_n - u_n \to 0$ quand $n \to \infty$.

Alors les suites (u_n) et (v_n) sont convergentes, vers la même limite.

Démonstration : Elle est très facile : la suite (u_n) est croissante, on va utiliser le théorème qui précède et il nous reste donc seulement à prouver qu'elle est majorée.

Comme (v_n) décroît ainsi que $(-u_n)$, la suite $(v_n - u_n)$ décroît. Soit un entier N fixé. On a donc pour tout entier $n \geq N$ l'inégalité $v_n - u_n \leq v_N - u_N$. Faisons tendre n vers l'infini dans cette inégalité : on obtient $0 \leq v_N - u_N$, soit $u_N \leq v_N$.

Pour tout entier $N \geq 0$, on a donc, en utilisant une nouvelle fois la décroissance de (v_n) : $u_N \leq v_N \leq v_0$, et la suite (u_n) est donc majorée, par le réel **fixe** v_0.

Croissante et majorée, la suite (u_n) admet donc une limite l. En écrivant que $v_n = (v_n - u_n) + u_n$, on voit aussitôt que pour sa part (v_n) converge vers $0 + l$ donc aussi vers l.

•

3 - Sous-suites

Encore un concept très simple, mais obscurci par un formalisme inévitable pour pouvoir en faire usage proprement...

Définition 19-3-153 : Soit (u_n) une suite (pas forcément de réels d'ailleurs) ; on dit qu'une suite (v_n) est une **sous-suite** (ou **suite extraite**) de (u_n) lorsqu'il existe une application φ de $\mathbf{N}$ vers $\mathbf{N}$, strictement croissante, telle que pour tout n entier, $v_n = u_{\varphi(n)}$.

En clair, une suite extraite de (u_n) est une suite où on a gardé certains des termes de la première suite et jetés les autres, mais en conservant bien l'ordre (la condition “φ croissante” l'exprime) et sans bégayer (c'est la condition de croissance “stricte”).

Pour toutes les démonstrations concernant des sous-suites, il sera confortable de connaître le très facile

Lemme 19-3-9 : Soit φ une application strictement croissante $\mathbf{N} \to \mathbf{N}$. Alors pour tout entier $n \in \mathbf{N}$, $\varphi(n) \geq n$. En particulier, $\varphi(n) \to +\infty$ quand $n \to \infty$.

Démonstration : C'est sans surprise une récurrence sur n : pour $n = 0$ il n'y a rien à montrer, l'entier $\varphi(0)$ étant évidemment supérieur ou égal à 0. Supposons le résultat prouvé pour un n fixé. Comme φ est strictement croissante, on a alors $\varphi(n+1) > \varphi(n)$ et comme ce sont tous les deux des entiers, $\varphi(n+1) \geq \varphi(n) + 1$. Comme par hypothèse de récurrence, $\varphi(n) \geq n$ on en déduit aussitôt que $\varphi(n+1) \geq n + 1$.

•

Une fois ce lemme mis en relier, il ne reste plus rien à prouver pour la

Proposition 19-3-96 : Toute sous-suite extraite d'une suite de réels convergente converge, vers la même limite. De même pour les suites extraites de suites tendant vers $+\infty$ ou $-\infty$.
Démonstration : Soit (u_n) une suite de réels convergente, vers un réel l, et $(u_{\varphi(n)})$ une suite extraite. Vu le lemme, $\varphi(n) \to \infty$ quand $n \to \infty$, donc par composition des limites (par la version concernant des limites éventuellement infinies, celle que j'ai eu la flemme d'écrire), $u_{\varphi(n)} \to l$ quand $n \to \infty$. On traiterait de même les limites infinies.
•

Exemple : Cette proposition avait une preuve presque vide, pourtant elle nous permet de prouver que la suite $((-1)^n)$ diverge (sans même tendre vers $+\infty$ ou $-\infty$) –ce qu'on pouvait faire plus haut à la main, mais plus lourdement. Supposons en effet que $(-1)^n \to l$ quand $n \to \infty$. La sous-suite $\left((-1)^{2n}\right)$ tendrait aussi vers l, d'où $l = 1$ et la sous-suite $\left((-1)^{2n+1}\right)$ itou, donc $l = -1$. Contradiction !

Le théorème suivant est fort élégant mais ne nous servira guère que d'étape sur la voie du théorème de Bolzano-Weierstrass.

Théorème 19-3-34 : (Ramsey) De toute suite de réels, on peut extraire une sous-suite monotone.
Démonstration : Soit (u_n) une suite de réels. On va noter $A = \{n \in \mathbf{N} \mid \text{pour tout } k > n, u_k < u_n\}$.

* Si A est fini. On va alors parvenir à extraire de (u_n) une sous-suite croissante.

Prenons $\varphi(0)$ strictement plus grand que tous les éléments de A. Dès lors, $\varphi(0) \notin A$ et il existe donc au moins un $n > \varphi(0)$ tel que $u_{\varphi(0)} \leq u_n$. Prenons pour $\varphi(1)$ un tel n : on a alors $u_{\varphi(0)} \leq u_{\varphi(1)}$ d'une part, et d'autre part $\varphi(0) < \varphi(1)$ dont on déduit que $\varphi(1) \notin A$. Ce dernier fait autorise à recommencer de même et construire un $\varphi(2)$ tel que $u_{\varphi(1)} \leq u_{\varphi(2)}$ et en même temps $\varphi(1) < \varphi(2)$, et donc $\varphi(2) \notin A$. On peut alors écrire si on y tient une récurrence formelle permettant de construire toute l'application φ, ou se contenter comme je le suggère d'être convaincu qu'on saurait le faire. La suite extraite croissante annoncée est alors construite.

* Si A est infini. On va alors parvenir à extraire de (u_n) une suite décroissante.

Prenons pour $\varphi(0), \varphi(1), \ldots$ les éléments de A, numérotés dans l'ordre croissant. Alors l'application φ est définie sur $\mathbf{N}$ –puisque A est infini–, strictement croissante par construction.

Pour tout $n \in \mathbf{N}$, comme $\varphi(n)$ est dans A et que $\varphi(n+1) > \varphi(n)$ on obtient par définition de A l'inégalité $u_{\varphi(n+1)} < u_{\varphi(n)}$. La suite extraite construite est donc bien décroissante (et même strictement décroissante).
•

On va en déduire un résultat étonnant, qui arrive à tirer une conclusion à partir de presque aucune hypothèse, le

Théorème 19-3-35 : (Bolzano-Weierstrass) De toute suite bornée de réels, on peut extraire une sous-suite convergente.
Démonstration : Par le théorème précédent, on peut extraire une sous-suite monotone ; si elle est croissante, elle est croissante majorée, donc convergente ; si elle est décroissante, elle est décroissante minorée, donc convergente.
•

4 - Le critère de Cauchy

C'est un critère, assez technique, permettant de prouver qu'une suite converge lorsqu'on n'a pourtant aucune idée *a priori* de sa limite. En fait, pour des suites de réels, on dispose déjà du critère relatif suites monotones qui permet un tel exploit, est beaucoup plus manipulable par un étudiant débutant, et, quoique ne marchant pas "à tous les coups" donne un résultat dans les problèmes pas trop artificiels, où la suite n'est pas totalement désordonnée.

Toutefois, pour des suites de réels trop agitées, ou surtout dans les années ultérieures pour des suites de fonctions, l'usage de la monotonie se révèle souvent insuffisant. Le critère de Cauchy sera alors précieux. Cette année, l'objectif est surtout de l'apprendre plus que de savoir s'en servir dans des cas difficiles –mais pour être réduit, l'objectif ne doit pas moins être tenu !

Définition 19-4-154 : Une suite (u_n) de réels est dite **de Cauchy** lorsqu'elle vérifie la propriété suivante :

$$\text{Pour tout } \epsilon > 0, \text{ il existe un } N \geq 0, \text{ tel que pour tous } p \geq q \geq N, |u_p - u_q| \leq \epsilon.$$

Nous aurons besoin du

Lemme 19-4-10 : Toute suite de Cauchy est bornée.

Démonstration : Soit (u_n) une suite de Cauchy de réels. Appliquons la définition de "suite de Cauchy" à $\epsilon = 1$: on obtient un entier N tel que pour tous $p \geq q \geq N, |u_p - u_q| \leq 1$. En particulier, en prenant $q = N$, pour tout $p \geq N$, on a : $|u_p - u_N| \leq 1$, qu'on peut préférer écrire : $u_N - 1 \leq u_p \leq u_N + 1$.

On en déduit que pour tout $n \geq 0$:

$$\mathrm{Min}(u_0, u_1, \ldots, u_{N-1}, u_N - 1) \leq u_n \leq \mathrm{Max}(u_0, u_1, \ldots, u_{N-1}, u_N + 1).$$

•

Théorème 19-4-36 : Une suite de réels est convergente si et seulement si elle est de Cauchy.
Démonstration : Soit (u_n) une suite de réels.
• Preuve de $\Rightarrow$. Supposons (u_n) convergente, et notons l sa limite. C'est le sens facile : on sait que les u_n se dirigent tous vers l, il faut prouver que leur comportement est grégaire.

Soit un $\epsilon > 0$ fixé. Appliquons la définition de "converger" au réel strictement positif $\frac{\epsilon}{2}$. On obtient un entier N tel que pour tout entier $n \geq N$, on ait l'inégalité : $|u_n - l| \leq \frac{\epsilon}{2}$.

Soit $p \geq q \geq N$ deux entiers ; on a alors :

$$|u_p - u_q| = |(u_p - l) - (u_q - l)| \leq |u_p - l| + |u_q - l| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

• Preuve de $\Leftarrow$. Supposons (u_n) de Cauchy. Ce sens peut paraître plus surprenant : on sait que les (u_n) tendent à se regrouper, mais comment connaître le lieu l vers lequel ils se dirigent ?

Le théorème de Bolzano-Weierstrass est la bonne piste, puisqu'il permet de produire quelque chose à partir de (presque) rien : d'après le lemme, la suite de Cauchy (u_n) est bornée. On peut donc en extraire une sous-suite $(u_{\varphi(n)})$ convergente, vers une limite l. On va montrer l est en fait la limite de toute la suite (u_n).

Pour ce faire, appliquons tout d'abord la définition de "suite de Cauchy" au réel $\frac{\epsilon}{2}$: elle produit un entier N_1 tel que pour tous $p \geq q \geq N_1$ on ait : $|u_p - u_q| \leq \frac{\epsilon}{2}$.

Appliquons ensuite la définition de "converger" à la sous-suite convergente $(u_{\varphi(n)})$ et encore au réel $\frac{\epsilon}{2}$: ceci fournit un entier N_2 tel que pour tout $n \geq N_2$ on ait : $|u_{\varphi(n)} - l| \leq \frac{\epsilon}{2}$.

Posons alors $N = \mathrm{Max}(N_1, N_2)$. Soit alors un $n \geq N$. Un lemme encore récent a montré que $\varphi(n) \geq n$, donc $\varphi(n) \geq n \geq N_1$; par construction de N on a aussi $n \geq N_2$. On en déduit que :

$$|u_n - l| = |(u_{\varphi(n)} - l) - (u_{\varphi(n)} - u_n)| \leq |u_{\varphi(n)} - u_n| + |u_{\varphi(n)} - l| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

•

Chapitre 20 - Quelques compléments sur les fonctions d'une variable réelle

Ce chapitre très court énonce quelques résultats qui me paraissaient un peu techniques pour être abordés au premier semestre –et étaient plus facilement énoncés, voire prouvés, en utilisant des suites. Maintenant que les suites sont connues et les résultats qui les concernent prouvés, il est temps de généraliser aux fonctions définies sur une partie de $\mathbf{R}$ autre que $\mathbf{N}$ des résultats connus sur les suites.

1 - Critère séquentiel pour l'étude des limites

Le résultat qui suit est bien pratique : il explique que si on sait étudier soigneusement les limites des suites, on sait travailler sur les limites de fonctions d'une variable réelle les plus générales.

Proposition 20-1-97 : Soit f une fonction réelle d'une variable réelle, définie sur un ensemble $\mathcal{D}_f$. Soit a un réel adhérent à $\mathcal{D}_f$ (ou soit a le symbole $+\infty$ ou $-\infty$, supposé adhérent à $\mathcal{D}_f$). Soit l un réel (ou le symbole $+\infty$, ou le symbole $-\infty$).

Alors $f(t) \to l$ quand $t \to a$ si et seulement si pour toute suite (u_n) de points de $\mathcal{D}_f$ vérifiant $\lim\limits_{n\to\infty} u_n = a$, $f(u_n) \to l$.

Démonstration : Comme j'en ai pris l'habitude, je ne traiterai pas tous les cas a infini ou l infini et les laisserai "au lecteur". Pour changer un peu, je laisserai cette fois le cas l fini au lecteur et traite le cas d'un a fini mais où l est le symbole $+\infty$.

• Preuve de $\Rightarrow$: Il n'y a presque rien à faire, c'est un simple problème de composition de limites. La suite (u_n) tend vers a quand $n \to \infty$ et $f(t) \to +\infty$ quand $t \to a$, donc la règle de composition des limites donne aussitôt le résultat.

• Preuve de $\Leftarrow$: On va prouver cette implication par contraposition. Supposons donc que $f(t)$ ne tende pas vers $+\infty$ quand $t \to a$. C'est donc qu'il existe un $A \in \mathbf{R}$ tel que pour tout $\epsilon > 0$, il existe un $t \in \mathcal{D}_f$ tel que $|t-a| \leq \epsilon$ mais que pourtant $f(t) < A$. Pour $n \geq 1$, appliquons cette propriété à $\epsilon = \dfrac{1}{n}$: elle fournit au moins un u_n tel que $|t-a| \leq \dfrac{1}{n}$ mais que pourtant $f(u_n) < A$. La suite formée de ces u_n (définie pour $n \geq 1$, mais on peut prendre u_0 n'importe comment si on tient à être indexé par $\mathbf{N}$) tend alors vers a et pourtant $f(u_n)$, constamment majoré par A, ne peut tendre vers l'infini.

•

2 - Propriété de la limite monotone

Il s'agit de répéter pour des fonctions le théorème déjà montré pour des suites soit en résumé "croissant et majoré entraîne convergent". Attention à un piège toutefois : pour des suites, la variable n tend vers $+\infty$ et le théorème marchera de même pour des fonctions d'une variable t tendant vers $+\infty$ et aussi pour une limite **à gauche** finie. En revanche, pour prouver le résultat analogue en $-\infty$ ou en une limite à droite finie, c'est la minoration qui devra intervenir : regardez l'exemple de $f(t) = -\dfrac{1}{t}$ sur $\mathbf{R}^{+*}$ qui est indéniablement croissante et majorée, et admet bien une limite en $+\infty$ mais n'en admet pas en 0. Regardez aussi l'exemple de l'exponentielle qui admet une limite en $-\infty$ sans être majorée mais bien parce qu'elle est croissante **minorée** sur $\mathbf{R}$.

Proposition 20-2-98 : Soit a un réel ou le symbole $+\infty$ et f une fonction réelle d'une variable réelle, définie sur un ensemble $\mathcal{D}_f$ tel que a soit adhérent à $]-\infty, a[\cap \mathcal{D}_f$. On suppose f croissante. Alors

* Si f n'est pas majorée sur $]-\infty, a[\cap \mathcal{D}_f$, $f(t) \to \infty$ quand $t \to a,\ t < a$.

* Si f est majorée sur $]-\infty, a[\cap \mathcal{D}_f$, $f(t)$ admet une limite quand $t \to a,\ t < a$. ; sa limite est égale à $\operatorname*{Sup}\limits_{\substack{t\in\mathcal{D}_f\\ t<a}} f(t)$.

Démonstration : C'est exactement la même que pour des suites. Je la laisse au lecteur (en l'invitant à ne pas se perdre dans les notations et décider s'il est en train de démontrer le résultat pour un a réel ou en l'infini).

•

3 - Le critère de Cauchy pour des fonctions

Je n'ai pas traité ce point en amphi ; un collègue me l'ayant fait remarquer, je l'ajoute dans la version papier. Il s'agit de choses un peu délicates et pourtant essentielles dès la deuxième année, qui serviront notamment en deuxième année pour montrer l'existence de limites pour des fonctions de la forme $f(x) = \int_a^x g(t)dt$. Si vous comprenez bien les énoncés analogues sur les suites, cette section ne devrait pas vous poser de problème ; si vous ne les avez pas encore bien assimilés, remettez plutôt le métier sur l'ouvrage et relisez le paragraphe consacré aux suites de Cauchy plutôt que de vous obstiner sur celui-ci qui n'en est qu'un clone un peu moins lisible.

Le critère peut être énoncé en un point fini ou en $+-\infty$, pour éviter toute confusion, voici les deux définitions :

Définition 20-3-155 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$, et soit a un réel adhérent à $\mathcal{D}_f$. On dit que f **vérifie le critère de Cauchy** en a lorsque :

$$\text{pour tout } \epsilon > 0, \text{ il existe } \eta > 0 \text{ tel que pour tous } s, t \in \mathcal{D}_f, (|s-a| \leq \epsilon \text{ et } |t-a| \leq \epsilon| \Rightarrow (|f(s) - f(t)| \leq \epsilon)\,.$$

Définition 20-3-156 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$,on suppose que $+\infty$ est adhérent à $\mathcal{D}_f$. On dit que f **vérifie le critère de Cauchy** en $+\infty$ lorsque :

$$\text{pour tout } \epsilon > 0, \text{ il existe un réel } A \text{ tel que pour tous } s, t \in \mathcal{D}_f, (A \leq s \text{ et } A \leq t) \Rightarrow (|f(s) - f(t)| \leq \epsilon)\,.$$

Comme pour les suites, cette notion équivaut à la convergence :

Théorème 20-3-37 : Soit f une fonction réelle d'une variable réelle, définie sur l'ensemble $\mathcal{D}_f$, et soit a un réel adhérent à $\mathcal{D}_f$. La fonction f admet une limite en a si et seulement si elle y vérifie le critère de Cauchy. De même en $\pm\infty$.

Démonstration : (ou plutôt indications de démonstration)

• Preuve de $\Rightarrow$. Comme pour les suites, c'est le sens "facile" : on applique la définition de "convergence" à $\frac{\epsilon}{2}$ et on applique une fois l'inégalité triangulaire.

• Preuve de $\Leftarrow$ Il y a de quoi rester perplexe : pour les suites, on s'en était tiré en pensant à Bolzano-Weierstrass, mais on ne connaît rien d'analogue pour des fonctions. L'astuce est tout simplement d'utiliser encore Bolzano-Weierstrass. Comme a est adhérent à $\mathcal{D}_f$, pour tout $n \geq 1$, on peut trouver un u_n qui soit dans $\mathcal{D}_f$ et tel que $|u_n - a| \leq \frac{1}{n}$. La suite (u_n) converge donc vers a. En regardant la définition du critère de Cauchy, on se convainc qu'elle entraîne que la suite $f(u_n)$ est une suite de Cauchy. Donc cette suite converge, vers une limite l. On vérifie ensuite, en utilisant de nouveau le critère de Cauchy, que l est bien la limite de f au point a (mêmes calculs que dans la démonstration analogue pour des suites).

Comme j'en ai pris l'habitude, je considère comme facile de modifier ces démonstrations pour une limite en l'infini.

•

Chapitre 21 - Fonctions continues, deuxième couche

Au premier semestre, on n'avait vu que des choses "faciles" sur les fonctions continues. Comme je ne pouvais me passer d'un théorème important mais difficile à démontrer –le théorème référencé 6-4-10– il avait été énoncé sans preuve ; la preuve arrive ainsi que tout plein d'autres.

1 - Critère séquentiel de continuité

En appliquant le premier résultat du chapitre précédent, de même qu'on a maintenant une méthode utilisant les suites pour prouver l'existence d'une limite, on a une méthode utilisant les suites pour prouver la continuité d'une fonction d'une variable réelle.

Proposition 21-1-99 : Soit f une fonction réelle d'une variable réelle définie sur un ensemble $\mathcal{D}_f$ et soit a un point de $\mathcal{D}_f$, supposé adhérent à $\mathcal{D}_f \setminus \{a\}$. Alors

f est continue en a si et seulement si pour toute suite (u_n) de points de $\mathcal{D}_f$ vérifiant $\lim\limits_{n\to\infty} u_n = a$, $f(u_n) \to f(a)$ quand $n \to \infty$.

Démonstration : Elle sera un peu lourde, je paie ici le choix que j'ai fait de prendre une définition de la continuité se voulant plus intuitive que celle de la plupart des sources, mais en échange un peu moins adroite. Ne vous cassez pas la tête à analyser de trop près les difficultés de cette preuve, elles me paraissent fort peu instructives.

Par défnition de la continuité, f est continue au point a signifie que $f(t) \to f(a)$ quand $t \to a,\ t \neq a$. En appliquant le critère séquentiel d'existence d'une limite du chapitre précédent, cette propriété équivaut à :

"pour toute suite (u_n) de points de $\mathcal{D}_f \setminus \{a\}$ vérifiant $\lim\limits_{n\to\infty} u_n = a$, $f(u_n) \to f(a)$ quand $n \to \infty$".

C'est presque ce qu'il fallait démontrer, mais pas tout à fait car ici il ne s'agit que de suites (u_n) auxquelles la valeur a est interdite, alors que dans l'énoncé les suites (u_n) vivent n'importe où dans $\mathcal{D}_f$, y compris sur le point a. Il est toutefois clair que si la propriété séquentielle est vraie pour des suites quelconques de $\mathcal{D}_f$ elle est vraie *a fortiori* pour des suites évitant la valeur a et donc f est continue en a.

Réciproquement, supposons connue la propriété pour les seules suites évitant la valeur a et soit (u_n) une suite dans $\mathcal{D}_f$ tendant vers a. S'il existe un N tel que pour tout $n \geq N$, $u_n \neq a$, on en déduit aussitôt que $f(u_n) \to f(a)$; s'il existe un N tel que pour tout $n \geq N$, $u_n = a$ il est encore plus idiot que $f(u_n) \to f(a)$. Le seul cas à problème est celui où l'ensemble des indices n tels que $u_n = a$ et celui des indices n tels que $u_n \neq a$ sont tous deux infinis ; dans ce cas on peut remarquer que sur chacun de ces ensembles $f(u_n) \to f(a)$ et conclure par le théorème de recollement de deux limites (celui numéroté 5-4-29).

•

2 - Fonctions continues sur les intervalles fermés bornés

Par un simple copier-coller, je rappelle le

Théorème 6-4-10 : Soit f une fonction réelle continue d'une variable réelle définie sur un intervalle **fermé borné** [a,b]. Alors il existe un $c_- \in [a,b]$ et un $c_+ \in [a,b]$ tel que pour tout $t \in [a,b]$ on ait :

$$f(c_-) \leq f(t) \leq f(c_+).$$

En d'autres termes, avec le vocabulaire désormais acquis concernant la relation d'ordre sur $\mathbf{R}$: la fonction continue f est bornée, et l'ensemble des valeurs qu'elle prend admet un plus grand et un plus petit élément.

La nouveauté par rapport au premier semestre, c'est la

Démonstration :

• Dans un premier temps, montrons que f est majorée (et minorée de façon analogue). Supposons qu'elle ne le soit pas. Alors pour chaque n de $\mathbf{N}$, n ne serait pas un majorant de f, donc il existerait un point x_n dans $[a,b]$ tel que $f(x_n) \geq n$. De la suite (x_n), le théorème de Bolzano-Weierstrass permet d'extraire une sous-suite $(x_{\varphi(n)})$ convergente dans $\mathbf{R}$ vers une limite l. Comme pour chaque $n \geq 0$ on a les inégalités $a \leq x_{\varphi(n)} \leq b$, par passage à la limite on obtient $a \leq l \leq b$. Ainsi le point l est dans l'ensemble de définition de f (c'est l'endroit où on utilise, discrètement mais efficacement, la fermeture de l'intervalle), et f est donc continue

en l. Comme f est continue en l et que $x_{\varphi(n)} \to l$ quand $n \to \infty$, par le critère séquentiel de continuité $f(x_{\varphi(n)}) \to f(l)$. Mais pourtant pour chaque n, $f(x_{\varphi(n)}) \geq \varphi(n)$ qui tend vers $+\infty$. Contradiction !
• La fonction f est bornée, et un segment fermé est non vide. L'ensemble $f([a,b])$ des valeurs prises par f est donc borné non vide, et admet donc un Sup qu'on notera M (et de façon analogue un Inf). Il nous reste à montrer que ce Sup est en fait un Max. Pour ce faire, il suffit de recommencer la construction qui nous a si bien réussi dans la première partie de la démonstration en l'adaptant d'un cheveu : comme M est le plus petit majorant de f, pour tout $n \geq 1$ il existe un x_n dans $[a,b]$ tel que $f(x_n) \geq M - \dfrac{1}{n}$. De la suite (x_n), le théorème de Bolzano-Weierstrass permet d'extraire une sous-suite $(x_{\varphi(n)})$ convergente dans $\mathbf{R}$ vers une limite c_+. Comme pour chaque $n \geq 0$ on a les inégalités $a \leq x_{\varphi(n)} \leq b$, par passage à la limite on obtient $a \leq c_+ \leq b$. Ainsi le point c_+ est dans l'ensemble de définition de f, et f est donc continue en c_+. Comme f est continue en c_+ et que $x_{\varphi(n)} \to c_+$ quand $n \to \infty$, par le critère séquentiel de continuité $f(x_{\varphi(n)}) \to f(c_+)$. Or pour chaque n, $M - \dfrac{1}{\varphi(n)} \leq f(x_{\varphi(n)}) \leq M$, et la limite des $f(x_{\varphi(n)})$ est donc égale à M. Ceci prouve que $f(c_+) = M$ et donc que f admet mieux qu'un Sup : un Max. On construit de même c_-.

•

3 - Le théorème des valeurs intermédiaires

Ce théorème terriblement intuitif assure que si un poste de péage est installé au milieu d'une autoroute, même les automobilistes les plus imaginatifs ne parviendront pas à faire tout le trajet sans verser leur obole.
Théorème 21-3-38 : (des valeurs intermédiaires) Soit I un **intervalle** de $\mathbf{R}$ et f une fonction continue de I vers $\mathbf{R}$. Soit $a \leq b \leq c$ trois réels. Si f prend les valeurs a et c, elle prend aussi la valeur b.
Démonstration :
• Débarrassons-nous d'un cas stupide : si l'une des inégalités $a \leq b$ ou $b \leq c$ n'est pas stricte, le théorème est évident. On supposera donc dans la suite $a < b < c$.
• Soit α et γ deux points de I tel que $f(\alpha) = a$ et $f(\gamma) = c$. Dans la suite, on supposera $\alpha < \gamma$ (si l'inégalité était dans l'autre sens, on peut regarder $-f$). Comme I est un intervalle, le segment $[\alpha,\gamma]$ est entièrement inclus dans I, et on pourra parler de $f(t)$ pour n'importe quel t de $[\alpha,\gamma]$ sans se faire de souci sur son existence.

Notons alors $F = \{t \in [\alpha,\gamma] \mid f(t) \leq b\}$. Cet ensemble F n'est pas vide, puisqu'il contient α (on a bien $f(\alpha) = a < b$). Il est majoré puisque tous ses points sont plus petits que γ. Il admet donc un Sup qu'on appellera β (et qui est inférieur ou égal à γ). Au vu de cette notation, le lecteur se doute bien qu'on va tenter de prouver que $f(\beta) = b$, ce qui prouvera le théorème.

Pour ce faire, on va prouver une double inégalité :

* Preuve que $f(\beta) \leq b$. Soit $n \geq 1$, alors $\beta - \dfrac{1}{n}$ est strictement inférieur à β, donc n'est pas un majorant de F ; il existe donc un point t_n dans F tel que $\beta - \dfrac{1}{n} < t_n$ –et, comme β majore F, on a en outre $t_n \leq \beta$. On en conclut que t_n tend vers β quand n tend vers $+\infty$. Mais pour chaque n, comme $t_n \in F$, $f(t_n) \leq b$. Enfin f est continue, donc continue au point β. On peut donc passer à la limite et obtenir : $f(\beta) \leq b$.

* Preuve que $b < f(\beta)$. Maintenant que nous savons que $f(\beta) \leq b$, nous en déduisons notamment que $\beta \neq \gamma$ et donc que $\beta < \gamma$.
Dès lors, pour n assez grand (très précisément pour n plus grand que $\dfrac{1}{\gamma - \beta}$), le réel $\beta + \dfrac{1}{n}$ est dans l'intervalle $[\alpha,\gamma]$ tout en étant strictement plus grand que β. Il n'est donc pas dans F, qui est majoré par β et donc $f\left(\beta + \dfrac{1}{n}\right) > b$ (en se référant à la définition de F). Par passage à la limite dans cette inégalité, en utilisant encore la continuité de f au point β, on obtient : $f(\beta) \geq b$.

On a donc prouvé que $f(\beta) = b$ qui est donc bien une valeur prise par f.

•

Remarque : On peut énoncer ce théorème de façon plus concise sous la forme : " l'image d'un intervalle par une fonction continue est un intervalle". Avec la définition que j'ai donnée du mot "intervalle" l'équivalence entre cette forme et celle j'ai mise en relief est de la simple traduction, sans aucun effort de réflexion.

Sur mon élan, je donne une deuxième démonstration du théorème des valeurs intermédiaires. Elle repose sur l'utilisation de concepts liés à la dérivation, donc inutilement savants pour utiliser des fonctions continues. Selon ce qu'on considère comme beau, on pourra donc la trouver bien plus laide que la première, comme utilisant des notions inutilement difficiles, ou au contraire bien plus jolie puisqu'allant chercher un outil inattendu. On vérifiera bien qu'il n'y a pas de tête-à-queue logique et que la démonstration n'utilise pas discrètement le théorème des valeurs intermédiaires : ce que nous savons sur les fonctions dérivables utilise le théorème de la section précédente, mais pas le théorème des valeurs intermédiaires.

Deuxième démonstration : Comme dans la première démonstration, on commence par se limiter au seul cas intéressant où $a < b < c$. On notera encore α un réel tel que $f(\alpha) = a$ et γ un réel tel que $f(\gamma) = c$.

On va faire une démonstration par l'absurde ; supposons donc que f ne prenne pas la valeur b.

Introduisons la fonction g définie sur $\mathbf{R} \setminus \{b\}$ par $g(y) = -1$ si $y < b$ et $g(y) = 1$ si $b < y$. La fonction g est continue sur son ensemble de définition, et, comme on a supposé que f ne prend pas la valeur b, $g \circ f$ existe. La fonction $g \circ f$ est donc une fonction continue de I vers $\mathbf{R}$. De plus, comme g ne prend que les valeurs -1 et 1, il en est de même de $g \circ f$.

Soit x un point de I. En appliquant la définition de "continuité" à $g \circ f$ au point x pour $\epsilon = 1$, on voit qu'il existe un $\eta > 0$ tel que pour tout h tel que $|h| < \eta$, $|(g \circ f)(x+h) - (g \circ f)(x)| \leq 1$. Il est donc impossible que $g(x)$ vaille 1 tandis que $g(x + h)$ vaut -1, ou réciproquement que $g(x)$ vaille -1 tandis que $g(x + h)$ vaille 1. Ainsi pour tout h tel que $|h| < \eta$, $(g \circ f)(x + h) = (g \circ f)(x)$.

De ce fait, le taux de variation $\dfrac{(g \circ f)(x + h) - (g \circ f)(x)}{h}$ est nul pour tout h tel que $|h| < \eta$. Sa limite quand h tend vers 0 existe donc de façon très évidente et est nulle. De façon fort suprenante, la fonction $g \circ f$ se révèle donc dérivable, et de dérivée nulle (alors qu'on n'avait aucune hypothèse de dérivabilité concernant f).

L'ensemble de départ étant un intervalle (c'est ici qu'on s'en sert), la fonction $g \circ f$, dérivable de dérivée nulle, est donc constante. On en déduit en particulier que $(g \circ f)(\alpha) = (g \circ f)(\gamma)$. Mais $(g \circ f)(\alpha) = g(a) = -1$ puisque $a < b$ et vu la définition de g, tandis que $(g \circ f)(\gamma) = g(c) = 1$. Contradiction !

•

4 - Fonctions continues et monotonie

Cette section est conçue pour que vous n'ayez à mémoriser que les trois théorèmes qui la terminent. Les étudiants très méticuleux pourront néanmoins faire des efforts pour tout apprendre, mais ce sont des efforts un peu gaspillés (les plus observateurs remarqueront toutefois que la deuxième proposition ne suppose qu'une monotonie alors que les théorèmes qui closent la section sont basés sur une stricte monotonie, et donc qu'elle contient quelques grammes d'information supplémentaires, pouvant justifier un effort de mémoire).

Une remarque aussi pour les puristes : dans le choix que j'ai fait pour définir les fonctions continues, les fonctions continues sur un singleton ne sont pas définies ; de ce fait, certains des théorèmes que j'énonce ci-dessous n'ont pas de sens dans l'hypothèse d'un intervalle I réduit à un point. Mais comme vous le verrez certainement bientôt, en définissant un peu autrement la continuité, ce problème disparaît. J'ai donc "anticipé" sur vos connaissances et n'ai pas restreint les hypothèses aux intervalles non réduits à des points. Le cas manquant étant bien sûr toujours stupide et sans intérêt.

Proposition 21-4-100 : Soit f une fonction réelle d'une variable réelle, définie sur une partie $\mathcal{D}_f$. Si f est strictement monotone, alors elle est injective.

Démonstration : C'est tellement évident que j'ai hésité à énoncer ce résultat : pour $x \neq x'$ dans $\mathcal{D}_f$, si $x < x'$ et f est strictement croissante, on en déduit que $f(x) < f(x')$ et donc que $f(x) \neq f(x')$, et on traite de même les trois autres cas.

•

Proposition 21-4-101 : Soit f une fonction réelle d'une variable réelle définie sur un intervalle I. Si f est monotone et $f(I)$ est un intervalle, alors f est continue sur I.

Démonstration : Soit x un point de I ; nous devons montrer que f est continue en x. Il nous suffit de montrer que f est continue à gauche et à droite en x, étant entendu que si x est l'extrémité gauche éventuelle de I nous n'avons à montrer que la continuité à droite –seule à avoir un sens– et de même si x est l'extrémité droite de I.

La démonstration sera écrite pour une f croissante, la preuve étant évidemment analogue pour f décroissante (ou, si on préfère, le résultat pouvant alors être appliqué à $-f$). De même, je n'écrirai que la preuve de la continuité à gauche, celle de la continuité à droite étant analogue (ou pouvant être obtenue en examinant l'application $x \mapsto -f(-x)$).

Notons I_- l'intervalle $I \cap]-\infty, x[$ (non vide dès lors que x n'est pas l'extrémité droite de I, donc auquel x est adhérent). Sur cet intervalle, la fonction f est croissante. De plus elle est majorée, par $f(x)$ puisque par croissance de f, pour tout $t \in I_-$, comme $t < x$, $f(t) \leq f(x)$. D'après la propriété de la limite monotone, $f(t)$ admet donc une limite à gauche quand t tend vers x, $t < x$, et cette limite l_- est inférieure ou égale à $f(x)$ puisque $l_= \operatorname*{Sup}_{t \in I_-} f(t)$ et que $f(x)$ est un majorant de tous les $f(t)$, $t \in I_-$.

Supposons que l_- ne soit pas égal à $f(x)$ et considérons alors un réel m strictement compris entre les deux, par exemple $m = \dfrac{l_- + f(x)}{2}$.

Alors pour tout $t < x$, $f(t)$ est inférieur ou égal à l_-, donc n'est pas égal à m. De plus il existe au moins un $t < x$ qui soit dans I, donc f prend effectivement au moins une valeur inférieure à m.

De l'autre côté de x, pour $x \leq t$, $f(x) \leq f(t)$ (par croissance) donc $f(t)$ n'est pas non plus égal à m. De plus au point x très précisément, f prend une valeur supérieure à m.

Mais alors f prend à la fois une valeur inférieure à m et une valeur supérieure à m tout en évitant m. Ceci contredit l'hypothèse selon laquelle l'ensemble des valeurs prises par f est un intervalle. C'est donc qu'on avait $l_- = f(x)$ et la continuité à gauche en x est prouvée.

•

Proposition 21-4-102 : Soit f une fonction réelle d'une variable réelle définie sur un intervalle I. Si f est continue et injective sur I, alors elle est strictement monotone sur I.

Démonstration : Elle se prête bien à être découpée en un certain nombre d'étapes intermédiaires ayant des énoncés clairement distingués.

• **Première étape** : Chaque fois qu'on prend trois points distincts $x_1 < x_2 < x_3$ dans I, la valeur $f(x_2)$ n'est ni la plus petite, ni la plus grande des trois valeurs prises par f en ces trois points.

Preuve de la première étape : Notons tout d'abord que par l'injectivité de f, les trois valeurs $f(x_1)$, $f(x_2)$ et $f(x_3)$ sont distinctes. Supposons que $f(x_2)$ soit la plus petite des trois, et notons la a. Notons b la suivante dans l'ordre croissant : ainsi $b \leq f(x_1)$ et $b \leq f(x_3)$ (avec égalité dans un des cas, et inégalité stricte dans l'autre).

Comme I est un intervalle, l'intervalle $[x_1, x_2]$ est inclus dans I. Comme f est continue sur I, sa restriction à $[x_1, x_2]$ est également continue et on peut lui appliquer le théorème des valeurs intermédiaires. Comme elle prend la valeur $a < b$ (en x_2) et une valeur supérieure ou égale à b (en x_1) elle prend aussi la valeur b quelque part ; comme ce ne peut être en x_2 on a montré qu'il existe un t dans $[x_1, x_2[$ tel que $f(t) = b$.

On fait exactement de même à droite de x_2 et on obtient un u dans $]x_2, x_3]$ tel que $f(u) = b$. Mais ceci contredit l'injectivité de f.

On éliminerait évidemment de même le cas où $f(x_2)$ serait la plus grande des trois valeurs.

• **Deuxième étape** : Sur toute partie finie de I à trois éléments, la restriction de f est strictement monotone.

Justification de la deuxième étape : ce n'est qu'une reformulation de la première étape. Soit en effet $A = \{x_1, x_2, x_3\}$ une partie finie de I à trois éléments, où $x_1 < x_2 < x_3$. Par la première étape, les seules façons dont peuvent s'agencer les trois valeurs prises par f sur cette partie sont soit $f(x_1) < f(x_2) < f(x_3)$ –et alors la restriction de f à A est strictement croissante, soit $f(x_3) < f(x_2) < f(x_1)$ –et alors la restriction de f à A est strictement décroissante.

• **Troisième étape** : Sur toute partie finie de I à quatre éléments au plus, la restriction de f est strictement monotone.

Preuve de la troisième étape : elle est évidente si la partie considérée a moins de deux éléments, et c'est la deuxième étape si elle en a trois. Soit donc une partie finie $A = \{x_1, x_2, x_3, x_4\}$ de I à exactement quatre éléments, où $x_1 < x_2 < x_3 < x_4$.

On fait deux cas selon la position relative de x_2 et x_3. Supposons $f(x_2) < f(x_3)$. En appliquant la deuxième étape sur $\{x_1, x_2, x_3\}$ on conclut que $f(x_1) < f(x_2)$ et en recommençant sur $\{x_2, x_3, x_4\}$ que $f(x_3) < f(x_4)$. Finalement f se révèle strictement croissante sur A. Réciproquement, si $f(x_3) < f(x_2)$, f se révèle de même strictement décroissante.

• **Fin de la preuve** : Prenons deux points distincts et fixes $\alpha < \beta$ dans I. Supposons $f(\alpha) < f(\beta)$. On va montrer que f est alors strictement croissante. Soit deux points $x < x'$ dans I. Appliquons l'étape précédente sur l'ensemble fini $\{\alpha, \beta, x, x'\}$ qui a au plus quatre éléments. Vu ce qui se passe sur α et β, c'est strictement croissante qu'est f sur cet ensemble. On en déduit que $f(x) < f(x')$: la croissance stricte de f est montrée. Bien évidemment, si on partait de $f(\beta) < f(\alpha)$, on obtiendrait de même la stricte décroissance. •

Comme promis, voilà maintenant les trois énoncés faciles à mémoriser qui découlent de ces propositions.

Théorème 21-4-39 : Soit f une fonction réelle d'une variable réelle définie sur un intervalle I.

On suppose que :

f est continue sur I.

f est strictement monotone sur I.

Alors il existe un intervalle J tel que la restriction de f de I vers J soit une bijection entre ces deux intervalles.

De plus, la bijection réciproque $f^{-1} : J \to I$, est alors continue.

Théorème 21-4-40 : Soit f une fonction réelle d'une variable réelle définie sur un intervalle I.

On suppose que :

f est strictement monotone sur I.

Il existe un intervalle J tel que la restriction de f de I vers J soit une bijection entre ces deux intervalles.

Alors f est continue sur I.

De plus, la bijection réciproque $f^{-1} : J \to I$, est alors continue.

Théorème 21-4-41 : Soit f une fonction réelle d'une variable réelle définie sur un intervalle I.

On suppose que :

Il existe un intervalle J tel que la restriction de f de I vers J soit une bijection entre ces deux intervalles.

f est continue sur I.

Alors f est strictement monotone sur I.

De plus, la bijection réciproque $f^{-1} : J \to I$, est alors continue.

Démonstration du théorème 21-4-39 : C'est une simple conséquence du théorème des valeurs intermédiaires. Comme f est une application continue, le théorème des valeurs intermédiaires (sous la forme donnée en remarque), assure que $f(I)$ est un intervalle. Posons $J = f(I)$. Alors par construction de J, la restriction de f de I vers J est surjective. Par la proposition 21-4-100, elle est injective. Et donc bijective. La preuve du complément sera commune aux trois théorèmes. •

Démonstration du théorème 21-4-40 : L'ensemble $f(I)$ est égal à J et est donc un intervalle. L'application f est en outre monotone : toutes les hypothèses de la proposition 21-4-101 sont donc remplies et on peut conclure que f est continue. La preuve du complément sera commune aux trois théorèmes. •

Démonstration du théorème 21-4-41 : La restriction de f de I vers J est une application injective, continue et définie sur un intervalle : par application de la proposition 21-4-102 elle est donc strictement monotone. Donc f aussi. La preuve du complément sera commune aux trois théorèmes. •

Preuve de la dernière phrase des trois théorèmes : l'application f^{-1} est définie sur un intervalle, à valeurs dans un intervalle, bijective, strictement monotone. On peut lui appliquer le deuxième théorème et obtenir la seule information non évidente à son sujet : sa continuité. •

5 - Dérivation d'une fonction réciproque

Ce résultat avait été (volontairement) oublié au premier semestre, car quoique la formule qui va être énoncée soit simple, importante, et couramment utilisée, les détails des hypothèses ne sont pas si simples (et la preuve plus subtile qu'il pourrait y paraître, utilisant les résultats qui précèdent).

Théorème 21-5-42 : Soit f une fonction réelle d'une variable réelle définie sur un intervalle I (non réduit à un point).

On suppose que f est continue sur I, et que la restriction de f de I vers l'intervalle $f(I)$ est bijective.

Soit a un point de I en lequel on suppose f dérivable, avec $f(a) \neq 0$.

Alors f^{-1} est dérivable en $f(a)$ et

$$\left(f^{-1}\right)'[f(a)] = \frac{1}{f'(a)}.$$

Démonstration : Examinons le taux de variation dont la dérivée éventuelle sera la dérivée de f^{-1} en $f(a)$, soit le quotient, défini pour $y \in J$, $y \neq f(a)$:

$$\tau(y) = \frac{f^{-1}(y) - f^{-1}[f(a)]}{y - f(a)} = \frac{f^{-1}(y) - a}{y - f(a)}.$$

Regardons concurremment le quotient dont on sait que la limite est la dérivée de f en a, soit le quotient, défini pour $x \in I$, $x \neq a$:

$$t(x) = \frac{f(x) - f(a)}{x - a}.$$

Calculons, pour $y \in J$, $y \neq f(a)$, l'expression $t[f^{-1}(y)]$ (comme f^{-1} est injective, $f^{-1}(y) \neq a$). On obtient :

$$t[f^{-1}(y)] = \frac{f(f^{-1}(y)) - f(a)}{f^{-1}(y) - a} = \frac{y - f(a)}{f^{-1}(y) - a} = \frac{1}{\tau(y)},$$

et donc

$$\tau(y) = \frac{1}{t[f^{-1}(y)]}$$

.

Mais quand y tend vers $f(a)$, $f^{-1}(y)$ tend vers a (on utilise ici, et c'est le point sensible de la démonstration, la continuité de f^{-1}), donc $t[f^{-1}(y)]$ tend vers $f'(a)$. Comme on a supposé ce réel non nul, $\tau(y)$ tend vers $\dfrac{1}{f'(a)}$.

•

Remarque : Pour énoncer la formule qui conclut ce théorème, on peut préférer donner une notation $b = f(a)$, et donc $a = f^{-1}(b)$; en centrant les notations sur b la formule devient :

$$\left(f^{-1}\right)'(b) = \frac{1}{f'[f^{-1}(b)]}.$$

En particulier dans le cas fréquent où f est dérivable en tous les points de I, et où sa dérivée ne s'annule nulle part dans I, on obtient une identité fonctionnelle :

$$\left(f^{-1}\right)' = f' \circ f^{-1}.$$

Chapitre 22 - Fonctions convexes

Ceux qui ont assisté au cours ont pu recopier les dessins que j'ai faits au tableau ; ceux qui ont préféré rester au lit et tentent en lisant ce polycopié de rattraper leur retard sont avertis : il s'agit d'un chapitre à contenu très géométrique, et sans dessiner tout ce qui est décrit ils n'y comprendront pas grand chose.

1 - Quelques préliminaires

Le fait suivant résulte de l'application à $\mathbf{R}$ de la théorie des barycentres, qui est connue si j'ai bien compris, mais il coûte peu cher de le redémontrer.

Proposition 22-1-103 : Soit $a < b$ deux réels ; $]a,b[= \{(1-t)a+tb \mid 0 < t < 1\}$.

Démonstration : Par double inclusion. Si $a < u < b$, posons $t = \dfrac{u-a}{b-a}$; comme $a < u$ et $a < b$, il est clair que $0 < u$. On calcule $1 - t = 1 - \dfrac{u-a}{b-a} = \dfrac{b-a}{b-a} - \dfrac{u-a}{b-a} = \dfrac{b-u}{b-a}$ et de $u < b$ et $a < b$ il est alors clair que $t < 1$. On vérifie stupidement que

$$(1-t)a + tb = \frac{b-u}{b-a}a + \frac{u-a}{b-a}b = \frac{ab - au + bu - ab}{b-a} = u.$$

Réciproquement, soit t un réel avec $0 < t < 1$ et notons $u = (1-t)a + tb$; on vérifie encore stupidement que $u - a = t(b-a)$ est strictement positif parce que t et $b - a$ le sont, et que $b - u = (1-t)(b-a)$ l'est parce que $1 - t$ et $b - a$ le sont.

•

Je pourrais sans doute me dispenser de donner la définition qui suit, sa clarté géométrique étant totale, mais il vaut mieux savoir ce qu'on suppose exactement pour donner des démonstrations bien vérifiables...

Définition 22-1-157 : Soit Δ une droite non verticale de $\mathbf{R}^2$ et $B = (x_B, y_B)$ un point de $\mathbf{R}^2$. Notons $P = (x_B, y_P)$ l'unique point de Δ de même abscisse que B. On dira que B est **au-dessous de** Δ lorsque $y_B \leq y_P$.

Le résultat qui suit est évident sur un dessin, et sa preuve est une simple vérification ; ce sera notre lien entre la lourde définition qui précède et son utilisation.

Proposition 22-1-104 : Soit $A = (x_A, y_A), B = (x_B, y_B), C = (x_C, y_C)$ trois points de $\mathbf{R}^2$. Alors

* Sous l'hypothèse supplémentaire $x_A < x_B$,

$$\text{Pente de } (AB) \leq \text{ Pente de } (AC) \iff B \text{ est au-dessous de } (AC)$$

* Sous l'hypothèse supplémentaire $x_B < x_C$,

$$B \text{ est au-dessous de } (AC) \iff \text{Pente de } (AC) \leq \text{Pente de } (BC)$$

Démonstration : Comme x_B est un point du segment de l'intervalle $]x_A, x_C[$ il existe un t avec $0 < t < 1$ tel que $x_B = (1-t)x_A + tx_C$. Soit P le point de (AC) qui a même abscisse que B ; notons $P = (x_B, y_P)$.

Alors la pente de (AC) est aussi la pente de (AP) : elle est donc égale à $\dfrac{y_P - y_A}{x_B - x_A}$; par ailleurs la pente de (AB) vaut $\dfrac{y_B - y_A}{x_B - x_A}$.

$$\begin{aligned}\text{Dès lors, Pente de } (AB) \leq \text{ Pente de } (AC) \iff \frac{y_B - y_A}{x_B - x_A} \leq \frac{y_P - y_A}{x_B - x_A} &\iff y_B - y_A \leq y_P - y_A \\ &\iff y_B \leq y_P \\ &\iff B \text{ est au-dessous de } (AC).\end{aligned}$$

(L'hypothèse $x_A < x_B$ ayant discrètement servi pour multiplier par $x_B - x_A$ sans changer le sens de l'inégalité).

La deuxième équivalence se traite de façon analogue.

•

2 - Définition des fonctions convexes

Définition 22-2-158 : On dit qu'une fonction réelle d'une variable réelle f, définie sur un intervalle est **convexe** lorsqu'elle vérifie la propriété suivante : pour tous points $A = (x_A, y_A)$, $B = (x_B, y_B)$ et $C = (x_C, y_C)$ du graphe de f tels que $x_A < x_B < x_C$, le point B est au-dessous de la droite (AC).

La caractérisation suivante est très visuelle géométriquement ; c'est une variante infinitésimale de la définition mais elle nous sera utile pour les preuves.

Proposition 22-2-105 : Soit f une fonction réelle d'une variable réelle, définie sur un intervalle. La fonction f est convexe si et seulement si elle vérifie la propriété suivante : pour tous points $A = (x_A, y_A)$, $B = (x_B, y_B)$ et $C = (x_C, y_C)$ du graphe de f tels que $x_A < x_B < x_C$,

$$\text{Pente de } (AB) \leq \text{Pente de } (BC).$$

Démonstration : Supposons f convexe, et soit A, B, C trois points comme dans l'énoncé. Par définition des fonctions convexes, B est au-dessous de (AC) donc, par la proposition qui a clôturé les préliminaires, la pente de (AB) est plus faible que celle de (AC), elle-même plus faible que celle de (BC). L'inégalité souhaitée entre pentes est bien réalisée.

Réciproquement, supposons que f vérifie la propriété de l'énoncé, et soit A, B, C trois points comme dans cette propriété. Appliquons la proposition clôturant les préliminaires à B, C et A (dans cet ordre) : comme $x_B < x_C$ et que la pente de (BC) est plus forte que celle de (AB), C est au-dessus de (AB). Appliquons la maintenant à A, C et B dans cet ordre ; puisque $x_A < x_C$ et que C est au-dessus de (AB), la pente de (AC) est plus forte que celle de (AB). Appliquons une dernière fois cet énoncé à A, B et C dans cet ordre : comme $x_A < x_B$ et que la pente de (AC) est plus forte que celle de (AB), on en déduit que B est au-dessous de (AC) : le critère choisi pour définition de la convexité est vérifié.

•

3 - La convexité vue à travers des formules

Je n'ai pas choisi d'insister sur ce point de vue, mais il est lui aussi tout à fait fondamental pour étudier les fonctions convexes...

Proposition 22-3-106 : Soit f une fonction réelle d'une variable réelle, définie sur un intervalle I. Alors f est convexe si et seulement si pour tous points u, v de I et tout λ tel que $0 \leq \lambda \leq 1$,

$$f((1-\lambda)u + \lambda v) \leq (1-\lambda)f(u) + \lambda f(v).$$

Démonstration : Remarquons tout d'abord que si l'inégalité est vraie pour tous les $u < v$ elle est vraie pour tous les u, v de I : si $u > v$, il suffit de l'appliquer pour v et u (dans cet ordre) et $1 - \lambda$ pour retomber sur le résultat cherché ; et si $u = v$ l'inégalité est évidente. Remarquons aussi que l'inégalité est vérifiée de façon évidente pour λ valant 0 ou 1 et que sa réalisation pour tous u, v et tout λ tel que $0 \leq \lambda \leq 1$ équivaut donc à sa seule réalisation pour tous $u < v$ et tout λ tel que $0 < \lambda < 1$.

Supposons f convexe et soit $u < v$ deux points de I et λ un élément de $]0, 1[$; notons A le point du graphe de f d'abscisse u, c'est-à-dire $A = (u, f(u))$, B le point du graphe de f d'abscisse $(1-\lambda)u + \lambda v$, c'est-à-dire $B = ((1-\lambda)u + \lambda v, f((1-\lambda)u + \lambda v))$ et C le point du graphe de f d'abscisse v, c'est-à-dire $C = (v, f(v))$.

D'après la proposition préliminaire décrivant paramétriquement les intervalles de $\mathbf{R}$, l'abscisse de B est comprise strictement entre celles de A et de C. La définition de la convexité s'applique donc à ces trois points. On a calculé l'ordonnée de B : c'est $f((1-\lambda)u + \lambda v)$, reste à connaître l'ordonnée du point de (AC) qui se trouve à la verticale de B. Soit Q le point $((1-\lambda)u + \lambda v, (1-\lambda)f(u) + \lambda f(v)) = (1-\lambda)A + \lambda C$. Il est clair que Q a même abscisse que B ; par ailleurs $Q - A = \lambda(C - A)$ est visiblement colinéaire à $C - A$ et donc Q est sur la droite (AC). Écrire que B est au-dessous de (AC) c'est écrire que son ordonnée est plus faible que celle de Q : c'est exactement écrire l'inégalité recherchée.

Réciproquement, supposons l'inégalité vraie pour tous $u < v$ points de I et tout λ de $]0, 1[$. Soit $A = (x_A, y_A)$, $B = (x_B, y_B)$ et $C = (x_C, y_C)$ trois points du graphe de f tels que $x_A < x_B < x_C$; notons $u = x_A$ et $v = x_C$, et prenons un λ strictement entre 0 et 1 tel que $x_B = (1-\lambda)u + \lambda v$. Alors en reprenant les considérations qui précèdent, on voit que l'inégalité connue sur u, v et λ a l'interprétation géométrique souhaitée : B est bien au-dessous de (AC).

•

4 - Convexité et continuité

Théorème 22-4-43 : Soit f une fonction réelle d'une variable réelle, définie sur un intervalle **ouvert** I. On suppose f convexe.

Alors f est dérivable à gauche et à droite en tout point de I, donc continue.

De plus pour tout t de I, $f'_g(t) \leq f'_d(t)$ et les fonctions f'_g et f'_d sont toutes deux croissantes.

Démonstration : Soit t_0 un point de I. On va montrer la dérivabilité à gauche en t_0, celle à droite étant parfaitement analogue.

On va devoir étudier le taux de variation $\tau(h) = \dfrac{f(t_0+h) - f(t_0)}{h}$ pour un réel h strictement **négatif** h. Notons $J = \{h \in \mathbf{R}^{-*} \mid t_0 + h \in I\}$, qui est un intervalle comme intersection de deux intervalles, et est non vide parce qu'on a supposé I ouvert. La fonction τ (dont la limite quand $h \to 0^-$ nous intéresse) est définie sur l'intervalle non vide J ; montrons qu'elle est croissante sur cet intervalle. Soit en effet $h_1 < h_2$ dans J ; notons A, B et C les points du graphe de f d'abscisses respectives $h_1 + t_0$, $h_2 + t_0$ et t_0. D'après la définition d'une fonction convexe, B est au-dessous de (AC). En appliquant cette fois la deuxième partie de la proposition clôturant les préliminaires, la pente de (AC) est inférieure ou égale à celle de (BC). Mais la première est égale à $\tau(h_1)$ tandis que la seconde vaut $\tau(h_2)$: la croissance de τ est bien prouvée.

τ est croissante, et c'est une limite à gauche qu'on recherche... on sent bien qu'il ne reste plus qu'à la majorer. Pour ce faire, prenons un k_0 strictement positif assez petit pour que $t_0 + k_0$ soit dans I. Soit maintenant un h de J ; appelons cette fois A, B et C les points du graphe de f d'abscisses respectives $t_0 + h$, t_0 et $t_0 + k_0$. En appliquant le critère de convexité donné juste après la définition, on sait que la pente de (AB) (c'est-à-dire $\tau(h)$) est plus petite que la pente de (BC). On a donc bien trouvé un réel fixe (la pente de (BC)) qui majore tous les $\tau(h)$.

La fonction τ est donc croissante majorée, donc admet une limite en 0^-. Ceci prouve que f est dérivable à gauche en t_0. Pour ceux qui auraient besoin d'une explication, la dérivabilité à gauche entraîne la continuité à gauche et la dérivabilité à droite la continuité à droite. On a donc montré la continuité de f.

Pour montrer l'inégalité annoncée entre f'_g et f'_d, remarquons que la construction qui précédait, avec B d'abscisse t_0, A à gauche de B et C à droite de B a prouvé que pour $h < 0 < k_0$, $\tau(h) \leq \tau(k_0)$ et d'autre part que $f'_g(t_0) = \operatorname*{Sup}_{h<0} \tau(h)$ tandis que, de façon analogue, $f'_d(t_0) = \operatorname*{Inf}_{k>0} \tau(k)$. L'inégalité $\tau(h) \leq \tau(k_0)$ affirme que $\tau(k_0)$ est un majorant de $\{\tau(h) \mid h < 0\}$, donc est plus grand que le plus petit majorant de cet ensemble : en d'autres termes $f'_g(t_0) \leq \tau(k_0)$. Maintenant comme ceci est vrai pour tout k_0, on en déduit ensuite que $f'_g(t_0) \leq \operatorname*{Inf}_{k>0} \tau(k) = f'_d(t_0)$.

Il nous reste enfin à prouver les croissances de f'_g et f'_d. Pour ce faire, prenons deux points $t_1 < t_2$ dans I et introduisons un nouveau réel s strictement compris entre les deux (par exemple $s = \dfrac{t_1 + t_2}{2}$). Notons A d'abscisse t_1, B d'abscisse s et C d'abscisse t_2 sur le graphe de f. Les inégalités qui précèdent entre pentes de demi-tangentes et pentes de sécantes montrent alors que

$$f'_g(t_2) \leq f'_d(t_1) \leq \text{Pente de } (AB) \leq f'_g(s) \leq f'_d(s) \leq \text{Pente de } (BC) \leq f'_g(t_2) \leq f'_d(t_2).$$

Les croissances de f'_g et de f'_d sont toutes deux prouvées.

•

Remarque : Ce théorème est faux pour des raisons stupides si l'intervalle a le droit de ne pas être ouvert : la fonction f définie sur $\mathbf{R}^+$ par $f(t) = 0$ pour $t > 0$ et $f(0) = 1$ est convexe sans être continue... mais n'est pas bien intéressante à regarder.

5 - Fonctions convexes dérivables

Théorème 22-5-44 : Soit f une fonction réelle d'une variable réelle, définie sur un intervalle I. On suppose f dérivable sur I.

Alors f est convexe si et seulement si f' est croissante.

Démonstration :

• Preuve de $\Rightarrow$. Dans le cas où l'intervalle I est **ouvert** il n'y a plus rien à faire, tout était dans le théorème précédent : puisque f'_d est croissante et que f' existe, donc $f' = f'_d$, on obtient la croissance de f', et on a fini.

Il reste quelques lignes de travail si l'intervalle n'est pas ouvert. Supposons que l'intervalle soit de la forme $[a, b]$ (ou $[a, b[$, ou $[a, +\infty[$) et traitons de près ce qui se passe en a. Soit t un point de I distinct de a. Pour tout $h > 0$ tel que $a + h \leq t$, par le théorème de Rolle, il existe un c_h dans l'intervalle $]a, a+h[$ tel que le taux d'accroissement $\tau(h) = \dfrac{f(a+h) - f(a)}{h}$ soit égal à la valeur de la dérivée $f'(c_h)$. On sait déjà que f' est croissante sur l'intervalle ouvert $]a, b[$ (ou $]a, +\infty[$) et donc que $f'(c_h) \leq f'(t)$. En faisant tendre h

vers 0^+ dans cette inégalité, on obtient $f'(a) \leq f'(t)$. On recommence si nécessaire en b, extrémité droite de l'intervalle.

• Preuve de $\Leftarrow$. On va utiliser le critère de convexité numéroté proposition 22-2-105. Soit trois points $A = (x_A, y_A)$, $B = (x_B, y_B)$ et $C = (x_C, y_C)$ sur le graphe de f, s'y succédant dans cet ordre. Par le théorème des accroissements finis, il existe un t_1 et un t_2 avec $x_A < t_1 < x_B < t_2 < x_C$ tels que les pentes de (AB) et de (BC) soient respectivement égales aux dérivées $f'(t_1)$ et $f'(t_2)$. Comme f' est supposée croissante, la première est plus petite que la deuxième : ceci prouve la convexité de f.

•

On préfère souvent insister sur le corollaire qui suit, légèrement plus utile en pratique (un signe est si facile à estimer...) mais qui "explique" moins bien le phénomène que le théorème qui précède.

Corollaire 22-5-2 : Soit f une fonction réelle d'une variable réelle, définie sur un intervalle I. On suppose f deux fois dérivable sur I.

Alors f est convexe si et seulement si f'' est positive.

Démonstration : f' est croissante si et seulement si sa dérivée est positive.

•

Chapitre 23 - Polynômes

Tout le monde connaît ce qu'on appelle souvent "fonctions polynômes" (les fonctions du style : $f(t) = t^3 + 5t$) et que j'appellerai désormais "fonctions polynomiales" ; les polynômes en sont une version plus algébrique –ses avantages sont assez subtils, mais j'ose espérer que vous finirez par les percevoir. Et même si vous ne les percevez pas, vous pourrez vous en servir.

1 - Définitions

Définition 23-1-159 : Soit A un anneau commutatif. On dit qu'un anneau commutatif B est un **anneau de polynômes** sur A lorsque :

i) A est inclus dans B ; (ou plus précisément il existe une application j injective de A dans B telle que pour tous a_1, a_2 de A, $j(a_1 + a_2) = j(a_1) + j(a_2)$ et $j(a_1 a_2) = j(a_1)j(a_2)$ et que $j(1_A) = 1_B$ pour les neutres des multiplications de A et B).

ii) il existe un élément X de B (l'"**indéterminée**" tel que pour tout P de B, (avec $P \neq 0$) il existe un unique $d \geq 0$ et un unique $d+1$-uplet $(a_d, a_{d-1}, \ldots, a_0)$ d'éléments de A tel que $a_d \neq 0$ et que :

$$P = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0.$$

Notation 23-1-63 : Un anneau de polynômes sur A dont l'indéterminée est notée X est noté $A[X]$.

Définition 23-1-160 : Soit A un anneau commutatif et $A[X]$ un anneau de polynômes sur A. Pour P non nul de $A[X]$, l'unique entier $d \geq 0$ intervenant dans l'écriture de P en fonction de l'indéterminée est appelé le **degré** de P. Par convention, le degré du polynôme nul est le symbole $-\infty$.

Notation 23-1-64 : Le degré d'un polynôme P sera noté $\mathrm{d}^\circ P$.

Définition 23-1-161 : Soit A un anneau commutatif et $A[X]$ un anneau de polynômes sur A. Pour P non nul de $A[X]$, le **coefficient dominant** de P sera le coefficient a_d du terme de plus haut degré dans l'écriture de P en fonction de l'indéterminée. Par convention, le coefficient dominant du polynôme nul sera 1.

Définition 23-1-162 : Un polynôme sera dit **unitaire** (ou normalisé, ou monique) lorsque son coefficient dominant est égal à 1.

Proposition 23-1-107 : Soit A un anneau commutatif et $A[X]$ un anneau de polynômes sur A. Soit P, Q deux polynômes de $A[X]$. Alors :

$$\mathrm{d}^\circ(P+Q) \leq \mathrm{Max}(\mathrm{d}^\circ P, \mathrm{d}^\circ Q).$$

Démonstration : Si P ou Q est nul, le résultat est évident. Sinon, notons d le degré de P et e le degré de Q puis $P = a_d X^d + \cdots + a_0$ et $Q = b_e X^e + \cdots + b_0$ pour des a_i et b_i dans A. Si $d > e$, on peut alors écrire : $P + Q = a_d X^d + \cdots + a_{e+1} X^{e+1} + (a_e + b_e) X^e + \cdots + (a_0 + b_0)$ et il apparaît alors que $\mathrm{d}^\circ(P+Q) = d = \mathrm{Max}(\mathrm{d}^\circ P, \mathrm{d}^\circ Q)$. Le cas où $d < e$ est similaire. Enfin, lorsque $d = e$, on a un regroupement : $P + Q = (a_d + b_d) X^d + \cdots + (a_0 + b_0)$; ou bien tous les coefficients y sont nuls, et $\mathrm{d}^\circ(P+Q) = -\infty$ rendant l'inégalité évidente, ou bien un au moins est non nul et le coefficient non nul de plus fort indice est le degré de $P + Q$ qui est bien inférieur ou égal à d.

•

Proposition 23-1-108 : Soit A un anneau commutatif **intègre** et $A[X]$ un anneau de polynômes sur A. Soit P, Q deux polynômes de $A[X]$. Alors :

$$\mathrm{d}^\circ(PQ) = \mathrm{d}^\circ P + \mathrm{d}^\circ Q.$$

Remarque : Pour un anneau non intègre, on a encore une inégalité, mais ce ne me semble pas indispensable à mémoriser... (d'autant que la preuve en est très facile)

Démonstration : Si P ou Q est nul, c'est évident ; sinon notons d le degré de P et e le degré de Q puis $P = a_d X^d + \cdots + a_0$ et $Q = b_e X^e + \cdots + b_0$ pour des a_i et b_i dans A. On a alors $PQ = a_d b_e X^{d+e} + (a_d b_{e-1} + a_{d-1} b_e) X^{d+e-1} + \cdots + a_0 b_0$ (si on n'est pas convaincu par les points de suspension, on écrira plus précisément :

$$PQ = \sum_{k=0}^{d+e} \left(\sum_{i=0}^{k} a_i b_{k-i} \right) X^k$$

en ayant préalablement convenu que $a_i = 0$ pour $i > d$ et $b_i = 0$ pour $i > e$).

Comme l'anneau a été supposé intègre, le produit $a_d b_e$ n'est pas nul, donc le degré de PQ est exactement égal à $d + e$.

•

Définition 23-1-163 : Soit A un anneau commutatif et $A[X]$ un anneau de polynômes sur A. Pour un polynôme $P = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0$ non nul dans $A[X]$, le **polynôme dérivé** de P est le polynôme :

$d a_d X^{d-1} + (d-1) a_{d-1} X^{d-2} + \cdots + a_1$.

Notation 23-1-65 : Le polynôme dérivé de P est noté P'. Par analogie avec les fonctions, on notera ensuite P'' la dérivée de P', puis $P^{(n+1)}$ la dérivée de la dérivée n-ème.

Proposition 23-1-109 : Soit A un anneau commutatif et $A[X]$ un anneau de polynômes sur A. Soit P, Q deux polynômes de $A[X]$. Alors :

$$(P + Q)' = P' + Q' \qquad \text{et} \qquad (PQ)' = P'Q + PQ'.$$

Démonstration : Simple vérification évidente pour l'addition et ennuyeuse pour la multiplication.

•

Définition 23-1-164 : Soit A un anneau commutatif, $A[X]$ un anneau de polynômes sur A, a un élément de A et P un polynôme de $A[X]$. La **valeur de P en** a est l'élément $a_d a^d + a_{d-1} a^{d-1} + \cdots + a_1 a + a_0$ de A, si on note $P = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0$.

Proposition 23-1-110 : Soit A un anneau commutatif, $A[X]$ un anneau de polynômes sur A, a un élément de A, P et Q deux polynômes de $A[X]$. Alors $(P + Q)(a) = P(a) + Q(a)$ et $(PQ)(a) = P(a)Q(a)$.

Démonstration : Simple vérification ; on pourrait aussi énoncer que $1(a) = 1$ qui est évident et complète la collection d'évidences.

•

Cette notation n'a pas que des avantages : elle incite hélas à confondre le polynôme P avec la fonction qu'il n'est pas... Bien que la notation soit la même, cette définition ne se confond pas avec celle de valeur d'une application en un point.

La définition qui suit cherche à reproduire la notion de composition des fonctions (encore une fois, j'insiste sur le fait que les polynômes ne sont pas des fonctions). Elle est utilisée une seule fois plus loin, pour écrire la formule de Taylor relative aux polynômes ; elle sera peut-être davantage utilisée en TD.

Définition 23-1-165 : Soit A un anneau commutatif, $A[X]$ un anneau de polynômes sur A, P et Q deux polynômes de $A[X]$, où on note $P = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0$. On appelle **composé** de P par Q le polynôme $P = a_d Q^d + a_{d-1} Q^{d-1} + \cdots + a_1 Q + a_0$.

Notation 23-1-66 : Ce composé est noté, selon le contexte $P \circ Q$ ou $P(Q)$ (typiquement, pour $Q = X^n$, la notation $P(X^n)$ s'impose et est d'ailleurs d'interprétation évidente).

2 - Les polynômes existent

Je pourrais même écrire, "les polynômes existent et sont uniques" à ceci près que l'énoncé d'unicité est plus précisément un énoncé d'isomorphisme.

Cette preuve est technique et ne mérite d'être lue que par les étudiants les plus acharnés. Pour aider à la rendre plus lisible, je glisse volontairement sur la nuance qu'il peut exister entre "A est inclus dans $A[X]$" et "A s'identifie par une injection à une partie de $A[X]$" –j'espère que le flou créé de ce fait ne gênera pas trop.

Théorème 23-2-45 : Soit A un anneau commutatif.

a) Il existe un anneau $A[X]$ de polynômes sur A.

b) Si $A[X]$ et $A[Y]$ sont deux anneaux de polynômes sur A, il existe un isomorphisme d'anneaux $\varphi: A[X] \to A[Y]$ tel que la restriction de φ à A soit l'identité et que $\varphi(X) = Y$.

Démonstration :

Le point a) demande de l'imagination, le point b) est une ennuyeuse vérification.

L'idée de la preuve sera peut-être compréhensible si on se demande comment stocker un polynôme dans une mémoire de machine : un bon procédé pour stocker le polynôme $X^2+X^3+X^5$ de $\mathbf{Z}/2\mathbf{Z}[X]$ sera de simplement stocker la suite de ses coefficients ; on entrera donc dans la machine la suite 001101 (le coefficient de X^0 est 0 ainsi que celui de X, celui de X^2 est 1, etc...

Ce procédé de stockage sera tout simplement la définition même des polynômes. Simplement, comme nos polynômes mathématiques peuvent être de degré gigantesque, bien plus grand que les capacités de stockage de toute machine, il faudra se résigner à stocker une infinité de coefficients, dont seuls les N premiers sont non nuls (la métaphore technologique s'écroule alors) : ainsi notre polynôme-exemple sera stocké comme $0011010000\ldots$, occupant inutilement une infinité de cases-mémoire.

Notons B l'ensemble des suites (a_n) d'éléments de A A vérifiant la propriété suivante : le nombre d'indices i pour lesquels a_i n'est pas nul est fini. C'est un exercice facile que de montrer que B est un sous-groupe du groupe abélien (additif) de toutes les suites d'éléments de A.

On définit sur B une multiplication par la formule :

$$(a_i)_{i\in\mathbf{N}}\cdot(b_j)_{j\in\mathbf{N}}=(c_k)_{k\in\mathbf{N}} \qquad \text{où} \quad c_k=\sum_{i=0}^{k}a_ib_{k-i}.$$

On a déjà à vérifier que (c_k) est bien une suite de B. Soit en effet M tel que a_i soit nul pour $i>M$ et N tel que b_j soit nul pour $j>N$; alors pour $k>M+N$, dans le calcul de $c_k=\sum_{i=0}^{k}a_ib_{k-i}=\sum_{i=0}^{M}a_ib_{k-i}+\sum_{i=M+1}^{k}a_ib_{k-i}$ tous les termes de la première somme sont nuls, car l'indice $k-i$ y vaut au moins $k-M>N$ donc $b_{k-i}=0$ et tous les termes de la deuxième somme sont nuls car $i>M$ donc $a_i=0$. Tous les coefficients c_k pour $k>M+N$ sont donc nuls et (c_k) est bien dans B.

On va ensuite vérifier que pour ces formules, B est un anneau commutatif. C'est peu engageant et il n'y a guère d'astuces... Il faut calculer brutalement.

* Commutativité ?

Soit $\underline{a}=(a_i)_{i\in\mathbf{N}}$ et $\underline{b}=(b_j)_{j\in\mathbf{N}}$ deux éléments de B ; notons $(c_k)_{k\in\mathbf{N}}=\underline{a}\cdot\underline{b}$. Alors pour tout $k\geq 0$, $c_k=\sum_{i=0}^{k}a_ib_{k-i}=\sum_{j=0}^{k}a_{k-j}b_j$ (en posant $j=k-i$) ; cette expression est clairement celle qu'on trouverait en faisant le produit dans l'autre sens (en utilisant la commutativité de A).

* Associativité ?

Soit $\underline{a}=(a_i)_{i\in\mathbf{N}}$, $\underline{b}=(b_k)_{k\in\mathbf{N}}$ et $\underline{c}=(c_j)_{j\in\mathbf{N}}$ trois éléments de B ; notons $d=(d_j)_{j\in\mathbf{N}}=\underline{b}\cdot\underline{c}=\underline{c}\cdot\underline{b}$ et $e=(e_n)_{n\in\mathbf{N}}=\underline{a}\cdot(\underline{b}\cdot\underline{c})$.

Pour $n\geq 0$, calculons

$$e_n=\sum_{i=0}^{n}a_id_{n-i}=\sum_{i=0}^{n}a_i\sum_{j=0}^{n-i}c_jb_{n-i-j}=\sum_{\substack{(i,j)\in\mathbf{N}^2\\ i+j\leq n}}a_ib_{n-i-j}c_j.$$

On trouverait la même chose en calculant de la même façon $(\underline{a}\cdot\underline{b})\cdot\underline{c}$.

* Existence d'un élément neutre ?

La suite $(1,0,0,0,\ldots)$ est évidemment neutre pour cette multiplication.

* Distributivité ?

Encore une vérification ennuyeuse, celle-là je la saute.

On a bien vérifié que B est un anneau commutatif.

Vérifier que c'est un anneau de polynômes sur A est encore assez lourd mais sans astuces. Construisons d'abord l'injection $j\colon A\to B$ requise par la définition. On posera pour tout $a\in A$, $j(a)=(a,0,0,0,\ldots)$. La vérification que j est un morphisme d'anneaux est très facile et son injectivité est évidente.

Posons maintenant $X=(0,1,0,0,\ldots)$. Une récurrence sans difficulté montre que $X^2=(0,0,1,0,0,\ldots)$ puis $X^3=(0,0,0,1,0,0,\ldots)$ et ainsi de suite...

L'élément $\underline{a}$ de B peut alors s'écrire en fonction de l'indéterminée sous la forme

$$\underline{a} = \sum_{i \in \mathbf{N}} a_i X^i.$$

(On notera que cette somme semble infinie vue de loin, mais ne l'est pas vue de près parce que la famille $\underline{a}$ est dans B).

Reste, dernière corvée, à prouver l'unicité de cette écriture ; c'est encore stupide : si l'élément $(a_i)_{i\in\mathbf{N}}$ de B s'écrit $(a_i) = \sum_{i=0}^{d} b_i X^i$, avec $b_d \neq 0$, on a donc $(a_i) = (b_i)$ (où on note (b_i) la suite dont le terme général est b_i pour $0 \leq i \leq d$ et où on a prolongé cette définition en posant $b_i = 0$ pour $d < i$). Il est alors clair que d est nécessairement le numéro du plus grand indice N tel que $a_N \neq 0$ et que le $d+1$-uplet $(b_0, \ldots, b_d)$ est égal à $(a_0, \ldots, a_d)$. L'écriture était donc la même que celle donnée pour prouver l'existence.

Passons au b). Il est très, très stupide : on construit $\varphi\colon A[X] \to A[Y]$ en définissant

$$\varphi\left(a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0\right) = a_d Y^d + a_{d-1} Y^{d-1} + \cdots + a_1 Y + a_0$$

(c'est juste changer les X en des Y !) et ψ dans l'autre sens en changeant les Y en X. Alors tout est stupide (mais à vérifier formellement...) : φ et ψ sont des bijections car réciproques l'une de l'autre ; φ et ψ sont des morphismes d'anneaux ; la restriction de φ à A est l'identité.

•

Tout cela justifie qu'à partir d'ici je ne parle plus d'"un" anneau de polynômes sur A mais de "l"'anneau de polynômes sur A.

3 - Quelques remarques d'algèbre linéaire

Dans cette section, on suppose que l'anneau commutatif utilisé est un corps commutatif $\mathbf{K}$.

Remarquons tout d'abord que $\mathbf{K}[X]$ est un espace vectoriel sur $\mathbf{K}$. Le plus simple est encore de vérifier à la main la définition des espaces vectoriels, ce que je me garde bien de faire explicitement...

La définition de l'anneau des polynômes devrait vous évoquer le concept de base, avec son existence et unicité d'écriture comme une sorte de combinaison linéaire. On peut bien voir une base là derrière, en connaissant bien les définitions techniques du début du semestre concernant les familles infinies, désormais incontournables pour traiter les polynômes.

Proposition 23-3-111 : Soit $\mathbf{K}$ un corps commutatif. La suite $(X^i)_{i\in\mathbf{N}}$ est une base de $\mathbf{K}[X]$.

Démonstration : C'est quasiment tautologique, il faut s'apercevoir que la définition de l'"anneau des polynômes" et celle de "base" contiennent les mêmes idées.

Soit P un polynôme de $\mathbf{K}[X]$; si P est nul, il est la somme de la famille vide. Sinon, on sait qu'il existe un entier $d \geq 0$ et un $d+1$-uplet $(a_0, \ldots, a_d)$ de scalaires tel que $p = a_0 + \cdots + a_d X^d$. Posons $a_i = 0$ pour $d < i$; on a ainsi écrit P comme combinaison linéaire des X^i, et ainsi montré que la famille (X^i) est génératrice.

La liberté devrait reposer sur l'énoncé d'unicité de la définition des anneaux de polynômes. Il faut montrer que toute sous-famille $(X^i)_{0\leq i\leq n}$ est libre. Assurons nous en ; pour ce faire il suffit de montrer qu'un polynôme appartenant à l'espace engendré par $(X^i)_{0\leq i\leq n}$ possède au plus une écriture relativement à cette famille génératirce. Soit un polynôme P possédant deux écritures $P = \sum_{0\leq i\leq n} a_i X^i = \sum_{0\leq i\leq n} b_i X^i$ pour deux familles de scalaires. Si le polynôme P est non nul, il a un degré uniquement défini, donc le dernier terme non nul doit exister et avoir le même indice dans les deux suites (a_n) et (b_n) (cet indice étant le degré d de P), puis par unicité de l'écriture de P en fonction de l'indéterminée, les $d+1$-uplets $(a_0, \ldots, a_d)$ et $(b_0, \ldots, b_d)$ sont égaux, donc les suites (a_n) et (b_n) sont égales (les autres termes sont tous nuls). Si P est nul, la définition même que j'ai donnée des anneaux de polynômes ne garantit pas l'unicité immédiatement, mais il suffit d'ajouter 1 au terme constant dans les deux écritures pour obtenir deux écritures du polynôme $1 \neq 0$ donc conclure à l'égalité des suites de coefficients.

•

Ainsi $\mathbf{K}[X]$ est votre premier exemple raisonnablement simple d'espace vectoriel ayant une base infinie.

Toutefois, on est toujours plus à l'aise dans les espaces de dimension finie... Il est donc intéressant d'introduire la

Notation 23-3-67 : Soit $\mathbf{K}$ un corps commutatif. On note $\mathbf{K}_n[X]$ l'ensemble des polynômes sur $\mathbf{K}$ de degré inférieur ou égal à n.
Proposition 23-3-112 : $\mathbf{K}_n[X]$ est un sous-espace vectoriel de $\mathbf{K}[X]$. Une base en est $(1, X, \dots, X^n)$ et sa dimension est donc $n+1$.
Démonstration : On remarque que $\mathbf{K}_n[X]$ est l'ensemble engendré par $(1, X, \dots, X^n)$: c'est donc un sous-espace vectoriel. De plus cette famille génératrice est libre, comme sous-famille de la base $(X^i)_{i\in\mathbf{N}}$ de $\mathbf{K}[X]$, c'est donc une base de $\mathbf{K}_n[X]$.
Définition 23-3-166 : La base $(X^i)_{i\in\mathbf{N}}$ de $\mathbf{K}[X]$ est appelée sa **base canonique**. La famille $(1, X, \dots, X^n)$ de $\mathbf{K}_n[X]$ est appelée sa **base canonique**.
Remarque : L'étudiant pourra avoir l'impression qu'on passe son temps à définir de partout des "bases canoniques" : on en a vu pour $\mathbf{K}^n$, puis pour les espaces de matrices, et maintenant pour les polynômes. C'est fini pourtant. J'insiste bien sur le fait qu'un espace "abstrait" n'a **pas** de base canonique : le mot est réservé à certaines bases, remarquables par leur simplicité, d'espaces très particuliers.

Le lemme qui suit servira pour prouver la formule de Taylor. J'ai hésité à en faire un énoncé séparé, mais ai jugé que, même si ce n'est pas indispensable, ce ne peut faire de mal de le connaître ; le plus important étant de comprendre et savoir refaire sa brève démonstration.
Lemme 23-3-11 : Soit $\mathbf{K}$ un corps commutatif et $(P_0, P_1, \dots, P_n)$ un système de polynômes de $\mathbf{K}[X]$ tel que $0 \leq \mathrm{d}^\circ P_0 < \mathrm{d}^\circ P_1 < \cdots < \mathrm{d}^\circ P_n$. Alors $(P_0, P_1, \dots, P_n)$ est un système libre.
Démonstration : Le système (P_0) est libre, car il résulte de l'hypothèse $0 \leq \mathrm{d}^\circ P_0$ que P_0 n'est pas nul. Puis le système (P_0, P_1) est libre puisque P_1, de degré strictement plus grand que P_0, ne peut lui être proportionnel. Puis (P_0, P_1, P_2) est libre, puisque toute combinaison linéaire de (P_0, P_1) est de degré inférieur ou égal à $\mathrm{d}^\circ P_1$ donc P_2 ne peut en être une. Et ainsi de suite (ou plus proprement on fait une récurrence sur n). •

4 - Arithmétique des polynômes

Il s'agit de répéter pour les polynômes des résultats tout à fait analogues à ceux qui ont été énoncés pour les entiers.

Premier point à observer : l'arithmétique sur les polynômes est tout à fait analogue à celle sur les entiers à condition de travailler sur des polynômes sur un corps commutatif. Sur un anneau commutatif quelconque (même intègre) se glissent quelques bizarreries.

Second point à observer : les énoncés donnés sur les entiers l'ont été sur des entiers positifs. Ils se modifient sans trop de mal pour des entiers de $\mathbf{Z}$ mais parfois en s'alourdissant un peu ; ainsi dans $\mathbf{Z}$ je ne peux plus dire qu' "il existe un D unique tel que d divise 10 et 6 si et seulement si d divise D" : il en existe toujours un, mais il n'est plus unique – je peux prendre $D = 2$ mais aussi $D = -2$. Les polynômes unitaires joueront un rôle analogue aux entiers positifs mais ils sont légèrement moins confortables, dans la mesure où la somme de deux entiers positifs est positive alors que la somme de deux polynômes unitaires n'est pas nécessairement unitaire. Attention à ces petits détails donc en apprenant les énoncés...

Commençons par donner une définition, à partir de laquelle on ne montrera guère de théorèmes que dans $\mathbf{K}[X]$ mais que ça ne coûte pas plus cher de donner sur un anneau commutatif quelconque.
Définition 23-4-167 : Soit A un anneau commutatif. On dit qu'un polynôme P_m $(\in A[X])$ est **multiple** d'un polynôme P_d $(\in A[X])$, ou que P_d est un **diviseur** de P_m lorsqu'il existe un polynôme Q $(\in A[X])$ tel que $P_m = P_d Q$.
La division euclidienne

Comme pour les entiers, tout repose sur la division euclidienne :
Théorème 23-4-46 : Soit $\mathbf{K}$ un corps commutatif. Soit A un polynôme de $\mathbf{K}[X]$ et B un polynôme non nul de $\mathbf{K}[X]$.

Alors il existe un couple (Q, R) unique (de polynômes) vérifiant la double condition :

$$A = BQ + R \qquad \text{et} \qquad \mathrm{d}^\circ R < \mathrm{d}^\circ B.$$

Démonstration : On prouvera successivement l'existence et l'unicité de (Q, R).
* Existence de (Q, R).

La preuve est significativement différente de celle utilisée pour les entiers. Elle est toujours basée sur une maximisation/minimisation, mais les polynômes n'étant pas totalement ordonnés, cette maximisation est un peu plus technique.
Dans le cas stupide où B divise A, prenons $R = 0$ et Q tel que $A = BQ$. Sinon, considérons l'ensemble $\mathcal{R} = \{A - BQ \mid Q \in \mathbf{K}[X]\}$, qui est donc un ensemble non vide de polynômes non nuls ; puis l'ensemble $E = \{d^\circ R \mid R \in \mathcal{R}\}$ qui est donc un ensemble d'entiers positifs non vide. Cet ensemble E possède donc un plus petit élément d ; prenons un R dans $\mathcal{R}$ dont le degré soit d et enfin un Q tel que $A - BQ = R$.
Nous devons vérifier que ces choix conviennent ; l'identité entre A, B, Q et R est claire, reste l'inégalité concernant les degrés. Vérifions la par l'absurde, en supposant $d^\circ B \leq d^\circ R$; notons e le degré de B et $B = b_e X^e + b_{e-1}X^{e-1} + \cdots + b_0$ et notons $R = r_d X^d + r_{d-1}X^{d-1} + \cdots + r_0$. Posons $Q_1 = Q + \dfrac{r_d}{b_e}X^{d-e}$ (en écrivant cette définition, on utilise d'un seul coup l'hypothèse $d^\circ B \leq d^\circ R$, qui justifie que X^{d-e} ait un sens, et le fait qu'on travaille dans un corps, qui justifie la possibilité de diviser par b_e).
Considérons alors $R_1 = A - BQ_1 = A - BQ - B\left(\dfrac{r_d}{b_e}X^{d-e}\right) = R - \left(b_e X^e + b_{e-1}X^{e-1} + \cdots + b_0\right)\left(\dfrac{r_d}{b_e}X^{d-e}\right)$
Dans cette dernière écriture, on voit se simplifier les termes en X^d de R et du produit qu'on lui a soustrait, et on constate donc avoir obtenu un polynôme R_1 de degré strictement plus petit que celui de R. Mais alors le degré de R_1 est dans E et contredit l'hypothèse de minimisation qui a fait choisir d. Contradiction !
* Unicité de (Q, R) : soit (Q_1, R_1) et (Q_2, R_2) deux couples vérifiant tous deux les deux conditions exigées dans l'énoncé du théorème.
On déduit de $A = BQ_1 + R_1 = BQ_2 + R_2$ que $B(Q_2 - Q_1) = R_1 - R_2$. Ainsi, $R_1 - R_2$ est un multiple de B. Des conditions $d^\circ R_1 < d^\circ B$ et $d^\circ R_2 < d^\circ B$, on déduit que $d^\circ(R_1 - R_2) < d^\circ B$.
Ainsi $R_1 - R_2$ est un multiple de B de degré strictement plus petit. La seule possibilité est que $R_1 - R_2$ soit nul. On en déduit $R_1 = R_2$, puis, en allant reprendre l'égalité $B(Q_2 - Q_1) = R_1 - R_2$, que $Q_1 = Q_2$. •

Remarque : J'ai choisi d'énoncer ce théorème sur un corps commutatif pour faciliter sa mémorisation et parce que je n'aurai presque jamais besoin d'un énoncé plus général. J'aurai toutefois besoin une fois de l'utiliser pour des polynômes sur un anneau ; remarquons donc que la démonstration montre que le résultat reste vrai sur un anneau commutatif quelconque à condition de supposer non seulement que B est non nul, mais même que son coefficient dominant est inversible : le seul endroit où on a utilisé qu'on s'était placé dans un corps commutatif a en effet été une division par ce coefficient dominant.

Le PGCD

Je ne donnerai pas ici d'énoncés concernant le PPCM, non qu'il n'y en n'ait pas (ce sont là aussi les mêmes qu'en arithmétique des entiers) mais parce qu'ils ne me semblent pas très importants. Les étudiants curieux les reconstitueront eux-mêmes.

Théorème 23-4-47 : Soit $\mathbf{K}$ un corps commutatif. Soit A et B deux polynômes de $\mathbf{K}[X]$. Alors il existe un unique polynôme **unitaire** $D \in \mathbf{K}[X]$ tel que pour tout $P_d \in \mathbf{K}[X]$

$$P_d \text{ divise } A \text{ et } B \iff P_d \text{ divise } D.$$

De plus il existe deux polynômes S et T tels que $D = SA + TB$. (identité de Bezout)

Et tant qu'on y est avant de passer aux démonstrations :

Définition 23-4-168 : Le **plus grand commun diviseur** de deux polynômes A et B est le polynôme unitaire D apparaissant dans l'énoncé du théorème précédent.

Notation 23-4-68 : Le plus grand commun diviseur de A et B sera noté $\mathrm{PGCD}(A, B)$.

Comme pour les entiers, plusieurs démonstrations sont possibles ; ici je ne donne que celle basée sur l'algorithme d'Euclide.

Démonstration du théorème :
La démonstration est une récurrence sur le degré de B.
Merveilles du copier-coller, voici de nouveau un "résumé de la preuve" sous forme de programme informatique récursif (le même que pour l'arithmétique des entiers) :

* Pour $B = 0$, $\mathrm{PGCD}(A, 0) = A/$coefficient dominant de A.
* En notant R le reste de la division euclidienne de A par B, les diviseurs communs de A et B sont les diviseurs communs de B et R, d'où :

PGCD(A, B) = PGCD(B, R).

Et voici, toujours par les vertus du copier-coller, la preuve récurrente formelle.

On va démontrer par "récurrence forte" sur le degré d de B l'hypothèse (H_d) suivante :

Pour tout polynôme A et tout polynôme B de degré d, il existe deux polynômes S et T tels que, pour tout polynôme P_d, P_d divise A et $B \iff P_d$ divise $SA + TB$.

* Vérifions $(H_{-\infty})$.

Soit A un polynôme ; tout polynôme P_d qui divise A divise aussi $B = 0$ puisque $0P_d = 0$. Pour tout P_d, on a donc : P_d divise A et $0 \iff P_d$ divise A. Prenons alors $S = 1$ et $T = 0$: on a donc bien pour tout P_d : P_d divise A et $0 \iff P_d$ divise $SA + T \times 0$.

* Soit d un entier fixé. Supposons la propriété (H_c) vraie pour tout c strictement inférieur à d et montrons (H_d).

Soit A un polynôme et B un polynôme de degré d. Notons $A = BQ + R$ la division euclidienne de A par B (qu'on peut réaliser puisque $B \neq 0$).

Vérifions l'affirmation intermédiaire suivante : pour tout P_d, P_d est un diviseur commun de A et $B \iff P_d$ est un diviseur commun de B et R. (Avec des mots peut-être plus lisibles : "les diviseurs communs de A et B sont les mêmes que ceux de B et R").

Soit P_d un diviseur commun de A et B, alors P_d divise aussi $R = A - BQ$; réciproquement soit P_d un diviseur commun de B et R, alors P_d divise aussi $A = BQ + R$.

L'affirmation intermédiaire est donc démontrée.

On peut alors appliquer l'hypothèse de récurrence $(H_{\mathrm{d}^\circ R})$ (puisque précisément $\mathrm{d}^\circ R < \mathrm{d}^\circ B$) en l'appliquant au polynôme B.

On en déduit qu'il existe deux polynômes S_1 et T_1 tels que pour tout P_d, P_d divise B et $R \iff P_d$ divise $S_1 B + T_1 R$.

Remarquons enfin que $S_1 B + T_1 R = S_1 B + T_1(A - BQ) = T_A A + (S_1 - Q)B$, et qu'ainsi, si on pose $S = T_A$ et $T = S_1 - Q$ on a bien prouvé que, pour tout P_d, P_d divise Q et $B \iff P_d$ divise $SA + TB$.

(H_d) est donc démontrée.

* On a donc bien prouvé (H_d) pour tout $d \in \mathbf{N} \cup \{-\infty\}$.

* Une fois qu'on en est arrivé là, il ne reste donc plus qu'à montrer que pour **un** polynôme P (le polynôme $SA + TB$) il existe un unique D unitaire tel que P_d divise P si et seulement si P_d divise D. L'existence est claire : comme le résumé le suggère, on divise P par son coefficient dominant et on obtient un polynôme D unitaire ayant les mêmes diviseurs que P. Pour ce qui est de l'unicité, elle est évidente pour P nul ; on supposera P non nul. Soit maintenant D_1 un polynôme unitaire ayant exactement les mêmes diviseurs que P. Alors comme P divise P, P divise D_1, et comme D_1 divise D_1, D_1 divise P. Les polynômes P et D_1 se divisent donc mutuellement ; soit Q_1 et Q_2 les quotients respectifs de P par D_1 et de D_1 par P. En utilisant la formule calculant le degré d'un produit, on voit que forcément, P a même degré que D_1 et que les polynômes Q_1 et Q_2 sont de degré nul, donc des constantes λ_1 et λ_2. Soit a_d le coefficient dominant de P ; le coefficient dominant de $Q_1 D_1 = P$ vaut $\lambda_1 \cdot 1$ donc $\lambda_1 = a_d$ et D_1 est égal à $P/$ (coefficient dominant de P), donc à D, ce qui prouve l'unicité.

•

PGCD d'un nombre fini de polynômes

Je n'avais pas mis en relief cette notion en arithmétique des entiers, où elle n'était pas primordiale ; en revanche dans les applications des raisonnements arithmétiques à des polynômes, on est souvent dans des cas où on s'intéresse à des PGCDs de plus de deux polynômes à la fois.

L'énoncé donné ci-dessus pour deux polynômes se généralise à un nombre fini, par récurrence sur ce nombre.

Proposition 23-4-113 : Soit $\mathbf{K}$ un corps commutatif. Soit $A_1, A_2, \ldots, A_n$ un nombre fini de polynômes de $\mathbf{K}[X]$. Alors il existe un unique polynôme unitaire $D \in \mathbf{K}[X]$ tel que pour tout $P_d \in \mathbf{K}[X]$

$$P_d \text{ divise } A_1, A_2, \ldots, A_n \iff P_d \text{ divise } D.$$

De plus il existe n polynômes $S_1, S_2, \ldots, S_n$ tels que $D = S_1 A_1 + S_2 A_2 + \cdots + S_n A_n$ (identité de Bezout).

Démonstration : C'est une récurrence facile sur n. Le cas $n = 2$ est l'objet du théorème précédent (et le cas $n = 1$ a été traité dans sa démonstration, ou on peut le ramener fictivement à $n = 2$ en disant que les diviseurs de A_1 sont les diviseurs communs de A_1 et de 0).

Soit $n \geq 2$ fixé, supposons la proposition vraie pour tout ensemble de n polynômes, et soit des polynômes $A_1, A_2, \ldots, A_{n+1}$. Notons Δ le PGCD des n premiers, qui existe par l'hypothèse de récurrence. Alors les diviseurs communs de $A_1, A_2, \ldots, A_{n+1}$ sont les diviseurs communs de Δ et de A_{n+1} ; donc prendre $D = \mathrm{PGCD}(\Delta, A_{n+1})$ répond à la question. L'unicité est claire : si D_1 répondait aussi à la question, les diviseurs de D_1 seraient exactement les mêmes que ceux de D avec D et D_1 tous deux unitaires, et comme dans la preuve du théorème précédent (ou en appliquant le théorème précédent à D et 0...) on conclut que $D = D_1$. La relation de Bezout est aussi le résultat d'une récurrence immédiate : il existe $S_1, S_2, \ldots, S_n$ tels que $\Delta = S_1A_1 + S_2A_2 + \cdots + S_nA_n$ et T_1 et T_2 tels que $D = T_1\Delta + T_2A_{n+1}$ donc $D = (T_1S_1)A_1 + (T_1S_2)A_2 + \cdots + (T_1S_n)A_n + T_2A_{n+1}$.

•

Définition 23-4-169 : Soit $\mathbf{K}$ un corps commutatif. On dira que n polynômes de $\mathbf{K}[X]$ sont **premiers entre eux** lorsque leurs seuls diviseurs communs sont constants (en d'autres termes, quand leur PGCD est 1).

On prendra garde à ne pas confondre "premiers entre eux" et "deux à deux premiers entre eux" : dans $\mathbf{R}[X]$, les polyômes $(X-1)(X-2)$, $(X-1)(X-3)$ et $(X-2)(X-3)$ sont premiers entre eux, mais pas deux à deux premiers entre eux.

Polynômes irréductibles

Les polynômes irréductibles sont les analogues des nombres premiers. Toutefois les usages étant ce qu'ils sont, il y a une petite nuance de vocabulaire un peu désagréable : alors que le mot "nombre premier" est réservé à des entiers positifs, le mot "polynôme irréductible" n'est pas réservé à des polynômes unitaires. On se méfiera de cette peu perceptible nuance qui crée de légères discordances entre énoncés analogues portant les uns sur les polynômes et les autres sur les entiers.

Définition 23-4-170 : Soit $\mathbf{K}$ un corps commutatif. On dira qu'un polynôme $P \in \mathbf{K}[X]$ est **irréductible** lorsqu'il possède exactement deux diviseurs unitaires.

On remarquera tout de suite que ces deux diviseurs unitaires sont alors forcément les polynômes 1 et $P/$(coefficient dominant de P).

La proposition suivante est évidente, mais donne un exemple fondamental de polynômes irréductibles :

Proposition 23-4-114 : Soit $\mathbf{K}$ un corps commutatif. Dans $\mathbf{K}[X]$ les polynômes du premier degré sont irréductibles.

Démonstration : Soit $P = aX + b$ avec $a \neq 0$ un polynôme du premier degré dans $\mathbf{K}[X]$. Cherchons ses diviseurs unitaires. Un diviseur de P doit avoir un degré inférieur ou égal à celui de P. Le seul diviseur unitaire constant de P est le seul polynôme constant unitaire : la constante 1. Cherchons les diviseurs unitaires de la forme $X + c$ de P. Si $X + c$ divise P, il existe un polynôme Q tel que $P = (X + c)Q$ et en comparant les degrés, Q est nécessairement constant. En comparant les coefficients dominants, nécessairement $Q = a$ donc $c = \frac{b}{a}$. Ainsi P possède exactement un diviseur unitaire du premier degré, le polynôme $X + \frac{b}{a}$. Le polynôme P est donc irréductible.

•

Sur un corps quelconque, déterminer quels polynômes sont irréductibles et lesquels ne le sont pas est un problème très sérieux ; dans quelques pages, nous verrons que ce problème a une solution simple dans les cas particuliers des polynômes à coefficients complexes ou réels.

Le résultat fondamental est, comme en arithmétique entière, l'existence et unicité de la décomposition en facteurs irréductibles. Elle repose là encore sur le "lemme de Gauss". Je ne réécris pas les démonstrations pour deux raisons totalement contradictoires : d'abord parce que ce sont exactement les mêmes, et ensuite parce que ce ne sont pas exactement les mêmes –une petite difficulté se pose pour énoncer l'unicité de la décomposition en facteurs irréductibles d'un polynôme. Pour des entiers, j'ai convenu de classer les facteurs dans l'ordre croissant : ainsi 6 se décompose en $2 \cdot 3$ et non en $3 \cdot 2$. Une telle convention ne peut être appliquée pour décomposer des polynômes, aucun ordre "raisonnable" n'étant à notre disposition sur l'ensemble des polynômes irréductibles ; ainsi dans $\mathbf{C}[X]$ peut-on écrire selon la fantaisie du moment $X^2 + 1 = (X - i)(X + i)$ ou $X^2 + 1 = (X + i)(X - i)$. Quand j'énonce ci-dessous que la décomposition est "unique" je sous-entends que je considère les deux exemples qui précèdent comme la même décomposition, ce qui peut s'énoncer rigoureusement mais lourdement. Voulant glisser sur ce détail, je me condamne à rester un peu vaseux.

Voici donc le lemme de Gauss :

Lemme 23-4-12 : Soit $\mathbf{K}$ un corps commutatif. Soit A, B, C trois polynômes de $\mathbf{K}[X]$. Si A divise BC et est premier avec C, alors A divise B.

Démonstration : La même que pour les entiers, avec des majuscules. •

et voici le théorème de décomposition en facteurs irréductibles.
Théorème 23-4-48 : (énoncé moyennement précis) Soit $\mathbf{K}$ un corps commutatif. Tout polynôme P non nul de $\mathbf{K}[X]$ peut s'écrire de façon "unique" en produit :

$$P = \lambda P_1^{\alpha_1} P_2^{\alpha_2} \cdots P_k^{\alpha_k}$$

dans lequel λ est le coefficient dominant de P, les P_i $(0 \leq i \leq k)$ sont des polynômes irréductibles unitaires deux à deux distincts, et les α_i sont des entiers strictement positifs.
Démonstration : À peu près la même que pour les entiers, avec un peu plus de soin pour l'unicité... •

5 - Racines des polynômes

Définition 23-5-171 : Soit A un anneau commutatif, P un polynôme de $A[X]$ et a un élément de A. On dit que a est une **racine** (ou un **zéro**) de P lorsque $P(a) = 0$.

Le résultat qui suit est fondamental, bien que très facile :
Proposition 23-5-115 : Soit A un anneau commmutatif, P un polynôme de $A[X]$ et a un élément de A. L'élément a est une racine de P si et seulement si $X - a$ divise P.
Démonstration :

Supposons que $X - a$ divise P, soit $P = (X - a)Q$. On obtient aussitôt $P(a) = (a - a)Q(a) = 0$.

Réciproquement, supposons que $P(a) = 0$. La remarque qui suit l'énoncé du théorème de division euclidienne montre que, même dans un anneau quelconque, on peut faire la division euclidienne de P par $X - a$; écrivons donc $P = Q(X - a) + R$, où le degré de R est strictement inférieur à $1 = \mathrm{d}^\circ(X - a)$ donc R est une constante c.

En appliquant cette relation à a, on obtient $0 = P(a) = c$. Ainsi, $P = (X - a)Q$ et donc $X - a$ divise P. •

Corollaire 23-5-3 : Soit A un anneau commutatif intègre. Un polynôme non nul de degré n possède au plus n racines.
Démonstration : Par récurrence sur n. Pour $n = 0$, un polynôme constant non nul possède évidemment zéro racine.

Soit n fixé, supposons le résultat vrai pour les polynômes de degré n ; soit maintenant P un polynôme de degré $n + 1$. Si P n'a aucune racine, le résultat est vrai pour P ; sinon soit a une racine de P ; par la proposition précédente on peut écrire $P = (X - a)Q$ pour un polynôme Q, qui est clairement de degré n. Maitenant, si b est une racine de P, $0 = P(b) = (b - a)Q(b)$ donc $b = a$ ou b est une racine de Q (c'est ici qu'on utilise l'hypothèse d'intégrité) ; or Q a au plus n racines, donc P en a au plus $n + 1$. •

On va ensuite définir un concept de "racine multiple".
Définition 23-5-172 : Soit A un anneau commutatif, P un polynôme de $A[X]$ et a un élément de A. On dit que a est racine **au moins k-ème** de P lorsque $(X - a)^k$ divise P ; qu'il est **racine k-ème** lorsqu'il est racine au moins k-ème sans être racine $k + 1$-ème. On dit alors que k est la **multiplicité** (ou l'**ordre**) de a comme racine de P.

La dérivation des polynômes est un outil qui permet d'étudier les racines multiples. Voilà tout d'abord un énoncé concernant les racines doubles (l'énoncé concernant les racines d'ordre supérieur cache une petite subtilité et est reporté plus loin).
Proposition 23-5-116 : Soit A un anneau commutatif, P un polynôme de $A[X]$ et a un élément de A. L'élément a est racine au moins double de P si et seulement s'il est simultanément racine de P et de son dérivé P'.
Démonstration : Supposons a racine au moins double de P et posons $P = (X - a)^2 Q$ On a alors $P' = 2(X - a)Q + (X - a)^2 Q'$ et il est clair que a est également racine de P'.

Réciproquement, supposons a racine de P et de P'. Comme a est racine de P, on peut écrire $P = (X - a)Q_1$, donc $P' = (X - a)Q_1' + Q_1$. En appliquant cette identité à a, on obtient $Q_1(a) = 0$. Donc Q_1 admet lui-même $X - a$ en facteur et peut s'écrire $Q_1 = (X - a)Q$ pour un polynôme Q. Donc $P = (X - a)^2 Q$. •

6 - Polynômes versus fonctions polynomiales

J'ai commencé par insister pour que vous ne confondiez pas les polynômes et les fonctions polynomiales ; il est temps de voir le rapport entre ces deux concepts.

Définition 23-6-173 : Soit A un anneau commutatif. Une application $f: A \to A$ est dite **polynomiale** lorsqu'il existe un entier $n \geq -1$ et un $n+1$-uplet $(a_0, \dots, a_n)$ d'éléments de A tel que pour tout $x \in A$, $f(x) = a_0 + a_1 x + \cdots + a_n x^n$.

On peut associer à chaque polynôme une fonction polynomiale de façon stupide, mais il n'est pas du tout évident d'associer un polynôme à une fonction polynomiale.

Définition 23-6-174 : Soit A un anneau commutatif et P un polynôme de $A[X]$. La **fonction polynomiale associée à** P est l'application $f: A \to A$ définie de la façon suivante : si P s'écrit $a_0 + a_1 X + \cdots + a_n X^n$, f est l'application définie par $f(x) = a_0 + a_1 x + \cdots + a_n x^n$.

Les morceaux "évidents" de la proposition suivante resteraient vrais sur des anneaux, mais je l'énonce sur des corps pour pouvoir prononcer des termes d'algèbre linéaire.

Proposition 23-6-117 : Soit $\mathbf{K}$ un corps commutatif et soit $U: \mathbf{K}[X] \to \mathbf{K}^{\mathbf{K}}$ l'application définie par :

$U(P)$ est la fonction polynomiale associée à P.

Alors U est une application linéaire, et vérifie en outre $U(PQ) = U(P)U(Q)$ pour tous P, Q et $U(1) = 1$ (le deuxième 1 désignant la fonction constante prenant la valeur 1).

L'image de U est le sous-espace vectoriel de $\mathbf{K}^{\mathbf{K}}$ formé des fonctions polynomiales.

Si $\mathbf{K}$ est **infini**, l'application U est injective, donc induit une bijection entre l'espace des polynômes et celui des fonctions polynomiales.

Démonstration : Les deux premiers paragraphes sont totalement évidents : il faut juste déplier successivement la définition de U, celle de fonction polynomiale associée à un polynôme et celle de valeur d'un polynôme en un point.

Le paragraphe intéressant est le dernier. Puisqu'il s'agit d'une application linéaire, on peut attaquer l'injectivité par l'étude du noyau. Soit $P \in \operatorname{Ker} U$. Cela signifie que l'application polynomiale associée à P est la fonction nulle, c'est-à-dire que pour tout a de A, $P(a) = 0$. Ainsi tous les éléments de $\mathbf{K}$ sont des racines de P. Comme on a supposé $\mathbf{K}$ infini, ceci entraîne que P a une infinité de racines. Mais on sait qu'un polynôme non nul n'a qu'un nombre fini de racines (leur nombre vaut au plus son degré). Donc $P = 0$ ce qui prouve la trivialité de $\operatorname{Ker} U$ donc l'injectivité de U.

•

Remarque : Ce que dit en gros cette proposition, pour ceux qui la trouveraient trop abstraites, c'est que si on ne comprend pas la différence entre les polynômes et les fonctions polynomiales et qu'on travaille sur un corps infini, on ne s'expose pas à des déboires sérieux. Mais ce laxisme apparent de ma part ne doit pas vous inviter à vous laisser aller : une telle confusion sur un corps fini serait irrémédiable. Pour voir un exemple simple, contemplez le bête polynôme $X + X^2$ de $\mathbf{Z}/2\mathbf{Z}[X]$; si on le code en machine comme indiqué plus haut, c'est la suite de bits 011 ce n'est manifestement pas 0. Pourtant si on regarde non le polynôme mais la fonction polynomiale $x \mapsto x + x^2$, sa valeur en $\dot{0}$ est $\dot{0} + (\dot{0})^2 = \dot{0}$ et sa valeur en $\dot{1}$ est $\dot{1} + (\dot{1})^2 = \dot{0}$ –c'est bien la fonction polynomiale nulle. Ce n'est donc définitivement pas de celle-ci que l'on parle quand on évoque le polynôme $X + X^2$.

7 - La formule de Taylor pour les polynômes

Alors que pour des fonctions d'une variable réelle, la formule de Taylor ne peut tomber juste, puisqu'elle consiste à approcher la fonction par une fonction polynomiale et que la fonction quelconque n'est précisément en général pas polynomiale, pour des polynômes, la formule analogue ne contient pas de reste.

Une petite subtilité apparaît dans les divisions par des factorielles qui enjolivent la formule. En effet dans un anneau commutatif quelconque, mais même dans un corps commutatif, on ne peut pas toujours diviser par une factorielle : dans le corps $\mathbf{Z}/3\mathbf{Z}$, la factorielle 3! qui vaut 6 vaut tout simplement 0 puisque 6 est divisible par 3. C'est pourquoi ce théorème nécessite une restriction technique ; j'ai choisi de l'énoncer pour des polynômes à coefficients complexes ; si vous utilisez ce cours comme référence et le relisez dans quelques années, notez que la "bonne" hypothèse est plutôt d'être en caractéristique nulle.

Théorème 23-7-49 : Soit P un polynôme de $\mathbf{C}[X]$ de degré inférieur ou égal à n, et a un élément de $\mathbf{C}$. Alors

$$P = P(a) + P'(a)(X-a) + \frac{P''(a)}{2!}(X-a)^2 + \cdots + \frac{P^{(n)}(a)}{n!}(X-a)^n.$$

Démonstration : On va travailler dans l'espace vectoriel $\mathbf{C}_n[X]$ et considérer dans cet espace le système $(1, X-a, (X-a)^2, \ldots, (X-a)^n)$. Ces polynômes sont de degrés successifs $0 < 1 < \cdots < n$ donc on peut appliquer le lemme fait exprès pour de la section d'observations d'algèbre linéaire et conclure que c'est un système libre dans $\mathbf{C}_n[X]$. Elle possède $n+1$ vecteurs dans cet espace de dimension $n+1$, donc en est une base, et en particulier un système générateur.

Il existe donc des coefficients $c_0, c_1, \ldots, c_n$ tels que

$$(*) \qquad P = c_0 + c_1(X-a) + c_2(X-a)^2 + \cdots + c_n(X-a)^n.$$

Appliquons $(*)$ au point a : on obtient $P(a) = c_0$.
Dérivons $(*)$; on obtient :

$$(*') \qquad P' = c_1 + 2c_2(X-a) + 3c_3(X-a)^2 \cdots + nc_n(X-a)^{n-1}.$$

Appliquons $(*')$ au point a : on obtient $P'(a) = c_1$.
Dérivons $(*')$; on obtient :

$$(*'') \qquad P'' = c_2 + 6c_3(X-a) + (4\times 3)c_3(X-a)^2 \cdots + n(n-1)c_n(X-a)^{n-2}.$$

Appliquons $(*'')$ au point a : on obtient $P''(a) = 2c_2$.
En écrivant formellement une récurrence on montre ainsi que pour tout k avec $1 \leq k \leq n$, $P^{(k)}(a) = k!c_k$.

Comme on est dans $\mathbf{C}$, on peut diviser par $k!$ et obtenir les relations $c_k = \dfrac{P^{(k)}(a)}{k!}$ donc la formule annoncée.

•

Remarque : J'ai énoncé ce théorème pour des polynômes à coefficients complexes. Mais si j'ai par exemple affaire à un polynôme réel, c'est en particulier un polynôme complexe et la formule est donc parfaitement vraie pour ce polynôme aussi.

De cette formule, on peut tirer un énoncé un peu technique sur les racines multiples, qui n'est vrai qu'en caractéristique nulle.

Proposition 23-7-118 : Soit P un polynôme de $\mathbf{C}[X]$, a un nombre complexe et k un entier supérieur ou égal à 1. Alors a est une racine au moins $k+1$-ème de P si et seulement si $P(a) = P'(a) = \ldots = P^{(k)}(a) = 0$.

Démonstration : Si P est nul, c'est évident, sinon notons n le degré de P et λ son coefficient dominant.

Considérons les indices $i \geq 0$ tels que $P^{(i)}(a) \neq 0$, en convenant que $P^{(0)} = P$. Il existe de tels indices, car le polynôme $P^{(n)}$ est égal à la constante $n!\lambda$, donc n'est pas nul en a. Cet ensemble non vide d'entiers positifs a donc un plus petit élément m, qui vérifie $0 \leq m \leq n$. Écrivons la formule de Taylor en mettant en relief cet entier m :

$$P = \frac{P^{(m)}(a)}{m!}(X-a)^m + \frac{P^{(m+1)}(a)}{(m+1)!}(X-a)^{m+1} + \cdots + \frac{P^{(n)}(a)}{n!}(X-a)^n.$$

On constate qu'on peut mettre $(X-a)^m$ en facteur, mais que le facteur obtenu , qui est le polynôme $Q = \dfrac{P^{(m)}(a)}{m!} + \dfrac{P^{(m+1)}(a)}{(m+1)!}(X-a) + \cdots + \dfrac{P^{(n)}(a)}{n!}(X-a)^{n-m}$ ne s'annule pas en a : la multiplicité de a comme racine de P est donc exactement m.

Dès lors, P est racine au moins $k+1$-ème de P si et seulement si $k < m$, et par définition de m ceci arrive bien si et seulement si $P(a) = P'(a) = \ldots = P^{(k)}(a) = 0$.

•

8 - Les spécificités de $\mathbf{C}[X]$ et de $\mathbf{R}[X]$

Toute cette section repose sur un théorème qu'il n'est pas possible de démontrer en DEUG :

Théorème 23-8-50 : (de d'Alembert-Gauss) Tout polynôme non constant de $\mathbf{C}[X]$ admet au moins une racine complexe.

Démonstration : Elle repose sur un peu d'analyse, mais d'analyse complexe, qui n'est traitée qu'en licence.
•

Corollaire 23-8-4 : Dans $\mathbf{C}[X]$ les polynômes irréductibles sont exactement les polynômes du premier degré.
Démonstration : On sait déjà que dans n'importe quel corps commutatif les polynômes du premier degré sont irréductibles ; il est très facile de voir que les constantes (non nulles) ne possèdent que 1 comme diviseur unitaire et que 0 en possède une infinité : les constantes ne sont donc irréductibles sur aucun corps.

Soit maintenant un P de degré supérieur ou égal à 2 dans $\mathbf{C}[X]$. Par le théorème précédent, P possède au moins une racine a. Mais on sait alors expliciter trois diviseurs unitaires de P : la constante 1, le polynôme du premier degré $X-a$ et le polynôme $P/$ (coefficient dominant de P), qui est de degré supérieur ou égal à deux. Ainsi P n'est pas irréductible.
•

Définition 23-8-175 : On dit qu'un polynôme est **scindé** lorsqu'il peut s'écrire sous forme de produit de facteurs du premier degré.
Corollaire 23-8-5 : Dans $\mathbf{C}[X]$ tout polynôme non nul est scindé.
Démonstration : Sa décomposition en facteurs irréductibles est une décomposition en produit de facteurs du premier degré.
•

Dans $\mathbf{R}[X]$, les choses sont légèrement plus compliquées, mais pas tant que ça.
Proposition 23-8-119 : Dans $\mathbf{R}[X]$ les polynômes irréductibles sont exactement les polynômes du premier degré et les polynômes du deuxième degré à discriminant strictement négatif.
Démonstration : On sait déjà que les polynômes du premier degré sont irréductibles. Soit maintenant P du deuxième degré ; s'il a un diviseur unitaire autre que les deux évidents, celui-ci est du premier degré, donc P a une racine et son discriminant est positif ou nul. Les polynômes du deuxième degré à discriminant strictement négatif sont donc irréductibles.

Réciproquement, il est clair que les polynômes du deuxième degré à discriminant positif ou nul sont factorisables, donc pas irréductibles. Soit enfin un P de degré supérieur ou égal à 3. Si P admet une racine réelle a, P n'est pas irréductible de façon quasi évidente. Sinon, considérons pendant quelques lignes P comme un polynôme à coefficients complexes. Par le théorème de d'Alembert-Gauss, il admet au moins une racine complexe a, qui n'est pas réelle puisqu'on a supposé P sans racine réelle. En profitant de ce que le conjugué de la somme est la somme des conjugués, que le conjugué du produit est le produit des conjugués et que chaque coefficient de P est invariant par conjugaison, on voit qu'on a aussi $P(\overline{a})=0$. Les polynômes $X-a$ et $X-\overline{a}$ étant deux irréductibles distincts dans $\mathbf{C}[X]$, le fait qu'ils divisent tous deux P entraîne que leur produit divise P dans $\mathbf{C}[X]$. Mais ce produit vaut $(X-a)(X-\overline{a})=X^2-2\,\mathcal{R}e(a)X+|a|^2$ et est donc un polynôme B du deuxième degré à coefficients réels.

Si on est distrait, on pourra croire qu'on a ainsi trouvé en B un diviseur unitaire non évident de P dans $\mathbf{R}[X]$ et conclure que P n'est pas irréductible. En réalité, on glisserait sur un détail en affirmant ceci : on sait en effet que B divise P dans $\mathbf{C}[X]$ mais il nous faut encore vérifier qu'il le divise dans $\mathbf{R}[X]$. Pour ce faire, effectuons la division euclidienne de P par B dans $\mathbf{R}[X]$: elle fournit des polynômes Q et R, avec $\mathrm{d}^\circ R<2$, tels que $P=BQ+R$. Ces polynômes de $\mathbf{R}[X]$ peuvent aussi être vus comme des polynômes à coefficients complexes, donc $P=BQ+R$ est aussi la division euclidienne de P par B dans $\mathbf{C}[X]$. Mais on sait que B divise P dans $\mathbf{C}[X]$ et que la division euclidienne est unique ; donc $R=0$, donc $P=BQ$ pour un Q à coefficients réels, et on a bien montré que B divise P dans $\mathbf{R}[X]$ aussi.

Une fois cet obstacle franchi, on conclut comme dit au début du paragraphe précédent : on a trouvé un diviseur unitaire non évident de P et celui-ci ne peut donc pas être irréductible.
•

9 - Division suivant les puissances croissantes

C'est beaucoup moins fondamental que la division euclidienne, mais c'est une technique utile pour produire des algorithmes dans des cadres assez variés.
Proposition 23-9-120 : Soit $\mathbf{K}$ un corps commutatif, A et B deux polynômes de $\mathbf{K}[X]$ et $n\geq 0$ un entier fixé. On suppose que $B(0)\neq 0$.

Alors il existe un couple (Q_n,S_n) unique (de polynômes) vérifiant la double condition :

$$A=BQ_n+X^{n+1}S_n \qquad \text{et} \qquad \mathrm{d}^\circ Q_n\leq n.$$

Démonstration : La démonstration d'existence n'est pas passionnante (simple description abstraite de l'algorithme de calcul) ; la démonstration d'unicité est plus agréable.

* Preuve de l'existence

C'est une récurrence sur l'entier $n \geq 0$.

- Pour $n = 0$, on note $a_0 = A(0)$ et $b_0 = B(0)$, puis on pose $Q_0 = a_0/b_0$ (qui existe puisque $B(0) \neq 0$). On constate alors que $A - BQ_0$ est par construction un polynôme sans terme constant, donc dans lequel X se factorise ; on peut donc mettre $A - BQ_0$ sous la forme XS_0.
- Soit n fixé et supposons le théorème vrai pour tous polynômes et tout $i \leq n$; montrons le pour les polynômes A et B de l'énoncé et pour l'entier $n+1$. Commençons par effectuer la division suivant les puissances croissantes de A par B à l'ordre n, et écrivons donc $A = BQ_n + X^{n+1}S_n$ (avec $d^\circ Q_n \leq n$), puis effectuons la division suivant les puissances croissantes de S_n par B à l'ordre 0 : on obtient une constante k et un polynôme T tels que $S_n = kB + XT$. On conclut que $A = BQ_n + kBX^{n+1} + X^{n+2}T$ et donc qu'on peut prendre $Q_{n+1} = Q_n + kX^{n+1}$ et $S_{n+1} = T$ pour répondre à la question.

* Preuve de l'unicité

Supposons qu'on ait deux écritures $A = BQ_n^{(1)} + X^{n+1}S_n^{(1)}$ et $A = BQ_n^{(2)} + X^{n+1}S_n^{(2)}$ remplissant les conditions $d^\circ Q_n^{(1)} \leq n$ et $d^\circ Q_n^{(2)} \leq n$.

Posons alors $Q_n = Q_n^{(1)} - Q_n^{(2)}$ et $S_n = S_n^{(1)} - S_n^{(2)}$ de telle sorte que $0 = BQ_n + X^{n+1}S_n$ (obtenue en soustrayant les deux écritures de A) avec en outre la condition $d^\circ Q_n \leq n$. Comme on a supposé $B(0) \neq 0$, X ne figure pas parmi les facteurs irréductibles de B, donc X^{n+1} est premier avec B. Mais d'après l'identité $0 = BQ_n + X^{n+1}S_n$, X^{n+1} divise BQ_n : on en déduit donc que X^{n+1} divise Q_n (lemme de Gauss) ; vu la condition sur le degré de Q_n, ceci entraîne que $Q_n = 0$. Dès lors $0 = X^{n+1}S_n$ donc $S_n = 0$. Les deux écritures fournies de A étaient donc la même.

•

10 - Relation entre les racines et les coefficients

Il s'agit d'écrire les coefficients d'un polynôme P scindé unitaire –ou, s'il n'est pas unitaire, les coefficients de $P/$ (coefficient dominant de P)– en fonction des racines.

Soit donc $P = \lambda(X - c_1)\cdots(X - c_d)$ un polynôme **scindé** de degré d sur un corps commutatif.

Commençons par regarder ce qui se passe pour des degrés assez petits.

En degré deux, soit $P = \lambda(X - c_1)(X - c_2)$. Écrivons par ailleurs $P = aX^2 + bX + c$. En développant la première écriture, on obtient l'identité : $\lambda X^2 + \lambda(c_1 + c_2) + \lambda c_1c_2 = aX^2 + bX + c$; en identifiant les coefficients, il vient alors $a = \lambda$ puis les deux identités bien connues :

$$c_1 + c_2 = -\frac{b}{a} \qquad c_1c_2 = \frac{c}{a}.$$

Recommençons avec du degré trois ; soit ainsi $P = \lambda(X - c_1)(X - c_2)(X - c_3)$ et écrivons par ailleurs $P = aX^3 + bX^2 + cX + d$. En développant et identifiant de la même façon, on obtient trois relations ; les deux extrêmes concernent encore la somme et le produit des racines ; celle du milieu est une nouveauté :

$$c_1 + c_2 + c_3 = -\frac{b}{a} \qquad c_1c_2 + c_1c_3 + c_2c_3 = \frac{c}{a} \qquad c_1c_2c_3 = -\frac{d}{a}.$$

Voyons enfin ce qui se passe en degré quelconque ; écrivons ici $P = a_dX^d + \cdots + a_0$ et identifions cette écriture avec le développement de $\lambda(X - c_1)\cdots(X - c_d)$; il tombe tout de suite $\lambda = a_d$; en comparant d'une part les termes en X^{d-1} et d'autre part les termes constants des deux écritures on obtient les deux relations importantes :

$$\sum_{i=1}^{d} c_i = -\frac{a_{d-1}}{a_d} \qquad \prod_{i=1}^{d} c_i = (-1)^d \frac{a_0}{a_d}$$

en regardant maintenant les termes en X^{d-k} pour un $k \leq d$ quelconque, on obtient une relation plus anecdotique à court terme ; la difficulté n'est pas de la prouver mais de l'écrire ! (la preuve rigoureuse serait une récurrence sur d) :

$$\sum_{i_1 < i_2 < \cdots < i_k} c_{i_1}c_{i_1}\cdots c_{i_k} = (-1)^k \frac{a_{d-k}}{a_d}.$$

Chapitre 24 - Fractions rationnelles

Le concept est très simple : les fractions rationnelles sont les expressions de la forme P/Q où P et Q sont des polynômes. Il reste à faire une mise en forme propre, ce qui demande un effort un peu disproportionné au caractère très intuitif de l'objet à construire.

Le chapitre se termine par le théorème de décomposition en , utilisé notamment pour le calcul des primitives de fractions rationnelles, et qui est un peu indigeste...

Dans tout le chapitre, $\mathbf{K}$ désigne un corps commutatif.

1 - Définition des fractions rationnelles

Définition 24-1-176 : On appelle **corps des fractions rationnelles** sur $\mathbf{K}$ tout corps commutatif $\mathbf{B}$ vérifiant :

i) l'anneau commutatif $\mathbf{K}[X]$ est inclus dans $\mathbf{B}$; (ou plus précisément il existe une application j injective de $\mathbf{K}[X]$ vers B telle qui soit un morphisme d'anneaux de $\mathbf{K}[X]$ vers B).

ii) tout élément de $\mathbf{B}$ peut s'écrire P/Q où P et Q sont deux éléments de $\mathbf{K}[X]$.

Notation 24-1-69 : Un corps de fractions rationnelles sur $\mathbf{K}$ est noté $\mathbf{K}(X)$.

2 - Les fractions rationnelles existent

Comme pour les polynômes, l'existence est intuitivement assez évidente, mais nécessite une preuve un peu abstraite ; et elle est complétée par un énoncé d'unicité un peu technique qui est implicitement utilisé par la suite pour parler "du" corps des fractions rationnelles.

Théorème 24-2-51 : Soit $\mathbf{K}$ un corps commutatif.

a) Il existe un corps de fractions rationnelles sur $\mathbf{K}$.

b) Si $\mathbf{B}_1$ et $\mathbf{B}_2$ sont deux tels corps de fractions rationnelles, il existe un isomorphisme de corps φ de $\mathbf{B}_1$ sur $\mathbf{B}_2$ dont la restriction à $\mathbf{K}[X]$ soit l'application identique.

Démonstration :

Notons $A = \mathbf{K}[X]$. La démonstration utilise simplement le fait que A est un anneau intègre, et nullement en réalité que A est l'anneau des polynômes (le théorème pourrait donc facilement être donné avec un énoncé plus général, mais je ne l'ai pas jugé utile).

La première idée qui peut venir à l'esprit est de tenter de modéliser la fraction P/Q par le couple (P, Q) qui contient à première vue la même information : ainsi la fraction $\dfrac{X}{X+1}$ correspondra au couple $(X, X+1)$. Une telle idée nous met sur la bonne piste, mais elle se heurte à un problème : le couple (X^2, X^2+X) représentera la fraction $\dfrac{X^2}{X^2+X} = \dfrac{X}{X+1}$; la même fraction correspond donc à plusieurs couples, et l'idée d'utiliser pour $\mathbf{B}$ l'ensemble des couples (P, Q) nous donne un ensemble trop gros.

On pourrait penser à n'utiliser que des couples (P, Q) avec P et Q premiers entre eux ; c'est vraisemblablement faisable, mais la preuve risque d'être extrêmement lourde, avec des PGCD à simplifier de partout.

Non, décidément, on ne fera rien de simple si on n'a pas compris ce qu'est un ensemble-quotient, alors que si on maîtrise cette notion, la preuve est longue à écrire, mais sans obstacles.

Attaquons formellement. Notons $D = A \times (A \setminus \{0\})$ (c'est la première idée suggérée. Sur D on introduit deux opérations $+$ et $\times$ définies comme suit : pour tous (P_1, Q_1) et (P_2, Q_2) de D, on pose

$$(P_1, Q_1) \times (P_2, Q_2) = (P_1P_2, Q_1Q_2) \qquad \text{et} \qquad (P_1, Q_1) + (P_2, Q_2) = (P_1Q_2 + P_2Q_1, Q_1Q_2).$$

(On notera qu'on utilise très discrètement l'intégrité de A pour justifier que le produit Q_1Q_2 qui intervient dans les formules n'est pas nul).

Ces deux formules se comprennent si on a en tête qu'un couple (P, Q) a vocation à décrire la fraction P/Q (qui n'aura un sens propre qu'une fois la démonstration terminée) : elles sont les reproductions des formules qu'on sait bien utiliser pour multiplier ou additionner des fractions.

L'ensemble obtenu a une bonne tête vu de loin, mais de près il est trop gros.

Pour le faire maigrir, introduisons une relation $\sim$ sur D définie par : pour tous (P_1, Q_1) et (P_2, Q_2) de D,

$$(P_1, Q_1) \sim (P_2, Q_2) \qquad \text{lorsque} \qquad P_1 Q_2 = P_2 Q_1.$$

(Si nous savions déjà donner un sens aux barres de fractions, nous aurions écrit la condition sous la forme $P_1/Q_1 = P_2/Q_2$, la rendant ainsi compréhensible, mais comme ce symbole ne nous sera disponible qu'une fois finie la démonstration, on a dû donner une forme moins limpide).

Il est très facile de vérifier que $\sim$ est une relation d'équivalence sur D. On note B l'ensemble-quotient $D/\sim$.

On va alors définir des opérations $+$ et X sur B ; le principe est le même que celui qui nous a permis de définir addition et multiplication sur $\mathbf{Z}/n\mathbf{Z}$: on définit simplement ces opérations sur des représentants des classes d'équivalence, et on vérifie méthodiquement que le résultat obtenu ne dépend pas de la classe utilisée.

Posons donc, pour $\overbrace{(P_1, Q_1)}^{\cdot}$ et $\overbrace{(P_2, Q_2)}^{\cdot}$ éléments de B :

$$\overbrace{(P_1, Q_1)}^{\cdot} \times \overbrace{(P_2, Q_2)}^{\cdot} = \overbrace{(P_1, Q_1) \times (P_2, Q_2)}^{\cdot} \qquad \text{et} \qquad \overbrace{(P_1, Q_1)}^{\cdot} + \overbrace{(P_2, Q_2)}^{\cdot} = \overbrace{(P_1, Q_1) + (P_2, Q_2)}^{\cdot}.$$

Il faut maintenant vérifier soigneusement que le résultat de ces opérations ne dépend pas des représentants choisis... Faisons-le soigneusement pour l'addition, avec "renvoi au lecteur" pour la multiplication. On notera que les "prime" dans les calculs qui suivent n'ont rien à voir avec des dérivées :

soit (P_1', Q_1') un autre représentant de la classe de (P_1, Q_1) et de même soit $(P_2', Q_2') \sim (P_2, Q_2)$. Il faut vérifier que $(P_1, Q_1) + (P_2, Q_2) = (P_1 Q_2 + P_2 Q_1, Q_1 Q_2) \sim (P_1' Q_2' + P_2' Q_1', Q_1' Q_2') = (P_1', Q_1') + (P_2', Q_2')$.

Cela revient très bêtement à comparer les produits $(P_1 Q_2 + P_2 Q_1) Q_1' Q_2'$ et $(P_1' Q_2' + P_2' Q_1') Q_1 Q_2$; on dispose pour ce faire des égalités $P_1 Q_1' = P_1' Q_1$ (issue de la relation $(P_1, Q_1) \sim (P_1', Q_1')$) et $P_2 Q_2' = P_2' Q_2$ (issue de la relation $(P_2, Q_2) \sim (P_2', Q_2')$).

La vérification est alors stupide : $(P_1 Q_2 + P_2 Q_1) Q_1' Q_2' = P_1 Q_1' Q_2 Q_2' + P_2 Q_2' Q_1 Q_1' = P_1' Q_1 Q_2 Q_2' + P_2' Q_2 Q_1 Q_1' = (P_1' Q_2' + P_2' Q_1') Q_1 Q_2$.

On a donc bien construit un ensemble $\mathbf{K}(X)$ puis construit sur celui-ci une addition et une multiplication.

L'étape suivante serait de vérifier que cette addition et cette multiplication en font un corps commutatif... C'est une simple vérification méthodique et lourde de toutes les propriétés de la définition d'un corps commutatif. Je me bornerai ici à justifier l'existence de l'inverse : si une classe $\overbrace{(P_1, Q_1)}^{\cdot}$ n'est pas nulle, on remarque d'abord que $P_1 \neq 0$ (puisque $(P_1, Q_1) \not\sim (0, 1)$). La classe $\overbrace{(Q_1, P_1)}^{\cdot}$ existe donc ; ce sera l'inverse de $\overbrace{(P_1, Q_1)}^{\cdot}$: en effet le produit des deux est $\overbrace{(Q_1 P_1, P_1 Q_1)}^{\cdot}$ qui est égal à la classe de $(1, 1)$ qui est le neutre pour la multiplication.

L'inclusion de $\mathbf{K}[X]$ dans $\mathbf{K}(X)$ s'obtient en envoyant un polynôme P sur la classe $\overbrace{(P, 1)}^{\cdot}$. Il est très facile de vérifier qu'elle transforme addition en addition, multiplication en multiplication ; son injectivité peut seule interpeller. Puisque cette transformation est un morphisme de groupes additifs, l'injectivité se laisse montrer à coup de noyaux ; et effectivement si un polynôme P est envoyé sur le neutre additif de $\mathbf{K}(X)$ qui est la classe de $(0, 1)$, c'est que $(P, 1) \sim (0, 1)$ et donc que $P = 0$: le noyau est bien réduit au seul polynôme nul.

Enfin tout élément de $\mathbf{K}(X)$ se met bien sous forme P/Q puisque :

$$\overbrace{(P, Q)}^{\cdot} = \overbrace{(P, 1)}^{\cdot}\overbrace{(1, Q)}^{\cdot} = \overbrace{(P, 1)}^{\cdot} \left[\overbrace{(Q, 1)}^{\cdot}\right]^{-1} = P/Q.$$

J'oublie malencontreusement d'écrire la preuve du b), personne ne s'en apercevra.

•

3 - Décomposition en éléments simples

Voici l'énoncé du

Théorème 24-3-52 :

Soit une fraction rationnelle $\dfrac{P}{Q} \in \mathbf{K}[X]$ et soit la décomposition en produits de polnômes unitaires irréductibles de Q :

$$Q = \lambda Q_1^{\alpha_1} Q_2^{\alpha_2} \cdots Q_k^{\alpha_k}.$$

Alors il existe une et une seule écriture :

$$\frac{P}{Q} = R + \frac{A_{1,1}}{Q_1} + \cdots + \frac{A_{1,\alpha_1}}{Q_1^{\alpha_1}} + \frac{A_{2,1}}{Q_2} + \cdots + \frac{A_{2,\alpha_2}}{Q_2^{\alpha_2}} + \cdots + \frac{A_{k,1}}{Q_k} + \cdots \frac{A_{k,\alpha_k}}{Q_k^{\alpha_k}}$$

dans laquelle R et les $A_{i,j}$ sont tous des polynômes de $\mathbf{K}[X]$ vérifiant en outre la condition suivante sur leurs degrés :

pour tout (i,j) (où $1 \leq i \leq k$, $1 \leq j \leq \alpha_i$), $\mathrm{d}^\circ A_{i,j} < \mathrm{d}^\circ Q_i$.

Démonstration :

Preuve de l'existence :

Dans un premier temps, on va considérer les polynômes :

$$T_1 = Q_2^{\alpha_2} Q_3^{\alpha_3} \cdots Q_k^{\alpha_k}, T_2 = Q_1^{\alpha_2} Q_3^{\alpha_3} \cdots Q_k^{\alpha_k}, \ldots, T_k = Q_1^{\alpha_1} Q_2^{\alpha_2} \cdots Q_{k-1}^{\alpha_{k-1}}.$$

Un éventuel diviseur irréductible unitaire commun à tous ces polynômes doit diviser T_k ; ce doit donc être un Q_i avec $i < k$. Mais Q_1 ne divise pas T_1, Q_2 ne divise pas T_2, et ce jusqu'à Q_{k-1} qui ne divise pas T_{k-1}. Les polynômes $T_1, \cdots, T_k$ ne possèdent donc aucun diviseur irréductible unitaire commun ; ils sont donc premiers entre eux.

On peut donc écrire une identité de Bezout :

$$1 = S_1 T_1 + S_2 T_2 + \cdots + S_k T_k$$

pour des polynômes $S_1, \cdots, S_k$ de $\mathbf{K}[X]$.

Multiplions cette identité par $\dfrac{P}{Q} = \dfrac{P}{\lambda Q_1^{\alpha_1} Q_2^{\alpha_2} \cdots Q_k^{\alpha_k}}$; on obtient :

$$\begin{aligned}
\frac{P}{Q} &= PS_1 \frac{T_1}{Q} + PS_2 \frac{T_2}{Q} + \cdots + PS_k \frac{T_k}{Q} \\
&= \frac{PS_1}{\lambda} \frac{\lambda T_1}{Q} + \frac{PS_2}{\lambda} \frac{\lambda T_2}{Q} + \cdots + \frac{PS_k}{\lambda} \frac{\lambda T_k}{Q} \\
&= \frac{PS_1}{\lambda} \frac{1}{Q_1^{\alpha_1}} + \frac{PS_2}{\lambda} \frac{1}{Q_2^{\alpha_2}} + \cdots + \frac{PS_k}{\lambda} \frac{1}{Q_k^{\alpha_k}}
\end{aligned}$$

En notant $B_1, \cdots, B_k$ les divers numérateurs intervenant dans cette formule, on a donc réussi à écrire :

$$\frac{P}{Q} = \frac{B_1}{Q_1^{\alpha_1}} + \frac{B_2}{Q_2^{\alpha_2}} + \cdots + \frac{B_k}{Q_k^{\alpha_k}}.$$

On va alors manipuler successivement chacun des termes de cette addition. Fixons un i avec $1 \leq i \leq k$ et travaillons l'expression $\dfrac{B_i}{Q_i^{\alpha_i}}$.

On commence par faire la division euclidienne de B_i par Q_i, en notant judicieusement le quotient et le reste :

$$B_i = B_{i,\alpha_i} Q_i + A_{i,\alpha_i} \text{ avec } \mathrm{d}^\circ A_{i,\alpha_i} < \mathrm{d}^\circ Q_i$$

En reportant cette division euclidienne en lieu et place de B_i on a réécrit :

$$\frac{B_i}{Q_i^{\alpha_i}} = \frac{B_{i,\alpha_i}}{Q_i^{\alpha_i - 1}} + \frac{A_{i,\alpha_i}}{Q_i^{\alpha_i}}.$$

On recommence une division euclidienne, cette fois-ci de B_{i,α_i} par Q_i, en notant toujours opportunément quotient et reste :

$$B_{i,\alpha_i} = B_{i,\alpha_i-1}Q_i + A_{i,\alpha_i-1} \text{ avec } \mathrm{d}^\circ A_{i,\alpha_i-1} < \mathrm{d}^\circ Q_i$$

et on reporte de nouveau dans l'expression la plus fraîche de $\dfrac{B_i}{Q_i^{\alpha_i}}$; on obtient :

$$\frac{B_i}{Q_i^{\alpha_i}} = \frac{B_{i,\alpha_i-1}}{Q_i^{\alpha_i-2}} + \frac{A_{i,\alpha_i-1}}{Q_i^{\alpha_i-1}} + \frac{A_{i,\alpha_i}}{Q_i^{\alpha_i}}.$$

On recommence jusqu'à ne plus pouvoir recommencer... On a finalement une expression :

$$\frac{B_i}{Q_i^{\alpha_i}} = B_{i,1} + \frac{A_{i,1}}{Q_i} + \cdots + \frac{A_{i,\alpha_i-1}}{Q_i^{\alpha_i-1}} + \frac{A_{i,\alpha_i}}{Q_i^{\alpha_i}}.$$

Il n'y a plus qu'à regrouper toutes ces expressions et noter

$$R = B_{1,1} + B_{2,1} + \cdots + B_{k,1}$$

pour avoir terminé la preuve d'existence.

Preuve de l'unicité :

Je l'écrirai (peut-être) l'année prochaine, elle n'est pas spécialement amusante... Contrairement à la preuve d'existence, il n'y a guère d'idées, seulement des décomptes de degrés.

•

Chapitre 25 - Intégration des fonctions continues par morceaux

Vous savez calculer l'intégrale de plus d'une fonction continue (enfin, je l'espère). L'objectif de ce chapitre est de montrer que l'intégrale existe même pour des fonctions continues pour lesquelles on ne saurait la calculer, et accessoirement de prouver quelques unes de ses propriétés simples.

À la fois pour des raisons techniques (les fonctions en escalier ne sont pas continues, et sont des outils bien pratiques pour construire l'intégrale) que pratiques (on a effectivement vraiment besoin dans des situations réelles d'intégrer des fonctions présentant quelques discontinuités) on va étendre ce projet à une classe de fonctions un peu plus large que les fonctions continues : les fonctions continues par morceaux.

1 - Fonctions continues par morceaux sur un segment fermé borné

Les définitions étant plus simples sur un segment fermé borné que sur un intervalle quelconque, on reportera le cas général à quelques remarques en fin de chapitre, et on ne travaillera dans cette section et la suivante que sur un intervalle $[a,b]$ fermé et borné (avec $a<b$).

Définition 25-1-177 : Soit $[a,b]$ un intervalle fermé borné (avec $a<b$) et f une fonction de $[a,b]$ vers $\mathbf{R}$. On dit que f est une **fonction en escalier** lorsqu'il existe un entier n (avec $1\leq n$) et des réels $a=c_0<c_1<\cdots<c_n=b$ tels que f soit constante sur chaque intervalle $]c_i,c_{i+1}[$ (où $0\leq i<n$).

Définition 25-1-178 : Soit $[a,b]$ un intervalle fermé borné (avec $a<b$) et f une fonction de $[a,b]$ vers $\mathbf{R}$. On dit que f est une fonction **continue par morceaux** lorsqu'il existe un entier n (avec $1\leq n$) et des réels $a=c_0<c_1<\cdots<c_n=b$ tels que f soit continue sur chaque intervalle $]c_i,c_{i+1}[$ (où $0\leq i<n$), que chaque limite à droite $\lim\limits_{\substack{t\to c_i\\ c_i<t}} f(t)$ (où $0\leq i<n$) existe et chaque limite à gauche $\lim\limits_{\substack{t\to c_i\\ t<c_i}} f(t)$ (où $0<i\leq n$) existe.

Il est évident que les fonctions continues et les fonctions en escalier sont des exemples simples de fonctions continues par morceaux. Tout s'y ramène par le

Lemme 25-1-13 : Soit $[a,b]$ un intervalle fermé borné (avec $a<b$) et f une fonction de $[a,b]$ vers $\mathbf{R}$. Alors f est continue par morceaux si et seulement s'il existe une fonction continue g de $[a,b]$ vers $\mathbf{R}$ et une fonction en escalier h de $[a,b]$ vers $\mathbf{R}$ telles que $f=g+h$.

Démonstration :

Preuve de $\Leftarrow$ (qui est le sens le plus facile). Soit $a=c_0<c_1<\cdots<c_n=b$ un découpage de l'intervalle $[a,b]$ tel que h soit constante sur chaque intervalle $]c_i,c_{i+1}[$. Alors sur chacun de ces intervalles h et g sont continues, donc aussi f. De plus en chaque point de ce découpage, g et h ont une limite à droite et à gauche, donc f aussi.

Preuve de $\Rightarrow$. Supposons f continue par morceaux et soit $a=c_0<c_1<\cdots<c_n=b$ un découpage associé à la définition de cette continuité par morceaux. On va construire h en escalier avec ce découpage. Pour cela, notons $l_i^+=\lim\limits_{\substack{t\to c_i\\ c_i<t}} f(t)$ (pour $0\leq i<n$) et $l_i^-=\lim\limits_{\substack{t\to c_i\\ t<c_i}} f(t)$ (où $0<i\leq n$) les limites à gauche et à droite respectives de f aux divers points du découpage. On construit alors h en posant $h(c_0)=f(c_0)$, puis $h(t)=l_0^+$ pour $a=c_0<t<c_1$, puis $h(c_1)=l_0^+-l_1^-+f(c_1)$, puis $h(t)=l_0^+-l_1^-+l_1^+$ pour $c_1<t<c_2$. Plus généralement, on pose $h(t)=l_0^+-l_1^-+l_1^++\cdots-l_i^-+l_i^+$pour $c_i<t<c_{i+1}$, et $h(c_i)=l_0^+-l_1^-+l_1^++\cdots-l_i^-+f(c_i)$. Par construction, h est en escalier.

Posons maintenant $g=f-h$: la relation $f=g+h$ ne pose pas de problèmes, la seule chose à vérifier est la continuité de g. Celle-ci est évidente en tout t autre qu'un des points c_i et en chacun de ceux-ci elle demande une vérification ; pour $i=0$ on compare $g(c_0)=f(c_0)-h(c_0)=0$ à $\lim\limits_{\substack{t\to c_0\\ c_0<t}} g(t)=\lim\limits_{\substack{t\to c_0\\ c_0<t}} f(t)-\lim\limits_{\substack{t\to c_0\\ c_0<t}} h(t)=l_0^+-l_0^+=0$ pour conclure à la continuité en $c_0=a$; pour $0<i<n$ on compare d'une part $g(c_i)=f(c_i)-h(c_i)=f(c_i)-\left(l_0^+-l_1^-+l_1^++\cdots-l_i^-+f(c_i)\right)=-l_0^++l_1^--l_1^++\cdots+l_i^-$, d'autre part $\lim\limits_{\substack{t\to c_i\\ t<c_i}} g(t)=\lim\limits_{\substack{t\to c_i\\ t<c_i}} f(t)-\lim\limits_{\substack{t\to c_i\\ t<c_i}} h(t)=l_i^--(l_0^+-l_1^-+l_1^++\cdots-l_{i-1}^-+l_{i-1}^+)=-l_0^++l_1^--l_1^++\cdots+l_i^-$ et enfin $\lim\limits_{\substack{t\to c_i\\ c_i<t}} g(t)=\lim\limits_{\substack{t\to c_i\\ c_i<t}} f(t)-\lim\limits_{\substack{t\to c_i\\ c_i<t}} h(t)=l_i^+-\left(l_0^+-l_1^-+l_1^++\cdots-l_i^-+l_i^+\right)=-l_0^++l_1^--l_1^++\cdots+l_i^-$ pour

conclure à la continuité de g en x_i ; et pour $i = n$ on fait la même vérification (mais à gauche seulement de $x_n = b$).

•

La simplicité de ce lemme explique je l'espère *a posteriori* pourquoi on a introduit dans la définition des fonctions continues par morceaux la compliquée condition d'existence de limites à droite et à gauche.

2 - Primitives et primitives par morceaux

Pour des fonctions continues, la bonne notion de "primitive" sera celle à laquelle on s'attend :

Définition 25-2-179 : Soit $[a, b]$ un intervalle fermé borné (avec $a < b$) et f, F deux fonctions de $[a, b]$ vers $\mathbf{R}$, où la fonction f est continue. On dit que F est une **primitive** de f lorsque f est dérivable sur $[a, b]$ et $F' = f$.

Pour des fonctions continues par morceaux, il faut renoncer à avoir la relation $F' = f$ en tout point de l'intervalle : c'est désespéré aux éventuelles discontinuités de f. On perdra donc la dérivabilité de F ; notez bien que la définition suivante exige la continuité de F.

Définition 25-2-180 : Soit $[a, b]$ un intervalle fermé borné (avec $a < b$) et f, F deux fonctions de $[a, b]$ vers $\mathbf{R}$, où la fonction f est continue par morceaux. On dit que F est une **primitive par morceaux** de f lorsque F est continue sur $[a, b]$, dérivable sauf peut-être en un nombre fini de points et on a l'identité $F'(t) = f(t)$, sauf peut-être en un nombre fini de points.

Ce concept est le bon pour pouvoir intégrer les fonctions en escalier.

Proposition 25-2-121 : Soit $[a, b]$ un intervalle fermé borné (avec $a < b$) et f une fonction en escalier sur $[a, b]$. La fonction f admet (au moins) une primitive par morceaux.

Démonstration : Soit Soit $a = c_0 < c_1 < \cdots < c_n = b$ un découpage de l'intervalle $[a, b]$ tel que f soit constante sur chaque intervalle $]c_i, c_{i+1}[$. Notons d_i la valeur constante de f sur $]c_i, c_{i+1}[$ (pour $O \leq i < n$).

On construit tout à fait explicitement F en posant $F(c_0) = 0$, puis pour $a = c_0 < t \leq c_1$, $F(t) = d_0(t - c_0)$, puis plus généralement pour $c_i < t \leq c_{i+1}$, $F(t) = d_0(c_1 - c_0) + d_1(c_2 - c_1) + \cdots + d_i(t - c_i)$.

Avec ces formules, il est clair que F est dérivable sauf peut-être aux points c_i (avec $0 \leq i < n$) (qui sont en nombre fini) et que sur chaque intervalle $]c_i, c_{i+1}[$ sa dérivée vaut d_i donc coïncide avec f. Le seul point à vérifier est la continuité de F en chaque c_i (et la continuité à gauche est évidente) ; la seule chose à faire est donc de comparer la valeur de $F(c_i)$ qui est $d_0(c_1 - c_0) + d_1(c_2 - c_1) + \cdots + d_{i-1}(c_i - c_{i-1})$ et sa limite à droite, qui est la limite quand t tend vers c_i de $d_0(c_1 - c_0) + d_1(c_2 - c_1) + \cdots + d_i(t - c_i)$, soit $d_0(c_1 - c_0) + d_1(c_2 - c_1) + \cdots + d_{i-1}(c_i - c_{i-1}) + 0$: on retrouve la même chose.

•

L'intérêt des primitives par morceaux est que certains résultats du cours de dérivation sont encore vrais avec cette notion un peu étendue.

Proposition 25-2-122 : Soit $[a, b]$ un intervalle fermé borné (avec $a < b$) et f, F deux fonctions de $[a, b]$ vers $\mathbf{R}$, où la fonction f est continue par morceaux et F est une primitive par morceaux de f.

Alors F est croissante si et seulement si f est positive sauf peut-être en un nombre fini de points.

Démonstration : Une implication est claire : si F est croissante, là où elle est dérivable sa dérivée est positive, donc f est positive sauf peut-être en un nombre fini de points.

Réciproquement, supposons f positive sauf peut-être en un nombre fini de points. En ajoutant à ces points les points éventuels où l'égalité $F'(t) = f(t)$ n'est pas vraie, et éventuellement aussi les points a et b, on en déduit qu'il existe un nombre fini de points $a = c_0 < c_1 < \cdots < c_n = b$ tels que sur chaque intervalle $]c_i, c_{i+1}[$ la fonction F est dérivable et de dérivée positive. On en déduit que la fonction F est croissante sur chaque intervalle $]c_i, c_{i+1}[$. Comme on a supposé en outre la fonction F continue (c'est là qu'on s'en sert de façon cruciale) on peut passer à la limite quand s tend vers c_i à droite dans l'inégalité $f(s) \leq f(t)$ pour $c_i < s \leq t < c_{i+1}$ et déduire que $f(c_i) \leq f(t)$; en agissant de même à gauche de c_{i+1} on montre ainsi la croissance de F sur chaque intervalle fermé $[c_i, c_{i+1}]$. Ceci entraîne évidemment la croissance de F sur $[a, b]$ tout entier.

•

Proposition 25-2-123 : Soit $[a, b]$ un intervalle fermé borné (avec $a < b$) et f, F deux fonctions de $[a, b]$ vers $\mathbf{R}$, où la fonction f est continue par morceaux et F est une primitive par morceaux de f.

Alors F est constante si et seulement si f est nulle sauf peut-être en un nombre fini de points.

Démonstration : Il suffit d'appliquer la proposition précédente d'une part à f et F et d'autre part à $-f$ et $-F$.

•

Corollaire 25-2-6 : Deux primitives par morceaux d'une même fonction continue par morceaux sur un intervalle fermé borné diffèrent d'une constante.
Démonstration : Soit F_1 et F_2 deux primitives d'une même f continue par morceaux. Alors $F_1 - F_2$ est continue, dérivable sauf peut-être en un nombre fini de points, et sa dérivée est égale à $f - f = 0$ sauf peut-être en un nombre fini de points : $F_1 - F_2$ est donc une primitive par morceaux de la fonction nulle, donc une constante.
•

Pour des primitives par morceaux, le théorème des accroissements finis dans sa version la plus précise (existence d'un c) peut échouer, mais il reste une inégalité.
Proposition 25-2-124 : Soit $[a, b]$ un intervalle fermé borné (avec $a < b$) et f, F deux fonctions de $[a, b]$ vers $\mathbf{R}$, où la fonction f est continue par morceaux et F est une primitive par morceaux de f.

Soit M un réel fixé ; on suppose que pour tout t de $[a, b]$ (ou même sauf peut-être un nombre fini de t) on a l'inégalité : $f(t) \leq M$. Alors :

$$\frac{F(b) - F(a)}{b - a} \leq M.$$

Démonstration : Posons $g(t) = M - f(t)$ et $G(t) = Mt - F(t)$. Il est alors immédiat de vérifier que G est une primitive par morceaux de g et que g est positive (sauf peut-être en un nombre fini de points). La fonction G est donc croissante, donc $G(a) \leq G(b)$, soit $Ma - F(a) \leq Mb - F(b)$, soit $F(b) - F(a) \leq M(b - a)$.
•

3 - Les fonctions continues ont des primitives

Théorème 25-3-53 : Soit $[a, b]$ un intervalle fermé borné (avec $a < b$) et f une fonction continue de $[a, b]$ vers $\mathbf{R}$.

Alors f admet au moins une primitive F.

Les primitives de f sont exactement les fonctions $F + c$ où c est une fonction constante.

Démonstration : Le dernier point est évident et découle simplement de la caractérisation des fonctions constantes comme fonctions de dérivée nulle. Toute la difficulté (et elle n'est pas petite) consiste à prouver l'existence d'au moins une primitive. Il faut bien être conscient qu'elle existe, mais qu'on ne la trouvera pas par une formule : pour des fonctions continues simples, comme $f(t) = \dfrac{e^t}{t}$ les primitives existent mais ne se laissent pas calculer.
La méthode va consister à approcher f par des fonctions en escalier, dont on sait trouver des primitives (par morceaux) puis passer à la limite à partir de ces primitives par morceaux.
Pour cela, il va falloir avaler quelques notations. Notons, pour $n \geq 1$ et $0 \leq k \leq n$, $x_n^{(k)} = a + \dfrac{k(b-a)}{n}$. Dit avec des mots :

"$x_n^{(k)}$ est l'extrémité droite du k-ème morceau du découpage de $[a, b]$ en n parts égales"

Soit maintenant f_n (pour $n \geq 1$) la fonction en escalier définie sur $[a, b]$ par :

$$\text{pour } x_n^{(k)} \leq t < x_n^{(k+1)},\ f_n(t) = f(x_n^{(k)}) \text{ (et } f_n(b) = f(b)).$$

Avec des mots :

"f_n est la fonction en escalier, constante sur chaque morceau du découpage de $[a, b]$ en n parts égales, qui prend sur chacun de ces morceaux la valeur que prend f à son extrémité gauche".

Préalablement à la construction, on va montrer l'**affirmation 1** suivante, cruciale pour la démonstration :
Pour tout $\epsilon > 0$ il existe un $N \geq 1$ tel que pour tout $n \geq N$ et tout $t \in [a, b]$, $|f_n(t) - f(t)| \leq \epsilon$.
(Avec un mot du programme de deuxième année, "f_n tend uniformément vers f").
Preuve de l'affirmation 1 (par l'absurde) : Supposons que ce soit faux. Il existerait alors un $\epsilon > 0$ tel que pour tout $N \geq 1$, il existe un $n_N \geq N$ et un $t_N \in [a, b]$ tels que $\epsilon < |f_{n_N}(t_N) - f(t_N)|$.
Pour chaque $N \geq 1$, notons s_N l'extrémité gauche de l'intervalle du découpage régulier de $[a, b]$ en n_N morceaux auquel appartient t_N. Ainsi $s_N \leq t_N$ et $t_N - s_N < \dfrac{b-a}{n_N} \leq \dfrac{b-a}{N}$, donc $s_N - t_N \to 0$ quand $N \to \infty$. De plus par définition des f_n, $f_{n_N}(t_N) = f(s_N)$. L'inégalité $\epsilon < |f_{n_N}(t_N) - f(t_N)|$ se réécrit donc plus brièvement $\epsilon < |f(s_N) - f(t_N)|$

Par le théorème de Bolzano-Weierstrass, il existe une suite-extraite $(t_{\varphi(N)})$ de (t_N) qui admette une limite l. Comme $s_N - t_N \to 0$ quand $N \to \infty$, la suite $(s_{\varphi(N)})$ converge aussi vers l. En passant à la limite dans l'inégalité $\epsilon < |f(s_{\varphi(N)}) - f(t_{\varphi(N)})|$, on obtient $\epsilon \leq |f(l) - f(l)| = 0$. Ce qui est contradictoire.
L'affirmation 1 est donc démontrée.

On sait désormais assez sur les f_n pour avancer dans la construction. Chaque f_n est une fonction en escalier ; elle admet donc des primitives par morceaux. Notons F_n la primitive de f_n telle que $F_n(a) = 0$.

Affirmation 2 : Pour chaque $t \in [a, b]$ fixé, la suite $(F_n(t))$ est une suite de Cauchy.
Preuve de l'affirmation 2 : Fixons un $t \in [a, b]$; si $t = a$ le résultat est évident (tous les $F_n(t)$ sont nuls) ; on supposera donc $a < t$.
Appliquons l'affirmation 1 au réel $\epsilon_1 = \dfrac{\epsilon}{2(t-a)}$. Elle nous fournit un $N \geq 1$ tel que pour tout $n \geq N$ et tout $s \in [a, b]$, on ait $|f_n(s) - f(s)| \leq \epsilon_1$.
On en déduit que pour tous p, q avec $p \geq q \geq N$, et tout $s \in [a, b]$, on a :

$$|f_p(s) - f_q(s)| = |(f_p(s) - f(s)) + (f(s) - f_q(s))| \leq |f_p(s) - f(s)| + |f(s) - f_q(s)| \leq 2\epsilon_1.$$

Notons maintenant que $F_p - F_q$ est une primitive par morceaux de $f_p - f_q$ et appliquons la proposition 25-2-124 aux inégalités $f_p - f_q \leq 2\epsilon_1$ et $f_q - f_p \leq 2\epsilon_1$, valables sur tout l'intervalle $[a, b]$, donc sur tout l'intervalle $[a, t]$. On en déduit que

$$(F_p - F_q)(t) - (F_p - F_q)(a) \leq 2\epsilon_1(t - a) \leq \epsilon$$

et symétriquement en échangeant p et q, c'est-à-dire exactement (en se souvenant que $F_p(a) = F_q(a) = 0$) l'inégalité $|F_p(t) - F_q(t)| \leq \epsilon$.
L'affirmation 2 est bien prouvée.

Arrivé à ce point des constructions, on en déduit que pour chaque t fixé, la suite de Cauchy $(F_n(t))_{n\geq 1}$ est une suite convergente. Notons $F(t)$ sa limite. La fonction F va être la primitive cherchée... Reste encore à le montrer. Ce n'est pas franchement astucieux, mais tout de même un peu indigeste parce qu'un peu lourd. Il faut en effet revenir à la définition même d'une dérivée comme limite...
Soit donc un $t_0 \in [a, b]$ fixé, et un $\epsilon > 0$ fixé. L'objectif sera de trouver un $\eta > 0$ tel que dès que $|t - t_0| \leq \eta$ (avec $t \in [a, b]$), on ait :

$$|f(t_0) - \frac{F(t) - F(t_0)}{t - t_0}| \leq \epsilon.$$

Pour ce faire, commençons par appliquer la définition de "continuité" à f au point t_0 et à $\dfrac{\epsilon}{2}$; ceci nous fournit un $\eta > 0$ tel que pour $|s - t_0| \leq \eta$ (avec $s \in [a, b]$), on ait : $|f(s) - f(t_0)| \leq \dfrac{\epsilon}{2}$.
Appliquons alors l'affirmation 1 à $\dfrac{\epsilon}{2}$; elle nous garantit l'existence d'un $N \geq 1$ tel que pour $n \geq N$, on ait pour tout $t \in [a, b]$, $|f(t) - f_n(t)| \leq \dfrac{\epsilon}{2}$. On a alors pour tout $s \in [a, b]$ tel que $|s - t_0| \leq \eta$ et tout $n \geq N$:

$$|f_n(s) - f(t_0)| = |(f_n(s) - f(s)) + (f(s) - f(t_0))| \leq |f_n(s) - f(s)| + |f(s) - f(t_0)| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Remarquons maintenant que la fonction auxiliaire G_n définie par $G_n(t) = F_n(t) - tf(t_0)$ est une primitive par morceaux de la fonction en escalier $f_n - f(t_0)$. Dès que l'on prend un $t \in [a, b]$ tel que $|t - t_0| \leq \eta$, on a pour tout point s du segment fermé d'extrémités t_0 et t les inégalités

$$f_n(s) - f(t_0) \leq \epsilon \qquad \text{et} \qquad f(t_0) - f_n(s) \leq \epsilon$$

dont on déduit (encore une fois par la proposition auxiliaire 25-2-124) les inégalités :

$$-\epsilon \leq \frac{G_n(t) - G_n(t_0)}{t - t_0} \leq \epsilon.$$

Mais $\dfrac{G_n(t) - G_n(t_0)}{t - t_0} = \dfrac{F_n(t) - F_n(t_0)}{t - t_0} - f(t_0)$: on a donc prouvé (pour tout $n \geq N$ et tout $t \in [a, b]$ tel que $|t - t_0| \leq \eta$) l'inégalité :

$$|f(t_0) - \frac{F_n(t) - F_n(t_0)}{t - t_0}| \leq \epsilon.$$

Il ne reste plus qu'à faire tendre n vers l'infini dans cette inégalité pour obtenir l'inégalité cherchée. •

Théorème 25-3-54 : Soit $[a, b]$ un intervalle fermé borné (avec $a < b$) et f une fonction continue par morceaux de $[a, b]$ vers $\mathbf{R}$.

Alors f admet au moins une primitive par morceaux F.

Les primitives par morceaux de f sont exactement les fonctions $F + c$ où c est une fonction constante.

Démonstration : Là aussi la dernière affirmation est facile (du moins une fois qu'on a préparé le terrain en ayant étudié les primitives par morceaux de zéro).

Pour la première affirmation on a aussi préparé le terrain : écrivons $f = g + h$ où g est continue et h en escalier. Le théorème précédent permet de trouver une primitive de g et on sait déjà (on s'en est abondamment servi pour prouver le théorème précédent...) que h admet des primitives par morceaux. On obtient F en additionnant primitive de g et primitive par morceaux de h. •

4 - Extensions à des intervalles autres que fermés bornés

Ces extensions sont utiles pour que le vocabulaire recouvre bien ce qui paraît raisonnable (il paraît raisonnable que la fonction partie entière soit continue par morceaux sur $\mathbf{R}$ bien qu'elle ait un nombre infini de discontinuités).

Définition 25-4-181 : Soit I un intervalle (ni vide, ni réduit à un point). On dira qu'une fonction f de I vers $\mathbf{R}$ est **continue par morceaux** sur I lorsque sa restriction à tout intervalle fermé borné inclus dans I est continue par morceaux.

Évidemment, on ne manquera pas de remarquer (c'est évident au vu des définitions) que cette définition ne modifie pas le sens de “continu par morceaux” lorsque I est fermé borné.

On définirait de même une fonction en escalier et une primitive par morceaux sur un intervalle quelconque. Je passe très vite car la notion ne contient guère de piège et je ne veux pas insister. On notera simplement que les résultats énoncés ci-dessus restent vrais sur des intervalles quelconques ; la preuve en est laissée en exercice (facile, mais d'un style assez inhabituel).

5 - La notation intégrale

Définition 25-5-182 : Soit a et b deux réels et f une fonction réelle d'une variable réelle, continue par morceaux sur un intervalle contenant a et b.

On appelle l'**intégrale** de f entre a et b le réel $F(b) - F(a)$ calculé à l'aide d'une primitive par morceaux de f.

Notation 25-5-70 : Ce réel est noté, comme tout le monde le sait bien, $\displaystyle\int_a^b f(t)\, dt$.

On notera aussitôt que cette définition a un sens, d'une part parce que les primitives par morceaux de f existent et d'autre part parce qu'elles diffèrent d'une constante, ce qui garantit que le résultat ne dépend pas de la primitive utilisée pour le calcul.

Les résultats suivants sont évidents avec ce choix de définition :

Proposition 25-5-125 : Soit a, b et c trois réels, et f une fonction réelle d'une variable réelle, continue par morceaux sur un intervalle contenant a, b et c. Alors :

$$\int_a^c f(t)\, dt = \int_a^b f(t)\, dt + \int_b^c f(t)\, dt.$$

Démonstration : C'est stupide : c'est simplement dire que $F(c) - F(a) = (F(c) - F(b)) + (F(b) - F(a))$. •

Proposition 25-5-126 : Soit f une fonction continue à valeurs réelles définie sur un intervalle I, et soit a un point de I. Alors la fonction $x \mapsto \int_a^x f(t)\,dt$ est une fonction dérivable sur I, dont la dérivée est f.

Démonstration : Cela découle de la définition, et de l'information supplémentaire selon laquelle les fonctions continues ont mieux que des primitives par morceaux, à savoir de "vraies" primitives. •

Les énoncés suivants ne sont pas aussi grossièrement évidents, mais sont de simples reformulations des résultats énoncés sur les fonctions continues par morceaux.

Proposition 25-5-127 : Soit $a \leq b$ deux réels, et f une fonction à valeurs réelles positive continue par morceaux sur l'intervalle $[a, b]$. Alors l'intégrale $\int_a^b f(t)\,dt$ est positive.

Démonstration : C'est parce que toute primitive par morceaux de la fonction positive f est croissante. •

Proposition 25-5-128 : Soit $a \leq b$ deux réels et f une fonction à valeurs réelles continue par morceaux sur l'intervalle $[a, b]$. On suppose que pour tout t de $[a, b]$ (ou même sauf peut-être un nombre fini de t) on a l'inégalité : $f(t) \leq M$. Alors :

$$\int_a^b f(t)\,dt \leq M(b-a).$$

Démonstration : C'est simplement la réécriture de la proposition 25-2-124 avec une nouvelle notation. •

Remarques : * On peut aussi obtenir une variante de la précédente si on suppose f continue (et non seulement continue par morceaux) et en utilisant la vraie égalité des accroissements finis et non la plus modeste inégalité 25-2-124. Je laisse en exercice la détermination de l'énoncé que l'on obtient.
* Il ne faut pas oublier l'hypothèse $a \leq b$ dans les énoncés ci-dessus : si $b < a$ c'est à $[b, a]$ qu'on peut appliquer la croissance de F et l'inégalité se renverse ; de même pour l'inégalité des accroissements finis, la multiplication par un $b - a$ strictement négatif renverserait l'inégalité. Méfiance donc.

Proposition 25-5-129 : Soit $a \leq b$ deux réels et f une fonction à valeurs réelles continue par morceaux sur l'intervalle $[a, b]$. Alors :

$$\left|\int_a^b f(t)\,dt\right| \leq \int_a^b |f(t)|\,dt.$$

Démonstration : On remarque que sur $[a, b]$ la fonction $|f(t)| - f(t)$ est positive ainsi que la fonction $|f(t)| + f(t)$. On obtient donc les inégalités

$$0 \leq \int_a^b (|f(t)| - f(t))\;dt \qquad \text{et} \qquad 0 \leq \int_a^b (|f(t)| + f(t))\;dt$$

d'où

$$-\int_a^b |f(t)|\,dt \leq \int_a^b f(t)\,dt \leq \int_a^b |f(t)|\,dt$$

d'où l'inégalité annoncée. •

Remarque : On remarquera l'analogie entre cette inégalité et l'inégalité triangulaire : même pour la "somme" infinie qu'est l'intégrale, la valeur absolue de la somme est plus petite que la somme des valeurs absolues.

J'ai aussi traité en amphi l'inégalité de Schwarz et la formule de changement de variables.

Chapitre 26 - Fonctions vectorielles d'une variable réelle

Nous avons jusqu'à présent étudié des fonctions dont l'ensemble de départ et l'ensemble d'arrivée étaient des sous-ensembles de $\mathbf{R}$. L'étape suivante sera de faire apparaître plusieurs coordonnées. Les faire apparaître au départ est toute une aventure (c'est une partie significative du programme de deuxième année) ; les multiplier à l'arrivée est au contraire essentiellement facile. Ce chapitre suffira à vous faire (presque) tout savoir sur cette problématique.

1 - Ce qu'on peut définir

On va étudier des applications f définies sur une partie $\mathcal{D}_f$ de $\mathbf{R}$ et à valeurs dans un espace vectoriel de dimension finie sur $\mathbf{R}$.

En fait on peut traiter la question à deux niveaux de complexité. Ce que nous ferons cette année, c'est utiliser une base de l'espace d'arrivée et se ramener à l'étude des coordonnées du point mobile $f(t)$. Ce n'est pas très satisfaisant pour l'esprit, car dans un modèle physique il arrive souvent qu'il n'y ait pas un repère spécialement privilégié dans l'espace où évolue le mobile, et faire le calcul dans un repère plutôt qu'un autre paraît bien arbitraire. Un autre inconvénient plus sérieux est que la généralisation à un espace d'arrivée de dimension infinie n'est guère possible par cette méthode, alors pourtant qu'en s'y prenant autrement on y arrive sans trop de mal. Le "autrement" nécessite toutefois de parler de "normes" quelconques sur un espace vectoriel réel – ce n'est pas très compliqué mais ça prend du temps– et cela attendra donc encore un peu...

Passons tout de suite à un exemple de ce qu'on peut faire avec des méthodes de calcul sur des coordonnées.

Définition 26-1-183 : Soit F un espace vectoriel réel de dimension finie, $\mathcal{D}_f$ une partie de $\mathbf{R}$ et f une application de $\mathcal{D}_f$ vers F. Soit a un point de $\mathcal{D}_f$ (adhérent à $\mathcal{D}_f \setminus \{a\}$). Soit $\underline{e} = (e_1, \ldots, e_n)$ une base de F ; pour chaque i $(1 \leq i \leq n)$ notons $f_i : \mathcal{D}_f \to \mathbf{R}$ l'application qui à un réel $t \in \mathcal{D}_f$ associe la i-ème coordonnée du vecteur $f(t)$ dans la base $\underline{e}$.

On dira que l'application f est **continue** au point a lorsque toutes les f_i $(1 \leq i \leq n)$ le sont.

Remarque : C'est très lourd à écrire (inconvénient de l'usage des bases) mais surtout, il n'est pas clair que c'est cohérent : il se pourrait en effet que pour une base $\underline{e}$ l'application f apparaisse comme continue, mais que dans une autre base $\underline{\epsilon}$ elle apparaisse discontinue... Ce n'est pas le cas mais cela demande une ennuyeuse vérification spécifique.

Vérification : Soit $\underline{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$ une autre base de F. Notons $\varphi_i(t)$ les coordonnées de $f(t)$ dans cette nouvelle base. Notons $P = (p_{ij})$ la matrice de passage de $\underline{e}$ à $\underline{\epsilon}$ et $Q = (q_{ij})$ la matrice (inverse de P) de passage de $\underline{\epsilon}$ à $\underline{e}$.

Par les formules de changement de base, pour tout $t \in \mathcal{D}_f$, on a :

$$\begin{pmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{pmatrix} = P \begin{pmatrix} \varphi_1(t) \\ \vdots \\ \varphi_n(t) \end{pmatrix} \qquad \text{et} \qquad \begin{pmatrix} \varphi_1(t) \\ \vdots \\ \varphi_n(t) \end{pmatrix} = Q \begin{pmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{pmatrix}.$$

Supposons toutes les fonctions f_j $(1 \leq j \leq n)$ continues au point a. De la formule, valable pour tout $t \in \mathcal{D}_f$ et tout i $(1 \leq i \leq n)$:

$$\varphi_i(t) = \sum_{j=1}^{n} q_{ij} f_j(t)$$

on déduit aussitôt que toutes les φ_i $(1 \leq i \leq n)$ sont également continues en a. Réciproquement, en utilisant P on voit que la continuité des φ_i entraîne celle des f_j.

La définition proposée est donc cohérente.

Pour bien comprendre la nécessité de cette vérification, un exemple "dans le mauvais sens", à savoir d'un concept qui a un sens pour les fonctions à valeurs réelles mais qu'il faut se résoudre à abandonner pour les fonctions à valeurs vectorielles.

Essai de définition voué à l'échec : Soit F un espace vectoriel réel de dimension finie, $\mathcal{D}_f$ une partie de $\mathbf{R}$ et f une application de $\mathcal{D}_f$ vers F. Soit $\underline{e} = (e_1, \ldots, e_n)$ une base de F ; pour chaque i $(1 \leq i \leq n)$ notons $f_i : \mathcal{D}_f \to \mathbf{R}$ l'application qui à un réel $t \in \mathcal{D}_f$ associe la i-ème coordonnée du vecteur $f(t)$ dans la base $\underline{e}$.

On **ne dira pas** que l'application f est **minorée** lorsque toutes les f_i $(1 \leq i \leq n)$ le sont.

Explication de l'échec : cette définition dépend de la base choisie... Un exemple permet de s'en convaincre. Soit $f : \mathbf{R} \to \mathbf{R}^2$ définie par $f(t) = (e^{-t}, e^t)$ (tracer la courbe de f sans calculs sera un bon exercice...). Utilisons pour $\underline{e}$ la base canonique ; avec ce choix les coordonnées de $f(t)$ sont simplement $f_1(t) = e^{-t}$ et $f_2(t) = e^t$ et sont toutes deux minorées par 0. Recommençons après avoir fait tourner la base d'un huitième de tour (je mets un coefficient un peu compliqué pour que les calculs ne soient pas trop stupides) ; en clair mettons nous dans la base $\underline{\epsilon} = (\epsilon_1, \epsilon_2)$ où $\epsilon_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ et $\epsilon_2 = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. La matrice de passage P de $\underline{e}$ à $\underline{\epsilon}$ est $\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$ donc son inverse, matrice de passage de $\underline{\epsilon}$ à $\underline{e}$ est la matrice $Q = P^{-1} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$.
Si on note $\varphi_1(t)$ et $\varphi_2(t)$ les coordonnées de $f(t)$ dans la nouvelle base $\underline{\epsilon}$, on en déduit aussitôt que $\varphi_1(t) = \frac{1}{\sqrt{2}}(e^{-t} + e^t) = \sqrt{2}\,\mathrm{ch}\, t$ (qui est bien minoré, pas de problème) mais que $\varphi_2(t) = \frac{1}{\sqrt{2}}(-e^{-t} + e^t) = \sqrt{2}\,\mathrm{sh}\, t$ prend toutes valeurs réelles et n'est donc nullement minoré. Le concept dépend donc de la base où on se trouve. Ce n'est pas une bonne notion.

Il ne me reste plus qu'à tenter d'énumérer en espérant ne pas en oublier trop toutes les notions qui passent bien aux fonctions à valeurs vectorielles... Il faudrait pour chacune des définitions qui suit vérifier soigneusement qu'elle ne dépend pas de la base utilisée, ce qui est à chaque fois infiniment facile.

Dans toutes les définitions qui suivent, F désigne un espace vectoriel réel de dimension finie, $\mathcal{D}_f$ une partie de $\mathbf{R}$ et f une application de $\mathcal{D}_f$ vers F ; $\underline{e} = (e_1, \ldots, e_n)$ est une base de F ; pour chaque i $(1 \leq i \leq n)$ on note $f_i : \mathcal{D}_f \to \mathbf{R}$ l'application qui à un réel $t \in \mathcal{D}_f$ associe la i-ème coordonnée du vecteur $f(t)$ dans la base $\underline{e}$.

Définition 26-1-184 : Soit a un réel adhérent à $\mathcal{D}_f$ et v un vecteur de F ; notons v_i la i-ème coordonnée de v dans $\underline{E}$. On dit que $f(t)$ **tend vers** v quand t tend vers a lorsque pour chaque i $(1 \leq i \leq n)$, $f_i(t)$ tend vers v_i quand t tend vers a.

Définition 26-1-185 : Supposons $\mathcal{D}_f$ non majoré. Soit v un vecteur de F ; notons v_i la i-ème coordonnée de v dans $\underline{E}$. On dit que $f(t)$ **tend vers** v quand t tend vers $+\infty$ lorsque pour chaque i $(1 \leq i \leq n)$, $f_i(t)$ tend vers v_i quand t tend vers $+\infty$.

Définition 26-1-186 : Soit a un réel adhérent à $\mathcal{D}_f \setminus \{a\}$ et v un vecteur de F ; notons v_i la i-ème coordonnée de v dans $\underline{E}$. On dit que $f(t)$ **tend vers** v quand t tend vers a $(t \neq a)$ lorsque pour chaque i $(1 \leq i \leq n)$, $f_i(t)$ tend vers v_i quand t tend vers a $(t \neq a)$.

Définition 26-1-187 : Soit a un réel adhérent à $\mathcal{D}_f \cap]a, +\infty[$ et v un vecteur de F ; notons v_i la i-ème coordonnée de v dans $\underline{E}$. On dit que $f(t)$ **tend vers** v quand t tend vers a **à droite** lorsque pour chaque i $(1 \leq i \leq n)$, $f_i(t)$ tend vers v_i quand t tend vers a à droite.

Définition 26-1-188 : Dans chacun des cas énumérés ci-dessus, v est appelé la **limite** de $f(t)$.

Définition 26-1-189 : Soit a un réel adhérent à $\mathcal{D}_f \cap]a, +\infty[$. On dit que f est **continue à droite** en a lorsque chaque f_i $(1 \leq i \leq n)$ est continue à droite en a.

Définition 26-1-190 : On dit que f est **continue** sur $\mathcal{D}_f$ lorsque toutes les f_i $(1 \leq i \leq n)$ le sont.

Définition 26-1-191 : Pour I intervalle ouvert inclus dans $\mathcal{D}_f$, on dit que f est **continue** sur I lorsque toutes les f_i $(1 \leq i \leq n)$ le sont.

Définition 26-1-192 : Soit a un réel adhérent à $\mathcal{D}_f \setminus \{a\}$. On dit que f est **dérivable** en a lorsque toutes les f_i $(1 \leq i \leq n)$ le sont.

Définition 26-1-193 : Soit a un réel adhérent à $\mathcal{D}_f \cap]a, +\infty[$. On dit que f est **dérivable à droite** en a lorsque toutes les f_i $(1 \leq i \leq n)$ le sont.

Définition 26-1-194 : Soit a un point où f est dérivable. Le vecteur dont les coordonnées dans $\underline{e}$ sont les réels $f_i'(a)$ est appelé **vecteur dérivé** de f. L'application (définie sur l'ensemble $\mathcal{E}$ des points de $\mathcal{D}_f$ où f est dérivable) à valeurs dans F qui à tout t associe le vecteur dérivé de f en t est appelée la **dérivée** de f (et notée f').

Définition 26-1-195 : Soit a un point où f est dérivable à droite. Le vecteur dont les coordonnées dans $\underline{e}$ sont les réels $f_i'(a)$ est appelé **vecteur dérivé à droite** de f. L'application (définie sur l'ensemble $\mathcal{E}$ des points de $\mathcal{D}_f$ où f est dérivable) à valeurs dans F qui à tout t associe le vecteur dérivé de f en t est appelée la **dérivée à droite** de f (et notée f_d', ou f_+').
Définition 26-1-196 : On dit que f est **dérivable** sur $\mathcal{D}_f$ lorsque toutes les f_i $(1 \leq i \leq n)$ le sont.
Définition 26-1-197 : Soit I un intervalle ouvert inclus dans $\mathcal{D}_f$. On dit que f est **dérivable** sur I lorsque toutes les f_i $(1 \leq i \leq n)$ le sont.
Définition 26-1-198 : Soit c un point de $\mathcal{D}_f$. On dit que c est un **point critique** pour f, mais aussi que f **admet un point singulier** au point c (le mot "point singulier" désignant la valeur $f(c)$) lorsque f est dérivable en c et $f'(c) = 0$.
Définition 26-1-199 : Soit $n \geq 2$ un entier et a un point de $\mathcal{D}_f$. On dit que f est n fois dérivable en a lorsque toutes les f_i $(1 \leq i \leq n)$ le sont.
Définition 26-1-200 : Soit $n \geq 2$ un entier et a un point de $\mathcal{D}_f$ en lequel f est n fois dérivable. On appelle **vecteur dérivée n-ème** de f en a (et on note $f^{(n)}(a)$) le vecteur dont les coordonnées dans $\underline{e}$ sont les réels $f_i^{(n)}(a)$. L'application (définie sur l'ensemble $\mathcal{E}_n$ des points de $\mathcal{D}_f$ où f est n fois dérivable) à valeurs dans F qui à tout t associe le vecteur dérivé n-ème de f en t est appelée la **dérivée n-ème** de f (et notée $f^{(n)}$).
Définition 26-1-201 : On dit que f est n **fois dérivable**, ou **de classe $\mathcal{C}^n$**, ou **de classe $\mathcal{C}^\infty$** sur $\mathcal{D}_f$ (ou sur un intervalle ouvert inclus dans $\mathcal{D}_f$) lorsque toutes les f_i $(1 \leq i \leq n)$ le sont.
Définition 26-1-202 : On dit que f est **bornée** lorsque toutes les f_i $(1 \leq i \leq n)$ le sont.
Définition 26-1-203 : Une suite de vecteurs $(v_k)_{k \in \mathbf{N}}$ est dite **de Cauchy** lorsque les n suites obtenues en prenant les i-èmes coordonnées de tous les v_k sont de Cauchy.
Définition 26-1-204 : Soit a un réel adhérent à $\mathcal{D}_f$. On dit que f **vérifie le critère de Cauchy** en a lorsque toutes les f_i $(1 \leq i \leq n)$ le vérifient.
Définition 26-1-205 : Supposons $\mathcal{D}_f$ non majoré. On dit que f **vérifie le critère de Cauchy** en $+\infty$ lorsque toutes les f_i $(1 \leq i \leq n)$ le vérifient.
Définition 26-1-206 : On dit que f est **continue par morceaux** lorsque toutes les f_i $(1 \leq i \leq n)$ le sont. (On pourrait aussi définir les fonctions en escalier, mais je ne vois pas bien quand on s'en servirait...)
Définition 26-1-207 : Soit G une autre fonction définie sur $\mathcal{D}_f = [a, b]$ (avec $a < b$) et notons G_i les coordonnées de G dans $\underline{e}$. On dit que G est une **primitive** de f, supposée continue, ou une **primitive par morceaux** de f, supposée continue par morceaux, quand chaque G_i $(1 \leq i \leq n)$ est une primitive de f_i.
Définition 26-1-208 : Soit a et b deux réels tels que f soit continue par morceaux sur un intervalle contenant a et b. L'**intégrale** de f entre a et b est le vecteur dont les coordonnées dans $\underline{e}$ sont les réels $\displaystyle\int_a^b f_i(t)\,dt$ $(1 \leq i \leq n)$. On la note $\displaystyle\int_a^b f(t)\,dt$.

Je devrais maintenant énumérer des théorèmes et des propositions, par exemple le fait que la dérivabilité entraîne la continuité et beaucoup d'autres... Le principe est simple : quand les résultats sont vrais, ils sont évidents (travailler coordonnée par coordonnée). Et je ne chercherai donc pas à les énumérer. J'insisterai au contraire un peu plus loin sur ce qui doit être légèrement modifié.

2 - Produit scalaire usuel sur $\mathbf{R}^n$

La section donne quelques définitions fort simples sur $\mathbf{R}^n$, que vous manipulez forcément en physique ou en mécanique. Il est à noter que vous apprendrez dès l'an prochain comment les généraliser à des espaces vectoriels de dimension finie réels quelconques (ce à quoi je ne m'essaierai pas cette année).
Définition 26-2-209 : Le **produit scalaire** ("produit scalaire canonique" si on redoute des confusions) de deux vecteurs $(x_1, \ldots, x_n)$ et $(y_1, \ldots, y_n)$ de $\mathbf{R}^n$ est le réel $x_1y_1 + \cdots + x_ny_n$.
Notation 26-2-71 : Le produit scalaire de x par y est noté $x \cdot y$, ou $\langle x, y \rangle$.
Définition 26-2-210 : La **norme** (ou "norme euclidienne canonique" si on redoute des confusions) d'un vecteur $(x_1, \ldots, x_n)$ est le réel $\sqrt{\langle x, x \rangle}$, soit $\sqrt{x_1^2 + \cdots + x_n^2}$.
Notation 26-2-72 : La norme euclidienne de x est notée $\|x\|$ (ou, si on redoute des confusions avec d'autres normes, $\|x\|_2$).

Proposition 26-2-130 : Pour tous vecteurs $x = (x_1, \ldots, x_n)$ et $y = (y_1, \ldots, y_n)$ de $\mathbf{R}^n$,

$$|\langle x, y\rangle| \leq \|x\|\|y\| \qquad \text{("inégalité de Cauchy-Schwarz")}$$
$$\|x+y\| \leq \|x\| + \|y\| \qquad \text{("inégalité triangulaire")}$$

Démonstration : Elle est assez courte, assez astucieuse, et franchement hors sujet ici. Voir votre cours de l'an prochain.

•

Ce qui est fort utile pour manipuler des fonctions à valeurs vectorielles (à valeurs dans $\mathbf{R}^n$ seulement, faute de connaître des produits scalaires sur tous les espaces) est que le produit scalaire se dérive comme le produit ordinaire (c'est un énoncé évident à montrer, mais très pratique à connaître pour faire des calculs synthétiques) :

Proposition 26-2-131 : Soit f et g deux fonctions d'une même partie $\mathcal{D}$ de $\mathbf{R}$ vers $\mathbf{R}^n$. Soit a un point de $\mathcal{D}$ (supposé adhérent à $\mathcal{D} \setminus \{a\}$). Si f et g sont toutes deux dérivables en a, alors $\langle f, g\rangle$ l'est aussi, et

$$\langle f, g\rangle'(a) = \langle f'(a), g(a)\rangle + \langle f(a), g'(a)\rangle$$

Démonstration : Simple calcul idiot en revenant aux coordonnées.

•

3 - Accroissements finis : attention, pas d'égalité, seulement une inégalité !

Ce qui marche bien avec des fonctions à valeurs réelles mais échoue tristement avec des fonctions à valeurs vectorielles, ce sont les théorèmes en "il existe c" : Rolle, égalité des accroissements finis, égalité de Taylor-Lagrange. Pour le premier, rien ne subsiste –si je sais qu'une trajectoire tracée dans un plan revient à son point de départ, cela ne m'enseigne rien sur la vitesse du mobile. Malgré ce revers, des variantes des accroissements finis (et aussi de Taylor-Lagrange, mais je n'en parlerai pas) existent ; simplement il faut renoncer à des théorèmes en "il existe c" et se contenter d'inégalités, comme on en a d'ailleurs déjà pris l'habitude en intégrant des fonctions continues par morceaux.

Théorème 26-3-55 : Soit f une fonction définie sur le segment $[a, b]$ (avec $a \neq b$) et à valeurs dans $\mathbf{R}^n$ supposée continue sur le segment $[a, b]$ et dérivable sur le segment ouvert $]a, b[$. Soit M une constante réelle telle que pour tout t du segment $]a, b[$ on ait $\|f'(t)\| \leq M$.

Alors $\left\|\dfrac{f(b) - f(a)}{b - a}\right\| \leq M$.

Démonstration : Si $f(b) = f(a)$, c'est évident ; sinon introduisons la fonction g à valeurs réelles définie sur le segment $[a, b]$ par $g(t) = \langle f(t), f(b) - f(a)\rangle$. La fonction g est continue sur le segment fermé, dérivable sur le segment ouvert de dérivée $g'(t) = \langle f'(t), f(b) - f(a)\rangle$. En appliquant Cauchy-Schwarz,

$$|g'(t)| = |\langle f'(t), f(b) - f(a)\rangle| \leq \|f'(t)\|\|f(b) - f(a)\| \leq M\|f(b) - f(a)\|.$$

Appliquons l'égalité des accroissements finis à la fonction à valeurs réelles g ; on obtient un c dans le segment ouvert $]a, b[$ tel que $g(b) - g(a) = g'(c)(b - a)$.

Mais $g(b) - g(a) = \langle f(b), f(b) - f(a)\rangle - \langle f(a), f(b) - f(a)\rangle = \langle f(b) - f(a), f(b) - f(a)\rangle = \|f(b) - f(a)\|^2$ donc

$$\|f(b) - f(a)\|^2 = |g'(c)||b - a| \leq M\|f(b) - f(a)\||b - a|.$$

En divisant par $\|f(b) - f(a)\||b - a|$ on obtient le résultat annoncé.

•

Remarque : Il faut faire attention à ce qu'il n'existe **pas** de proposition analogue avec une minoration des dérivées : si je ne vais pas trop vite, je suis sûr de ne pas arriver très loin, en revanche, même si je ne laisse pas ma vitesse fléchir, pour peu que je tourne un peu en rond, je peux me retrouver à mon point de départ.

On démontrerait de même la variante suivante (en jouant sur une primitive par morceaux de f et f au lieu de jouer sur f et f')

Proposition 26-3-132 : Soit $a \leq b$ deux réels et f une fonction continue par morceaux sur l'intervalle $[a, b]$ et à valeurs dans $\mathbf{R}^n$. Soit M une constante réelle telle que pour tout t de $[a, b]$ (ou même sauf un nombre fini de tels t) on ait $\|f'(t)\| \leq M$. Alors :

$$\left\|\int_a^b f(t)\,dt\right\| \leq M(b - a).$$

Démonstration : C'est une conséquence immédiate de la proposition qui suit.

•

Proposition 26-3-133 : Soit $a \leq b$ deux réels et f une fonction continue par morceaux sur l'intervalle $[a, b]$ et à valeurs dans $\mathbf{R}^n$.
Alors

$$\left\| \int_a^b f(t)\, dt \right\| \leq \int_a^b \|f(t)\| \, dt.$$

Démonstration : Le principe de la preuve est le même que pour le théorème de la page précédente : si $\int_a^b f(t)\, dt = 0$, c'est évident ; sinon on introduit (pour $s \in [a, b]$) la primitive par morceaux de f qu'est

$$F(s) = \int_a^s f(t)\, dt$$

(on notera que $F(a) = 0$ et que $F(b) = \int_a^b f(t)\, dt$) puis la fonction auxiliaire

$$\varphi(s) = \langle F(s), F(b) \rangle \, .$$

On constate alors que φ est continue, dérivable sauf peut-être en un nombre fini de points et qu'en les points où f est continue, $\varphi'(s) = \langle F'(s), F(b) \rangle = \langle f(s), F(b) \rangle$: φ est donc une primitive par morceaux de la fonction $s \mapsto \langle f(s), F(b) \rangle$, d'où

$$\int_a^b \langle f(s), F(b) \rangle \; ds = \varphi(b) - \varphi(a) = \langle F(b), F(b) \rangle - \langle F(a), F(b) \rangle = \langle F(b), F(b) \rangle = \|F(b)\|^2.$$

Donc

$$\begin{aligned} \|F(b)\|^2 = \int_a^b \langle f(s), F(b) \rangle \; ds &\leq \int_a^b |\langle f(s), F(b) \rangle| \; ds \\ &\leq \int_a^b \|f(s)\| \, \|F(b)\| \; ds \\ &= \|F(b)\| \int_a^b \|f(s)\| \; ds \end{aligned}$$

.

Il ne reste plus qu'à diviser les deux membres par $\|F(b)\|$ et remplacer le $\|F(b)\|$ restant par sa valeur $\int_a^b f(t)\, dt$ pour conclure.

•

Chapitre 27 - Déterminants

Comme sur tous les objets mathématiques importants, le déterminant a plusieurs interprétations possibles, et sa théorie peut être présentée de diverses façons.

L'idée qu'il me semble la première à connaître est que le déterminant est lié aux volumes : le déterminant d'une application linéaire u de $\mathbf{R}^n$ vers $\mathbf{R}^n$ (ou plus exactement sa valeur absolue) est la quantité par laquelle u multiplie les volumes.

La théorie sera plus facile à écrire pour des matrices carrées, on passera aux applications linéaires dans la dernière ligne droite.

1 - Matrices-transvections

Tous les calculs que nous allons voir exécuter sur les déterminants sont basés sur des manipulations simples sur les lignes et les colonnes de matrices, celles même qu'on a déjà vu en usage pour résoudre les systèmes par la méthode du pivot.

Formaliser un peu ces transformations se révèlera donc rentable.

Notation 27-1-73 : Pour $1 \leq i \leq n$ et $1 \leq j \leq n$ avec $i \neq j$ et λ un scalaire, on note $M_{ij}(\lambda) = I + \lambda E_{ij}$ (où on rappelle que la notation E_{ij} désigne la matrice élémentaire pour l'emplacement (i, j)).

Définition 27-1-211 : Ces matrices $M_{ij}(\lambda)$ (pour $\lambda \neq 0$) seront appelées **matrices-transvections**.

L'utilité de ces matrices-transvections est qu'elles formalisent les bidouillages qu'on sait pratiquer sans leur aide, mais qu'on aurait du mal à utiliser dans des preuves vérifiables sans cette nouvelle notation auxiliaire.

Affirmation : pour toute matrice carrée $A \in \mathcal{M}_n(\mathbf{K})$, la matrice $B = A \times M_{ij}(\lambda)$ s'obtient à partir de A de la façon suivante : on note C_i et C_j les i-ème et j-ème colonnes de A ; les colonnes de B sont identiques à celles de A sauf la j-ème qui vaut $C_j + \lambda C_i$; la matrice $B_1 = M_{ij}(\lambda) \times A$ s'obtient à partir de A de la façon suivante : on note L_i et L_j les i-ème et j-ème lignes de A ; les lignes de B_1 sont identiques à celles de A sauf la i-ème qui vaut $L_i + \lambda L_j$.

Vérification : C'est la simple application de la définition du produit de deux matrices, et l'interprétation de la phrase française qui précède...

On utilisera aussi plus occasionnellement une matrice moins technique que $M_{ij}(\lambda)$.

Notation 27-1-74 : Pour $1 \leq i \leq n$, on note $D_i(\lambda)$ et λ un scalaire, on note $D_i(\lambda)$ la matrice diagonale dont tous les termes diagonaux valent 1 sauf le (i, i)-ème qui vaut λ (si on préfère les formules, on écrira : $D_i(\lambda) = I - E_{ii} + \lambda E_{ii}$).

Affirmation : pour toute matrice carrée $A \in \mathcal{M}_n(\mathbf{K})$, la matrice $B = A \times D_i(\lambda)$ s'obtient à partir de A de la façon suivante : on note C_i la i-ème colonne de A ; les colonnes de B sont identiques à celles de A sauf la i-ème qui vaut λC_i ;la matrice $B_1 = D_i(\lambda) \times A$ s'obtient à partir de A de la façon suivante : on note L_i la i-ème ligne de A ; les lignes de B_1 sont identiques à celles de A sauf la i-ème qui vaut λL_i.

Vérification : C'est encore la simple application de la définition de la multiplication matricielle.

Les lemmes suivants concernant les matrices-transvections nous seront utiles.

Lemme 27-1-14 : Pour $1 \leq i \leq n$ et $1 \leq j \leq n$ avec $i \neq j$ et λ et μ deux scalaires.

$$M_{ij}(\lambda) \times M_{ij}(\mu) = M_{ij}(\lambda + \mu).$$

Démonstration : Remarquons (simple calcul...) que $E_{ij}^2 = 0$. On en déduit que :

$$M_{ij}(\lambda) \times M_{ij}(\mu) = (I + \lambda E_{ij})(I + \mu E_{ij}) = I + (\lambda + \mu)E_{ij} + 0 = M_{ij}(\lambda + \mu).$$

•

Corollaire 27-1-7 : Toute matrice-transvection est inversible, et son inverse est une matrice-transvection.

Démonstration : L'inverse de $M_{ij}(\lambda)$ est $M_{ij}(-\lambda)$.

•

Lemme 27-1-15 : Pour $1 \leq i \leq n$ et $1 \leq j \leq n$ avec $i \neq j$ et λ et μ deux scalaires non nuls, $M_{ij}(\lambda)$ et $M_{ij}(\mu)$ sont semblables

Démonstration : Simple vérification (pénible...). Le lecteur méticuleux vérifiera qu'en posant $P = D_i(\lambda/\mu)$, inversible d'inverse $P^{-1} = D_i(\mu/\lambda)$, on a bien $P^{-1}M_{ij}(\lambda)P = M_{ij}(\mu)$.
•

Nous allons maintenant voir que toute matrice peut être ramenée à la forme diagonale par des opérations élémentaires sur les lignes et les colonnes. Pour les besoins des preuves qui suivent, mélanger des bidouillages sur les lignes et les colonnes n'est pas gênant, et l'énoncé obtenu est par voie de conséquence très simple.
Lemme 27-1-16 : Soit $n \geq 0$ un entier. Pour toute matrice carrée (n,n) A, il existe une matrice diagonale D et des matrices-transvections $S_1, \ldots, S_k$, $T_1, \ldots, T_l$ telles que

$$A = S_1 S_2 \cdots S_k D T_l \cdots T_2 T_1.$$

Démonstration : C'est une récurrence sur n.
* Pour $n = 0$ (ou en commençant à $n = 1$ si on trouve les matrices vides trop effrayantes), c'est évident : la matrice A est directement diagonale.
* Soit $n \geq 2$ fixé, et supposons le théorème vrai pour toute matrice $(n-1, n-1)$. Soit A une matrice (n,n).
  * Premier cas : si la première ligne et la première colonne de A sont toutes deux nulles, à l'exception possible du coefficient a_{11}.
  Dans ce cas, A s'écrit par blocs :

$$A = \left(\begin{array}{c|c} a_{11} & 0 \\ \hline 0 & B \end{array}\right)$$

  où B est une matrice carrée $(n-1, n-1)$. On peut appliquer l'hypothèse de récurrence à B et écrire

$$B = S'_1 S'_2 \cdots S'_k D' T'_l \cdots T'_2 T'_1$$

  pour une D' diagonale et des S' et T' matrices-transvections. Prolongeons alors chaque S', chaque T' et D' en une matrice (n,n) en posant :

$$S_i = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & S'_i \end{array}\right) \quad D = \left(\begin{array}{c|c} a_{11} & 0 \\ \hline 0 & D' \end{array}\right) \quad T_i = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & T'_i \end{array}\right).$$

  Les matrices ainsi construites sont respectivement des matrices-transvections et une matrice diagonale, et le produit $B = S_1 S_2 \cdots S_k D T_l \cdots T_2 T_1$ vaut bien A (simple calcul par blocs).
  * Second cas : si certains a_{i1} $(1 \leq i \leq n)$ ou certains a_{1j} ne sont pas nuls.
    * Premier sous-cas : si $a_{11} \neq 0$. Dans ce sous-cas, en ajoutant à chaque colonne un multiple approprié de la première colonne, on peut tuer tous les a_{1j} (par exemple pour tuer a_{12}, on ajoutera à la deuxième colonne la première multipliée par $-a_{12}/a_{11}$). De la même façon, par des opérations sur les lignes, on pourra tuer tous les a_{i1}. En termes matriciels, la matrice

$$A' = M_{n1}(-\frac{a_{n1}}{a_{11}}) \cdots M_{31}(-\frac{a_{31}}{a_{11}}) M_{21}(-\frac{a_{21}}{a_{11}})\, A\, M_{12}(-\frac{a_{12}}{a_{11}}) M_{13}(-\frac{a_{13}}{a_{11}}) \cdots M_{1n}(-\frac{a_{1n}}{a_{11}})$$

    est de la forme traitée au premier cas ; on peut donc la décomposer en produit de matrices-transvections et de matrice diagonale ; puis en écrivant que

$$A = M_{21}(\frac{a_{21}}{a_{11}}) M_{31}(\frac{a_{31}}{a_{11}}) \cdots M_{n1}(\frac{a_{n1}}{a_{11}})\, A'\, M_{1n}(\frac{a_{1n}}{a_{11}}) \cdots M_{13}(\frac{a_{13}}{a_{11}}) \cdots M_{12}(\frac{a_{12}}{a_{11}})$$

    on obtient bien une décomposition de A.
    * Deuxième sous-cas : si $a_{11} = 0$. Comme on a supposé un autre des coefficients de la première ligne (ou de la première colonne) disons a_{1j} non nul, il suffit d'ajouter préalablement cette j-ème colonne à la première pour être ramené au premier sous-cas.

Le lemme est donc vrai pour toutes les matrices (n,n), ce qui clôt la récurrence.
•

2 - La définition

Définition 27-2-212 : Soit $\mathbf{K}$ un corps commutatif et $n \geq 0$ un entier. On appelle **déterminant** toute application $f : \mathcal{M}_n(\mathbf{K}) \to \mathbf{K}$ vérifiant les deux propriétés suivantes :
(i) Pour toutes matrices $A, B \in \mathcal{M}_n(\mathbf{K})$, $f(AB) = f(A)f(B)$.
(ii) Pour toute matrice diagonale $D \in \mathcal{M}_n(\mathbf{K})$, $f(D)$ est le produit des termes diagonaux de D.

Avec si peu de connaissances, il est déjà possible de montrer un résultat très simple

Proposition 27-2-134 : Soit $\mathbf{K}$ un corps commutatif et $n \geq 0$ un entier ; soit det un déterminant sur $\mathcal{M}_n(\mathbf{K})$. Si A et B sont deux matrices semblables de $\mathcal{M}_n(\mathbf{K})$, $\det A = \det B$.

Démonstration : On notera que selon l'usage on omettra le plus souvent les parenthèses pour écrire $\det A$ au lieu de $\det(A)$.

Remarquons tout d'abord que comme I est diagonale, on sait calculer $\det I = 1$.

Soit alors P inversible telle que $B = P^{-1}AP$. Alors $\det B = \det P^{-1} {\cdot} \det A {\cdot} \det P = \det P^{-1} {\cdot} \det P {\cdot} \det A = \det P^{-1}P \cdot \det A = \det I \cdot \det A = \det A$.

Le gros morceau du chapitre sera de montrer le

Théorème 27-2-56 : Soit $\mathbf{K}$ un corps commutatif et $n \geq 0$ un entier. Il existe sur $\mathcal{M}_n(\mathbf{K})$ un et un seul déterminant.

La démonstration va reposer sur la possibilité prouvée à la section précédente d'écrire toute matrice à partir de matrices diagonales et de matrices-transvections. On sait déjà calculer le déterminant des matrices diagonales, par définition ; on saura très bientôt calculer celui des matrices-transvections. On saura donc calculer celui de toutes les matrices, ce qui prouvera l'unicité du déterminant. Pour l'existence, il faut sortir une formule de sa manche...

3 - Déterminant et matrices inversibles

Proposition 27-3-135 : Soit $\mathbf{K}$ un corps commutatif et $n \geq 0$ un entier. Soit det un déterminant sur $\mathcal{M}_n(\mathbf{K})$. Alors pour toute matrice carrée $A \in \mathcal{M}_n(\mathbf{K})$,

$$A \text{ est inversible} \iff \det A \neq 0.$$

Démonstration :
* Preuve de $\Rightarrow$. Supposons A inversible. On a $1 = \det I = \det A \cdot \det A^{-1}$ donc $\det A \neq 0$ (et accessoirement son inverse est $\det A^{-1}$).
* Preuve de $\Leftarrow$ (par contraposition). Supposons A non inversible. Soit $r < n$ son rang. Comme dans le chapitre "matrices", introduisons la matrice (n, n) J_r dont les coefficients a_{ij} sont définis par $a_{ii} = 1$ pour $1 \leq i \leq r$ et $a_{ij} = 0$ pour tous les autres coefficients. Remarquons que J_r est diagonale, donc on sait calculer son déterminant, et on trouve 0 (elle contient des termes nuls sur la diagonale, puisque $r < n$). Comme A est de même rang que J_r, A est équivalente à J_r ; introduisons des matrices inversibles P et Q telles que $A = Q^{-1}J_rP$. Alors $\det A = \det Q^{-1} \cdot \det J_r \cdot \det P = \det Q^{-1} \cdot 0 \cdot \det P = 0$.

•

4 - Déterminants des matrices-transvections

Proposition 27-4-136 : Soit $\mathbf{K}$ un corps commutatif et $n \geq 0$ un entier, et soit det un déterminant sur $\mathcal{M}_n(\mathbf{K})$. Alors pour tous $1 \leq i \leq n$ et $1 \leq j \leq n$ avec $i \neq j$ et tout scalaire λ, $\det M_{ij}(\lambda) = 1$.

Démonstration : Notons tout d'abord que si $\lambda = 0$, $M_{ij}(\lambda) = I$ et le résultat est évident, on pourra donc supposer $\lambda \neq 0$.

Écrivons l'identité issue du premier lemme :

$$(E) \qquad (M_{ij}(\lambda))^2 = M_{ij}(2\lambda)$$

et notons $d = \det M_{ij}(\lambda)$.
* Premier cas : on est dans un corps où $2 \neq 0$.

Dans ce cas, on a aussi $2\lambda \neq 0$, donc d'après le second lemme, la matrice $M_{ij}(2\lambda)$ est semblable à $M_{ij}(\lambda)$ et a donc le même déterminant.

En appliquant det aux deux expressions liées par l'égalité *(E)* on obtient donc :

$$d^2 = d$$

soit $d^2 - d = 0$, soit $d(d-1) = 0$ donc d vaut 0 ou 1.

Comme $M_{ij}(\lambda)$ est inversible, $d \neq 0$. D'où $d = 1$.

* Second cas : on est dans un corps où $2 = 0$.

Les choses sont plus troublantes mais plus faciles, car l'identité *(E)* s'écrit alors plus simplement

$$(M_{ij}(\lambda))^2 = I$$

donc, en appliquant det on obtient :

$$d^2 = 1$$

soit $d^2 - 1 = 0$, soit $(d-1)(d+1) = 0$, soit $d = 1$ ou $d = -1$. Mais comme $2 = 0$, $1 = -1$, et donc là encore $d = 1$.

•

Preuve de l'unicité dans le théorème 27-2-56

Soit $\det_1$ et $\det_2$ deux déterminants sur $\mathcal{M}_n(\mathbf{K})$. Soit A une matrice (n,n).

Écrivons $A = S_1 S_2 \cdots S_k D T_l \cdots T_2 T_1$ pour D diagonale, et les S_i et T_j matrices-transvections. Alors $\det_1 A = \det_1 S_1 \det_1 S_2 \cdots \det_1 S_k \det_1 D \det_1 T_l \cdots \det_1 T_2 \det_1 T_1 = 1 \cdots 1 \cdots \det_1 D \cdot 1 \cdots 1 = \det_1 D$ est égal au produit des termes diagonaux de D. Il en est de même avec $\det_2$. D'où l'égalité des deux applications $\det_1$ et $\det_2$.

•

Maintenant que nous savons que le déterminant, s'il existe, est unique, nous parlerons "du" déterminant et le noterons det. On ne perdra pas de vue que nous ne savons pas encore que det existe ; les prochains énoncés ne sont donnés que sous réserve d'existence (mais comme on le construira dans quelques pages, tout ira bien).

5 - Opérations sur les colonnes

Notation 27-5-75 : Pour $C_1, \ldots, C_n$ des matrices colonnes (chacune formée de n scalaires), on notera $\det(C_1, \ldots, C_n)$ pour $\det A$, où A est la matrice carrée (n,n) dont les colonnes successives sont $C_1, \ldots, C_n$.

Proposition 27-5-137 : Le déterminant ne change pas quand on ajoute à une colonne un multiple d'une (autre) colonne. Avec des formules, pour tout n, toutes matrices-colonnes $C_1, \ldots, C_n$ à coefficients dans un même corps commutatif, tous $1 \leq i \leq n$ et $1 \leq j \leq n$ avec $i \neq j$ et tout scalaire λ

$$\det(C_1, \ldots, C_n) = \det(C_1, \cdots, C_{j-1}, C_j + \lambda C_i, C_{j+1}, \cdots, C_n).$$

Démonstration : Notons A la matrice carrée ayant pour colonnes $C_1, \ldots, C_n$ et B celle ayant pour colonnes $C_1, \cdots, C_{j-1}, C_j + \lambda C_i, C_{j+1}, \cdots, C_n$.

D'après l'affirmation de la section précédente $B = A \times M_{ij}(\lambda)$. On en déduit que

$$\det B = \det A \cdot \det M_{ij}(\lambda) = \det A.$$

•

Proposition 27-5-138 : Le déterminant est multiplié par λ quand on multiplie une (seule) colonne par λ. Avec des formules : pour tout n, toutes matrices-colonnes $C_1, \ldots, C_n$ à coefficients dans un même corps commutatif, tout $1 \leq i \leq n$ et tout scalaire λ

$$\lambda \det(C_1, \ldots, C_n) = \det(C_1, \cdots, C_{i-1}, \lambda C_i, C_{i+1}, \cdots, C_n).$$

Démonstration : Notons A la matrice carrée ayant pour colonnes $C_1, \ldots, C_n$ et B celle ayant pour colonnes $C_1, \cdots, C_{i-1}, \lambda C_i, C_{i+1}, \cdots, C_n$.

D'après l'affirmation de la section précédente $B = A \times D_i(\lambda)$, donc $\det B = \det A \cdot \det D_i(\lambda) = \lambda \det A$.

•

Proposition 27-5-139 : Pour i fixé, le déterminant est linéaire par rapport à chaque colonne. Avec des formules : voir la proposition précédente pour la propriété multiplicative, et, par ailleurs, pour tout n, tout

i tel que $1 \leq i \leq n$, toutes matrices-colonnes $C_1, \ldots, C_{i-1}, C_{i+1}, C_n$, C'_i et C''_i à coefficients dans un même corps commutatif

$$\det(C_1, \ldots, C_{i-1}, C'_i + C''_i, C_{i+1}, C_n) = \det(C_1, \ldots, C_{i-1}, C'_i, C_{i+1}, C_n) + \det(C_1, \ldots, C_{i-1}, C''_i, C_{i+1}, C_n).$$

Démonstration : On distingue selon que le système de $n-1$ colonnes $(C_1, \ldots, C_{i-1}, C_{i+1}, C_n)$ est libre ou non.
* Premier cas : si $(C_1, \ldots, C_{i-1}, C_{i+1}, C_n)$ est lié.
Dans ce cas les trois matrices carrées ayant respectivement pour colonnes $(C_1, \ldots, C_{i-1}, C'_i + C''_i, C_{i+1}, C_n)$, $C_1, \ldots, C_{i-1}, C'_i, C_{i+1}, C_n)$ et $C_1, \ldots, C_{i-1}, C"'_i, C_{i+1}, C_n)$ ont chacune $n-1$ colonnes liées, donc *a fortiori* ont toutes leurs colonnes liées. Elles ne sont donc pas inveresibles, et leurs déterminants sont donc tous trois nuls, et la formule a prouver se réduit à $0 = 0 + 0$.
* Second cas : si $(C_1, \ldots, C_{i-1}, C_{i+1}, C_n)$ est libre.
Le théorème de la base incomplète permet alors de le compléter en une base $(C_1, \ldots, C_{i-1}, C_i, C_{i+1}, C_n)$ de l'espace $\mathcal{M}_{n1}$ des matrices-colonnes. Développons dans cette base les deux colonnes C'_i et C''_i, soit :

$$C'_i = \sum_{k=1}^{n} \lambda_k C_k \qquad \text{et} \qquad C'_i = \sum_{k=1}^{n} \mu_k C_k.$$

On a alors $\det(C_1, \ldots, C_{i-1}, C'_i, C_{i+1}, C_n) = \det(C_1, \ldots, C_{i-1}, \sum_{k=1}^{n} \lambda_k C_k, C_{i+1}, C_n)$.
Dans cette dernière expression, retranchons à la i-ème colonne $\lambda_1 C_1$, puis $\lambda_2 C_2, \ldots, \lambda_{i-1} C_{i-1}$, puis pour terminer $\lambda_{i+1} C_{i+1}, \ldots, \lambda_n C_n$.
Il reste $\det(C_1, \ldots, C_{i-1}, C'_i, C_{i+1}, C_n) = \det(C_1, \ldots, C_{i-1}, \lambda_i C_i, C_{i+1}, C_n) = \lambda_i \det(C_1, \ldots, C_n)$.
Le même calcul montre par ailleurs que $\det(C_1, \ldots, C_{i-1}, C''_i, C_{i+1}, C_n) = \mu_i \det(C_1, \ldots, C_n)$ et enfin que $\det(C_1, \ldots, C_{i-1}, C'_i + C''_i, C_{i+1}, C_n) = (\lambda_i + \mu_i) \det(C_1, \ldots, C_n)$.
L'égalité annoncée est donc prouvée.

•

Remarque : On ne confondra pas cette propriété de linéarité "colonne par colonne" (la "multilinéarité" si on veut faire savant) avec la linéarité ordinaire ! Le déterminant n'est pas du tout une application linéaire. Pour A et B deux matrices carrées (n, n) sur un même corps commutatif, l'expression $\det(A + B)$ ne s'arrange en général absolument PAS, tandis que pour λ scalaire, $\det(\lambda A) = \lambda^n \det A$ (on fait sortir λ successivement de chaque colonne).
Proposition 27-5-140 : Quand on échange deux colonnes dans une matrice carrée, le déterminant change de signe. Avec des formules : pour tout n, toutes matrices-colonnes $C_1, \ldots, C_n$ à coefficients dans un même corps commutatif, tous indices i, j avec $1 \leq i < j \leq n$,

$$\det(C_1, \ldots, C_n) = -\det(C_1, \cdots, C_{i-1}, C_j, C_{i+1}, \cdots, C_{j-1}, C_i, C_{j+1}, \ldots, C_n).$$

Démonstration : Calculons de deux façons le déterminant

$$\det(C_1, \cdots, C_{i-1}, C_i + C_j, C_{i+1}, \cdots, C_{j-1}, C_i + C_j, C_{j+1}, \ldots, C_n).$$

Dans un premier calcul, on constate que les colonnes numérotées i et j sont les mêmes, donc il s'agit du déterminant d'une matrice carrée non inversible, donc ce déterminant est nul.

Dans un deuxième calcul, on développe en utilisant les linéarités par rapport à la i-ème et par rapport à la j-ème colonne.

On obtient :

$$\det(C_1, \cdots, C_{i-1}, C_i, C_{i+1}, \cdots, C_{j-1}, C_i, C_{j+1}, \ldots, C_n) +$$
$$\det(C_1, \cdots, C_{i-1}, C_i, C_{i+1}, \cdots, C_{j-1}, C_j, C_{j+1}, \ldots, C_n) +$$
$$\det(C_1, \cdots, C_{i-1}, C_j, C_{i+1}, \cdots, C_{j-1}, C_i, C_{j+1}, \ldots, C_n) +$$

$$\det(C_1, \cdots, C_{i-1}, C_j, C_{i+1}, \cdots, C_{j-1}, C_j, C_{j+1}, \ldots, C_n).$$

Dans cette formule , les premier et quatrième déterminants sont tous deux nuls (encore la répétition de colonnes). On en déduit donc finalement que

$$0 = \det(C_1, \cdots, C_{i-1}, C_i, C_{i+1}, \cdots, C_{j-1}, C_j, C_{j+1}, \ldots, C_n) +$$
$$\det(C_1, \cdots, C_{i-1}, C_j, C_{i+1}, \cdots, C_{j-1}, C_i, C_{j+1}, \ldots, C_n).$$

•

6 - Développement d'un déterminant par rapport à la première ligne

Définition 27-6-213 : Soit A une matrice à coefficients dans un corps commutatif. On appelle **mineurs** de A les déterminants des sous-matrices carrées de A.
Définition 27-6-214 : Soit A une matrice carrée (n, n) à coefficients dans un corps commutatif et $1 \leq i \leq n$, $1 \leq j \leq n$. Le **mineur associé à** (i, j) est le déterminant de la matrice carrée $(n-1, n-1)$ obtenue par ablation de la i-ème ligne et de la j-ème colonne de A.
Définition 27-6-215 : Soit A une matrice carrée (n, n) à coefficients dans un corps commutatif et $1 \leq i \leq n$, $1 \leq j \leq n$. Le **cofacteur associé à** (i, j) est obtenu en multipliant par $(-1)^{i+j}$ le mineur associé à (i, j).
Définition 27-6-216 : Soit A une matrice carrée (n, n) à coefficients dans un corps commutatif. La **comatrice** de A est la matrice formée des cofacteurs de A.
Notation 27-6-76 : La comatrice de A sera notée $\operatorname{com} A$.
Lemme 27-6-17 : Soit $m \leq n$ deux entiers, soit B une matrice carrée $(n-m, n-m)$ à coefficients dans un corps commutatif $\mathbf{K}$ et soit la matrice (n, n) qui s'écrit par blocs :

$$A = \left(\begin{array}{c|c} I_m & 0 \\ \hline 0 & B \end{array}\right).$$

On a l'égalité $\det A = \det B$.
Démonstration : Soit f l'application définie sur $\mathcal{M}_{n-m}(\mathbf{K})$ par :

$$f(M) = \left|\begin{array}{c|c} I_m & 0 \\ \hline 0 & M \end{array}\right|.$$

Pour M et N deux matrices de $\mathcal{M}_{n-m}(\mathbf{K})$, par multiplication par blocs des matrices,

$$\left(\begin{array}{c|c} I_m & 0 \\ \hline 0 & M \end{array}\right) \left(\begin{array}{c|c} I_m & 0 \\ \hline 0 & N \end{array}\right) = \left(\begin{array}{c|c} I_m & 0 \\ \hline 0 & MN \end{array}\right)$$

donc, en prenant les déterminants des trois termes, $f(M)f(N) = f(MN)$.

Par ailleurs lorsque $M = D$ est diagonale, la matrice $\left(\begin{array}{c|c} I_m & 0 \\ \hline 0 & D \end{array}\right)$ est elle-même diagonale donc son déterminant est égal au produit de ses termes diagonaux. Ainsi $f(D)$ est égal au produit des termes diagonaux de D.
L'application f est donc le déterminant sur $\mathcal{M}_{n-m}(\mathbf{K})$.
En écrivant que $f(B) = \det B$ on obtient le résultat annoncé.

•

Lemme 27-6-18 : Soit B_1 et B_2 deux matrices ayant chacune $n-1$ lignes et ayant à elles deux $n-1$ colonnes à coefficients dans un même corps commutatif $\mathbf{K}$; notons B la matrice $(n-1, n-1)$ obtenue par juxtaposition côte à côte de B_1 et B_2 et soit la matrice (n, n) :

$$A = \left(\begin{array}{c|c|c} 0 \cdots 0 & 1 & 0 \cdots 0 \\ \hline & 0 & \\ B_1 & \vdots & B_2 \\ & 0 & \end{array}\right).$$

En notant k l'indice de la colonne commençant par le 1, $\det A = (-1)^{k+1} \det B$.
Démonstration : Si $k = 1$, c'est le lemme précédent (quand $m = 1$). La démonstration se prête bien à écrire une récurrence sur k ; soit donc un k fixé, supposons le résultat vrai pour une colonne intermédiaire en k-ème position, et montrons le quand la colonne intermédiaire est en $k+1$-ème position, c'est-à-dire quand B_1 est formée de k colonnes. Dans ce contexte, notons X_k la k-ème colonne de B_1 et notons B_1' la matrice $(n-1, k-1)$ formée des $k-1$ premières colonnes de B_1. Avec ces notations,

$$A = \left(\begin{array}{c|c|c|c} 0\cdots 0 & 0 & 1 & 0\cdots 0 \\ \hline & & 0 & \\ B_1' & X_k & \vdots & B_2 \\ & & 0 & \end{array}\right).$$

On sait que l'échange de deux colonnes change le signe du déterminant ; on obtient donc :

$$\det A = \left|\begin{array}{c|c|c|c} 0\cdots 0 & 0 & 1 & 0\cdots 0 \\ \hline & & 0 & \\ B_1' & X_k & \vdots & B_2 \\ & & 0 & \end{array}\right| = - \left|\begin{array}{c|c|c|c} 0\cdots 0 & 1 & 0 & 0\cdots 0 \\ \hline & 0 & & \\ B_1' & \vdots & X_k & B_2 \\ & 0 & & \end{array}\right|.$$

En utilisant l'hypothèse de récurrence, on obtient donc $\det A = -(-1)^k \det B = (-1)^{k+1} \det B$.

•

Théorème 27-6-57 : Soit $A = (a_{ij})$ une matrice carrée (n,n) à coefficients dans un corps commutatif. Notons m_{ij} le mineur de A associé à (i,j). Alors :

$$\det A = a_{11}m_{11} - a_{12}m_{12} + a_{13}m_{13} + \cdots + (-1)^{n+1}a_{1n}m_{1n}.$$

Démonstration : Notons C_i la i-ème colonne de A (pour $1 \leq i \leq n$). Notons ensuite X_i le vecteur-colonne à n lignes obtenu en remplaçant le premier coefficient de C_i par un zéro (c'est à dire $X_i = \begin{pmatrix} 0 \\ a_{2i} \\ \vdots \\ a_{ni} \end{pmatrix}$) et notons enfin Y le vecteur colonne $\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ (avec n coefficients).

Avec ces notations, pour chaque i $(1 \leq i \leq n)$, $C_i = a_{1i}Y + X_i$.

Le déterminant $\det A = \det(C_1, \ldots, C_n) = \det(a_{11}Y + X_1, \ldots, a_{1n}Y + X_n)$ peut être développé en utilisant successivement la linéarité par rapport à chaque colonne. L'expression complète est alors une gigantesque sommation de 2^n termes. Mais dans la plupart de ces termes, Y apparaît au moins deux fois dans le déterminant. Dès lors qu'il y a répétition de colonnes, la matrice carrée correspondante n'est pas inversible et son déterminant est nul. La sommation s'allège donc indiciblement et il ne reste que l'expression

$$\det A = \det(X_1, \ldots, X_n) + a_{11}\det(Y, X_2, \ldots, X_n) + a_{12}\det(X_1, Y, X_3 \ldots, X_n) + \cdots + a_{1n}\det(X_1, \ldots, X_{n-1}, Y).$$

La matrice carrée obtenue par juxtaposition des colonnes $X_1, \ldots, X_n$ commence par une ligne de zéros : elle n'est donc pas inversible, et le premier terme dans la somme ci-dessus est lui aussi nul.

Enfin pour chaque i $(1 \leq i \leq n)$, la matrice carrée formée des colonnes $X_1, \ldots, X_{i-1}, Y, X_{i+1}, \ldots, X_n$ a exactement la forme préparée par le lemme : son déterminant est donc $(-1)^{i+1}m_{1i}$.

Il reste donc précisément la formule annoncée.

•

Remarque : Avec les mêmes efforts et un peu de concentration sur les signes, on pourrait écrire une formule analogue pour développer un déterminant par rapport à n'importe quelle ligne. Cela ne me paraît pas indispensable, dans la mesure où un simple échange de lignes permet de ramener en haut la ligne intéressante (au prix d'un changement de signe du déterminant).

Définition 27-6-217 : On dira qu'une matrice carrée est **triangulaire inférieure** lorsque tous ses coefficients au-dessus de la diagonale principale (en formules : ceux des termes (i,j) avec $i<j$) sont nuls.
Corollaire 27-6-8 : Le déterminant de toute matrice triangulaire inférieure est égal au produit de ses termes diagonaux.
Démonstration : Simple récurrence sur la taille de la matrice : pour une matrice $(1,1)$ c'est évident ; pour une matrice $(n+1,n+1)$, la première ligne ne contenant qu'un terme non nul, le premier, le développement par rapport à la première ligne ramène aussitôt au calcul d'un déterminant de matrice triangulaire inférieure (n,n).
•

Proposition 27-6-141 : Soit A une matrice inversible. L'inverse de A est donné par la formule :

$$A^{-1}=\frac{1}{\det A}{}^t\mathrm{com}\,A.$$

Démonstration : Notons m_{ij} les mineurs de A et c_{ij} ses cofacteurs (pour $1\leq i\leq n$, $1\leq j\leq n$).
Notons B la matrice $\dfrac{1}{\det A}{}^t\mathrm{com}\,A$: ainsi pour $1\leq i\leq n$ et $1\leq j\leq n$, $b_{ij}=\dfrac{1}{\det A}c_{ji}$.
Calculons le produit $P=AB$.
Calculons

$$\begin{aligned}p_{11}&=a_{11}b_{11}+a_{12}b_{21}+\cdots+a_{1n}b_{n1}\\&=\frac{1}{\det A}(a_{11}c_{11}+a_{12}c_{12}+\cdots+a_{1n}c_{1n})\\&=\frac{1}{\det A}(a_{11}m_{11}-a_{12}m_{12}+\cdots+(-1)^{n+1}a_{1n}m_{1n}).\end{aligned}$$

D'après la formule du développement du déterminant de A par rapport à la première ligne, $p_{11}=1$.
Calculons maintenant

$$\begin{aligned}p_{21}&=a_{21}b_{11}+a_{22}b_{21}+\cdots+a_{2n}b_{n1}\\&=\frac{1}{\det A}(a_{21}c_{11}+a_{22}c_{12}+\cdots+a_{2n}c_{1n})\\&=\frac{1}{\det A}(a_{21}m_{11}-a_{22}m_{12}+\cdots+(-1)^{n+1}a_{2n}m_{1n}).\end{aligned}$$

On remarque alors que cette dernière parenthèse est la formule qui apparaît dans le développement par rapport à la première ligne du déterminant

$$\begin{vmatrix}a_{21}&a_{22}&\ldots&a_{2n}\\a_{21}&a_{22}&\ldots&a_{2n}\\a_{31}&a_{32}&\ldots&a_{3n}\\\vdots&\vdots&&\vdots\\a_{n1}&a_{n2}&\ldots&a_{nn}\end{vmatrix}$$

et ce déterminant est nul (les deux premières lignes étant identiques) donc $p_{21}=0$.
On calculerait de même tous les p_{ij} mais la débauche d'indices me pousse à laisser le calcul "au lecteur" (d'autant qu'il utilise le développement par rapport à une ligne quelconque, dont j'ai fait remarquer que je pourrais l'écrire, mais que je n'ai pas écrit...)
•

7 - Existence du déterminant

Le moment est venu de prouver le théorème 27-2-56. Le principe est de définir une application par la formule de développement par rapport à la première ligne, et de vérifier que celle-ci est bien un déterminant. La formule appelant des déterminants plus petits –les mineurs qui y interviennent– la démonstration se fera raisonnablement par récurrence.

Démonstration du théorème 27-2-56
* Sur l'ensemble des matrices $(0,0)$, réduit à la matrice vide, la fonction constante valant 1 est "évidemment" un déterminant. Si ça ne vous convainc pas, commençons la récurrence à $n=1$: en posant $\det(a)=a$, on obtient manifestement un déterminant sur l'ensemble des matrices $(1,1)$.
* Soit $n \geq 2$ fixé. Supposons le théorème vrai sur $\mathcal{M}_{n-1}(\mathbf{K})$ et démontrons le sur $\mathcal{M}_n(\mathbf{K})$.
Pour ce faire, pour A matrice carrée (n,n), notons m_{ij} le mineur de A associé à (i,j) –ce mineur a un sens puisque le déterminant existe pour les matrices $(n-1,n-1)$. Définissons alors une application f de $\mathcal{M}_n(\mathbf{K})$ vers $\mathbf{K}$ en posant, pour tout $A \in \mathcal{M}_n(\mathbf{K})$:

$$f(A) = a_{11}m_{11} - a_{12}m_{12} + a_{13}m_{13} + \cdots + (-1)^{n+1}a_{1n}m_{1n}.$$

Il reste à vérifier que f est bien un déterminant.
Un premier point est évident à vérifier : pour une matrice D diagonale, la formule se réduit à $f(D) = d_{11}m_{11}$ et on voit aussitôt que $f(D)$ est bien le produit des termes diagonaux de D.
La difficulté concerne le produit.
Lemme 27-7-19 : Pour toute matrice-transvection T et toute matrice carrée A, $f(AT) = f(A)$.
Démonstration : Soit $T = M_{ij}(\lambda)$ et notons $B = AT$. On sait que B s'obtient à partir de A en ajoutant λ fois la i-ème colonne de A à la j-ème colonne de A.
Notons M' la matrice des mineurs de B et comparons ceux-ci aux mineurs de A. Pour un indice k (avec $1 \leq k \leq n$), soit A_k la sous-matrice $(n-1,n-1)$ de B obtenue par ablation de la première ligne et de la k-ème colonne et de même B_k. Lorsque l'indice k est à la fois distinct de i et de j, la matrice B_k s'obtient à partir de la matrice A_k en ajoutant un multiple d'une colonne à une autre colonne, donc $m'_{1k} = m_{1k}$. Par ailleurs pour ces indices k, $a_{1k} = b_{1k}$. Pour conclure que $f(A) = f(B)$ il suffit donc de prouver que

$$(-1)^{i+1}a_{1i}m_{1i} + (-1)^{j+1}a_{1j}m_{1j} = (-1)^{i+1}b_{1i}m'_{1i} + (-1)^{j+1}b_{1j}m'_{1j}$$

Pour $k = j$ les choses sont encore plus simples : les modifications faites pour passer de A à B ont concerné la j-ème colonne, celle qu'on a enlevée pour construire A_j puis B_j. Ces deux sous-matrices sont exactement les mêmes, et on a encore $m'_{1j} = m_{1j}$. Par ailleurs $b_{1j} = a_{1j} + \lambda a_{1i}$
En revanche, les choses se compliquent pour $k = i$. Pas de problème pour $b_{1i} = a_{1i}$ mais pour calculer m'_{1i} ça se corse. Écrivons :

$$B_i = \begin{pmatrix} a_{2,1} & \cdots & a_{2,i-1} & a_{2,i+1} & \cdots & & a_{2,j-1} & a_{2,j}+\lambda a_{2,i} & a_{2,j+1} & \cdots & a_{2,n} \\ a_{3,1} & \cdots & a_{3,i-1} & a_{3,i+1} & \cdots & & a_{3,j-1} & a_{3,j}+\lambda a_{3,i} & a_{3,j+1} & \cdots & a_{3,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & & & & \\ a_{n,1} & \cdots & a_{n,i-1} & a_{n,i+1} & \cdots & & a_{n,j-1} & a_{n,j}+\lambda a_{n,i} & a_{n,j+1} & \cdots & a_{n,n} \end{pmatrix}$$

(On notera que cette écriture n'est correcte que pour $i < j$, si $j < i$ il faut tout réécrire en conséquence, ce qu'on "laissera au lecteur").
En jouant sur la linéarité du déterminant $(n-1,n-1)$ par rapport à chacune de ses colonnes,

$$m'_{1i} = \det B_i = \begin{vmatrix} a_{2,1} & \cdots & a_{2,i-1} & a_{2,i+1} & \cdots & & a_{2,j-1} & a_{2,j} & a_{2,j+1} & \cdots & a_{2,n} \\ a_{3,1} & \cdots & a_{3,i-1} & a_{3,i+1} & \cdots & & a_{3,j-1} & a_{3,j} & a_{3,j+1} & \cdots & a_{3,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & & & & \\ a_{n,1} & \cdots & a_{n,i-1} & a_{n,i+1} & \cdots & & a_{n,j-1} & a_{n,j} & a_{n,j+1} & \cdots & a_{n,n} \end{vmatrix}$$
$$+ \lambda \begin{vmatrix} a_{2,1} & \cdots & a_{2,i-1} & a_{2,i+1} & \cdots & & a_{2,j-1} & a_{2,i} & a_{2,j+1} & \cdots & a_{2,n} \\ a_{3,1} & \cdots & a_{3,i-1} & a_{3,i+1} & \cdots & & a_{3,j-1} & a_{3,i} & a_{3,j+1} & \cdots & a_{3,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & & & & \\ a_{n,1} & \cdots & a_{n,i-1} & a_{n,i+1} & \cdots & & a_{n,j-1} & a_{n,i} & a_{n,j+1} & \cdots & a_{n,n} \end{vmatrix}$$

Le premier déterminant dans cette égalité n'est autre que m_{1i} ; le second ressemble à m_{1j} mais ses colonnes sont désordonnées : la colonne avec des indices i n'est pas au bon endroit ! Pour l'y ramener, il suffit de l'échanger avec ses voisines : soit successivement avec la colonne immédiatement à sa gauche, puis celle un

peu plus à gauche, et ainsi de suite jusqu'à un dernier échange, celui avec la colonne portant les numéros $i+1$. Au total, on aura fait $j-1-i$ échanges ; le second déterminant vaut donc $(-1)^{j-i-1}m_{1j}$.
Finalement

$$\begin{aligned}(-1)^{i+1}b_{1i}m'_{1i}+(-1)^{j+1}b_{1j}m'_{1j} &= (-1)^{i+1}a_{1i}(m_{1i}+(-1)^{j-i-1}\lambda m_{1j})+(-1)^{j+1}(a_{1j}+\lambda a_{1i})m_{1j}\\ &= (-1)^{i+1}a_{1i}m_{1i}+(-1)^{j+1}a_{1j}m_{1j}+[(-1)^{j}+(-1)^{j+1}]\lambda a_{1i}m_{1j}\\ &= (-1)^{i+1}a_{1i}m_{1i}+(-1)^{j+1}a_{1j}m_{1j}\end{aligned}$$

ce qui prouve bien que $f(A)=f(B)=f(AT)$.

•

Lemme 27-7-20 : Pour toute matrice diagonale D et toute matrice carrée A, $f(AD)=f(A)f(D)$.
Démonstration : C'est beaucoup plus facile que le lemme précédent. Posons

$$D=\begin{pmatrix}\lambda_1 & 0 & \dots & \dots & 0\\ 0 & \lambda_2 & \ddots & & \vdots\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ \vdots & & \ddots & \lambda_{n-1} & 0\\ 0 & \dots & \dots & 0 & \lambda_n\end{pmatrix}$$

et $B=AD$. La matrice B s'obtient donc en multipliant la première colonne de A par λ_1, la deuxième par λ_2 et ainsi de suite.
Notons là encore m'_{ij} les mineurs de B. Quand on calcule une expression $b_{1i}m'_{1i}$, le coefficient b_{1i} est égal à $\lambda_i a_{1i}$ tandis que le mineur m'_{1i} vaut $\lambda_1\cdots\lambda_{i-1}\lambda_{i+1}\cdots\lambda_n m_{1i}$: le terme $b_{1i}m'_{1i}$ est donc exactement égal à $\lambda_1\cdots\lambda_n a_{1i}m_{1i}$.
En sommant sur tous les termes, $f(AD)=f(B)=\lambda_1\cdots\lambda_n f(A)=f(A)f(D)$. D'où $f(AB)=f(A)f(B)$.

•

Nous sommes maintenant armés pour montrer la multiplicativité de la fonction f. Soit A et B deux matrices carrées (n,n). Comme on a appris à le faire dans la première section, décomposons B en produit de matrices-transvections et d'une matrice diagonale, soit

$$B=S_1S_2\cdots S_kDT_l\cdots T_2T_1.$$

Les deux lemmes qui précèdent permettent alors successivement de calculer $f(S_1)=f(IS_1)=f(I)=1$, puis $f(S_1S_2)=f(S_1)=1$, puis $f(S_1S_2S_3)=f(S_1S_2)=1$ jusqu'à $f(S_1S_2\cdots S_kD)=f(S_1S_2\cdots S_k)f(D)=f(D)$ puis $f(S_1S_2\cdots S_kDT_l)=f(S_1S_2\cdots S_kD)=f(D)$ et jusqu'à $f(B)=f(D)$.
et en recommençant à partir de $f(AS_1)=f(A)$ puis $f(AS_1S_2)=f(AS_1)=f(A)$ et ainsi de suite, on arrive à $f(AB)=f(A)f(D)$.

•

8 - Déterminant et transposition

Théorème 27-8-58 : Soit A une matrice carrée à coefficients dans un corps commutatif. On a l'identité $\det A=\det {}^tA$.

•

Démonstration : Soit $\mathbf{K}$ le corps commutatif des coefficients de A et n son côté. Considérons sur $\mathcal{M}_n(\mathbf{K})$ d'une part l'application déterminant, et d'autre part l'application f définie par $f(M)=\det {}^tM$. Les matrices diagonales étant égales à leur transposée, f a la même valeur que det sur les matrices diagonales ; pour M et N deux matrices de $\mathcal{M}_n(\mathbf{K})$, $f(MN)=\det {}^t(MN)=\det {}^tN{}^tM=\det {}^tN\det {}^tM=f(N)f(M)$.

L'application f est donc un déterminant, donc $f=\det$; en particulier $f(A)=\det A$.

•

Remarque : En conséquence, tout ce qu'on a dit sur les colonnes reste valable sur les lignes ; le développement par rapport à la première ligne peut être remplacé par un développement par rapport à la première colonne ; le déterminant des matrices triangulaires supérieures s'arrange aussi bien que celui des matrices triangulaires inférieures.

9 - Calcul du déterminant par blocs

Théorème 27-9-59 : Soit A et B des matrices carrées respectivement (m,m) et (n,n) et C une matrice (m,n) sur un même corps commutatif $\mathbf{K}$. Notons

$$M = \left(\begin{array}{c|c} A & C \\ \hline 0 & B \end{array}\right)$$

Alors $\det M = \det A \cdot \det B$.

•

Démonstration : Traitons tout d'abord le cas où A n'est pas inversible. Dans ce cas, comme il y a une relation de liaison entre les colonnes de A, il y a une relation de liaison entre les premières colonnes de M, donc M n'est pas non plus inversible. Dès lors $\det M$ et $\det A$ sont nuls et la formule est vraie. On traiterait de même le cas où B n'est pas inversible (en raisonnant sur les lignes).

On peut donc supposer que A et B sont inversibles.

Remarquons préalablement que

$$M = \left(\begin{array}{c|c} A & C \\ \hline 0 & B \end{array}\right) = \left(\begin{array}{c|c} A & 0 \\ \hline 0 & I_n \end{array}\right) \left(\begin{array}{c|c} I_m & 0 \\ \hline 0 & B \end{array}\right) \left(\begin{array}{c|c} I_m & A^{-1}C \\ \hline 0 & I_n \end{array}\right)$$

(simple calcul par blocs)

La dernière des trois matrices de ce produit est triangulaire supérieure ; nous savons donc calculer son déterminant, égal au produit des termes diagonaux : il vaut 1.

Pour la deuxième, le lemme 27-6-17 est prêt à servir : elle est égale à $\det B$. Pour la première, il faudrait réécrire la démonstration du lemme 27-6-17 pour voir qu'il marche aussi dans ce sens et qu'elle vaut $\det A$.

On a donc $\det M = \det A \cdot \det B \cdot 1$.

•

10 - Quelques définitions complémentaires

Définition 27-10-218 : Soit E un espace vectoriel de dimension finie et u un endomorphisme de E. Soit $\underline{e}$ une base de E. Le **déterminant** de l'endomorphisme u est le déterminant de la matrice de u dans $\underline{e}$.

Remarque : Cette définition n'a de sens que si le résultat ne dépend pas de la base $\underline{e}$ utilisée pour calculer le déterminant. C'est bien le cas : notons en effet A la matrice de u dans $\underline{e}$ et soit $\underline{f}$ une autre base de E et B la matrice de u dans $\underline{f}$. On sait que si on note P la matrice de passage de $\underline{e}$ à $\underline{f}$, A et B sont liées par la relation $B = P^{-1}AP$, et donc sont semblables ; elles ont donc même déterminant, et la définition est sans ambiguïté.

Cette notion permet de transposer immédiatement certains des résultats énoncés pour les matrices carrées, on voit aussitôt par exemple que pour tous endomorphismes u et v d'un même espace vectoriel,

$$u \text{ est bijectif} \iff \det u \neq 0$$

$$\det(u \circ v) = \det u \det v.$$

Définition 27-10-219 : Soit E un espace vectoriel de dimension n finie, $(e_1, \ldots, e_n) = \underline{e}$ une base de E et $(f_1, \ldots, f_n)$ un système de n vecteurs de E. Le **déterminant** dans $\underline{e}$ de $(f_1, \ldots, f_n)$ est le déterminant de la matrice carrée dont la i-ème colonne $(1 \leq i \leq n)$ est la colonne des coordonnées de f_i dans $\underline{e}$.

Notation 27-10-77 : Ce déterminant sera noté $\det_{\underline{e}}(f_1, \ldots, f_n)$.

Remarque : Contrairement au déterminant d'un endomorphisme, celui-ci dépend de la base utilisée pour le calcul ! Il peut être intéressant pour ne pas l'oublier de savoir que sa valeur absolue a une interprétation géométrique simple : c'est le rapport du volume du parallélépipède (peut-être aplati) construit sur les vecteurs $f_1, \ldots, f_n$ par le volume du parallélépipède (forcément non aplati) construit sur $e_1, \ldots, e_n$.

Proposition 27-10-142 : Soit E un espace vectoriel de dimension n finie, $(e_1, \ldots, e_n) = \underline{e}$ une base de E et $(f_1, \ldots, f_n)$ un système de n vecteurs de E.

$(f_1, \ldots, f_n)$ est libre si et seulement si $\det_{\underline{e}}(f_1, \ldots, f_n) \neq 0$.

Démonstration : Notons C_i la colonne des coordonnées de f_i dans $\underline{e}$ $(1 \leq i \leq n)$. Alors $(f_1, \ldots, f_n)$ est lié si et seulement si $(C_1, \ldots, C_n)$ est lié, c'est-à-dire si et seulement si $(C_1, \ldots, C_n)$ n'est pas un système générateur de l'espace des matrices-colonnes $(n, 1)$, c'est-à-dire si et seulement si la matrice carrée A ayant pour colonnes les C_i n'est pas de rang n, donc si et seulement si $\det A = 0$.

•

Chapitre 28 - Diagonalisation ; vecteurs propres

La question de la diagonalisation peut être présentée comme concernant un endomorphisme (d'un espace vectoriel de dimension finie) ou une matrice. Le choix sera le contraire de celui du chapitre précédent : presque tout est écrit pour des endomorphismes, et les matrices sont mentionnées pour mémoire en fin de chapitre.

1 - Quelques définitions

Définition 28-1-220 : Soit u un endomorphisme d'un espace vectoriel E de dimension finie. On dit que u est **diagonalisable** lorsqu'il existe une base de E dans laquelle la matrice de u est diagonale.
Remarque : Soit e_i un vecteur d'une telle base, et λ_i le terme diagonal correspondant de la matrice diagonale associée à cette base. On a donc $u(e_i) = \lambda_i(e_i)$. Ceci justifie l'intérêt éminent de l'équation

$$u(x) = \lambda x$$

dans laquelle coexistent deux inconnues : le scalaire λ et le vecteur x.

Cette équation admet des solutions stupides : prendre n'importe quel λ et $x = 0$. Elles sont hélas trop simples pour être exploitables : pas question de mettre le vecteur nul dans une base ! Les solutions non évidentes ont droit à un nom :
Définition 28-1-221 : Soit u un endomorphisme d'un espace vectoriel E, soit λ un scalaire et soit x un vecteur. On dit que x est un **vecteur propre** pour la **valeur propre** λ lorsque $x \neq 0$ et $u(x) = \lambda x$.
Notation 28-1-78 : Soit u un endomorphisme d'un espace vectoriel E et λ un scalaire. On note

$$E_\lambda = \{x \in E \mid u(x) = \lambda x\}.$$

Remarque : On voit aussitôt sur la définition que $E_0 = \operatorname{Ker} u$. C'est tout bête, mais combien d'étudiants ai-je vu qui n'en étaient pas conscients !
Proposition 28-1-143 : Pour tout endomorphisme u d'un espace vectoriel E et tout scalaire λ, E_λ est un sous-espace vectoriel de E.
Démonstration : Il est clair que $u(0) = 0 = \lambda \cdot 0$ donc $0 \in E_\lambda$. Par ailleurs, si $x, y \in E_\lambda$ et α est un scalaire, comme

$$u(\alpha x + y) = \alpha u(x) + u(y) = \alpha(\lambda x) + \lambda y = \lambda(\alpha x + y)$$

le vecteur $\alpha x + y$ est aussi dans E_λ.

•

Faisons le point de notre vocabulaire : deux situations peuvent se produire pour le sous-espace E_λ :
* ou bien λ n'est pas une valeur propre de u ; dans ce cas E_λ est réduit à $\{0\}$ et n'est pas bien intéressant. S'autoriser à manipuler la notation E_λ est pratique, mais l'usage ne lui donne pas de nom à cet espace dégénéré.
* ou bien λ est une valeur propre de u et dans ce cas E_λ n'est pas réduit à $\{0\}$. On lui donne un nom :
Définition 28-1-222 : Soit u un endomorphisme d'un espace vectoriel E et λ une valeur propre de u. L'espace E_λ est appelé l'**espace propre** associé à λ.

La proposition suivante est très facile à démontrer, mais bien ingénieuse : elle nous apprend que pour résoudre $u(x) = \lambda x$ il faut d'abord se consacrer à résoudre l'équation en l'inconnue λ et garder la recherche de x pour la suite.
Proposition 28-1-144 : Soit u un endomorphisme d'un espace vectoriel E de dimension finie et λ un scalaire.

Alors λ est une valeur propre de u si et seulement si $\det(u - \lambda Id) = 0$.
Démonstration : λ est une valeur propre de u si et seulement si l'équation d'inconnue x : $u(x) = \lambda x$ a une solution autre que la solution nulle, c'est-à-dire si et seulement si l'équation d'inconnue x : $(u - \lambda Id)x = 0$ a une solution autre que la solution nulle, c'est-à-dire si et seulement si $\operatorname{Ker}(u - \lambda Id) \neq \{0\}$, c'est-à-dire

si et seulement si $u - \lambda Id$ est un endomorphisme non injectif, c'est-à-dire si et seulement si $u - \lambda Id$ est un endomorphisme non bijectif, c'est-à-dire si et seulement si $\det(u - \lambda Id) = 0$.

•

2 - Le polynôme caractéristique

Comme on s'en convainc après avoir traité quelques exemples, l'expression $\lambda \mapsto \det(u - \lambda Id)$ est manifestement polynomiale en λ, et commence par un terme en $(-1)^n\lambda^n$, où n est la dimension de l'espace vectoriel ambiant.

Avançant d'un pas dans l'abstraction, on va introduire un polynôme dont la fonction précédente est la fonction polynomiale associée. Ici le gain que fournit cette abstraction est très réel, et c'est la première occasion que j'ai de mettre en relief qu'elle justifie à plein le travail du chapitre "polynômes". En effet tout ce qui va suivre va être centré sur la notion de "racines simples" ou de "racines multiples" des polynômes, et cette notion ne se laisse pas approcher facilement (voire n'a aucun sens, si le corps est fini) pour des fonctions polynomiales.

Définition 28-2-223 : Soit A une matrice carrée (n,n) à coefficients dans un corps commutatif. On appelle **polynôme caractéristique** de A le polynôme $\det(A - XI)$.

Notation 28-2-79 : Le polynôme caractéristique de A est noté χ_A.

Proposition 28-2-145 : Deux matrices semblables ont le même polynôme caractéristique.

Démonstration : Soit A et B deux matrices semblables, et P une matrice inversible telle que $B = P^{-1}AP$. On a alors $P^{-1}(A - XI)P = P^{-1}AP - X(P^{-1}P) = B - XI$. Les matrices $A - XI$ et $B - XI$ sont donc semblables, donc ont le même déterminant.

•

Remarques : * Le calcul qui précède semble très naturel mais cache des subtilités : que signifie exactement "la matrice $A - XI$" ? Dans les conventions qui ont été choisies, les matrices doivent avoir leurs éléments dans un corps commutatif. Mais l'indéterminée X qui apparaît dans les coefficients de $A - XI$ n'est pas un élément du corps $\mathbf{K}$ des scalaires de E. Pour donner un sens précis à cette définition en restant cohérent avec les conventions de ce cours, il faut préalablement avoir assimilé la notion de "corps des fractions rationnelles" sur $\mathbf{K}$ –que je n'ai que brièvement évoquée en amphi– et considérer la matrice $A - XI$ comme une matrice à coefficients dans ce corps $\mathbf{K}(X)$. (Une autre solution serait d'écrire la théorie des matrices et des déterminants dans le cadre d'un anneau commutatif et non d'un corps commutatif, mais les démonstrations ne pourraient toutes être gardées telles quelles et il faudrait faire un tri attentif pour savoir quels résultats restent utilisables et lesquels sont à jeter).

* Dans le droit fil de la remarque précédente, l'énoncé "le polynôme caractéristique est un polynôme" mériterait une démonstration ! Avec sa définition comme déterminant d'une matrice dont les coefficients sont des fractions rationnelles, le polynôme caractéristique n'est *a priori* qu'une fraction rationnelle, et montrer que c'est un polynôme demande un calcul. Pour ne pas trop troubler le lecteur, je ferai semblant d'oublier ce détail.

* Le passage par les matrices semble un peu maladroit, mais il est significativement plus délicat de donner un sens à $u - XId$ qu'il l'est de donner un sens à $A - XI$.

Définition 28-2-224 : Soit u un endomorphisme d'un espace vectoriel E de dimension finie et soit A la matrice de u dans une base de E. Le **polynôme caractéristique** de u est le polynôme caractéristique de la matrice A.

Notation 28-2-80 : Le polynôme caractéristique de u est noté χ_u.

Remarque : L'usage d'une base nécessite de vérifier que le résultat ne dépend pas de la base choisie ! Comme on sait que deux matrices du même endomorphisme sont semblables (c'est la théorie des matrices de passage qui l'affirme), cela découle aussitôt de la proposition précédente.

Proposition 28-2-146 : Soit u un endomorphisme d'un espace vectoriel E de dimension finie n. Le polynôme caractéristique χ_u est de degré n ; son coefficient dominant est $(-1)^n$; son terme constant est $\det u$.

Démonstration : L'écrire proprement serait pas loin d'un enfer. On va se contenter donc d'une explication en agitant plus ou moins les bras : regardez ce qui se passe quand je développe χ_u par rapport à la première ligne. Les termes autres que le premier sont le produit d'une constante par des mineurs $(n-1, n-1)$ qui sont manifestement de degré inférieur ou égal à $n-1$; le premier terme est le produit de $a_{11} - X$ par un mineur $(n-1, n-1)$ dont le terme de plus haut degré vient manifestement du $a_{22} - X$ et ainsi de suite...

Le terme de plus haut degré dans χ_u est donc de degré n et son coefficient vient du produit des -1 qui s'alignent devant les X sur la diagonale de $A-XI$. Ce coefficient dominant est donc $(-1)^n$.

Enfin pour le terme constant, on se convainc sans difficulté (ou on montre proprement avec bien des difficultés) que cela revient au même de calculer d'abord $\det(A-XI)$ puis de substituer 0 à X dans le polynôme obtenu que de substituer d'abord 0 à X dans tous les polynômes apparaissant comme coefficients de $A-XI$ puis d'appliquer la fonction déterminant. Dès lors le terme constant de χ_u, valeur obtenue par le première méthode décrite, est égal à $\det u$, valeur obtenue par la deuxième méthode décrite.

•

Théorème 28-2-60 : Soit u un endomorphisme d'un espace vectoriel de dimension finie et λ un scalaire. Alors

λ est une valeur propre de u si et seulement si λ est une racine de χ_u

et dans ce cas, en notant m_λ la multiplicité de λ comme racine de χ_u,

$$1 \leq \dim E_\lambda \leq m_\lambda.$$

Démonstration : La première équivalence n'est qu'une redite de la proposition de la section 1 en utilisant le langage des polynômes au lieu du langage des fonctions polynomiales. L'inégalité $1 \leq \dim E_\lambda$ n'est que la traduction de la définition de “valeur propre”. En revanche la deuxième inégalité est nouvelle, et on notera qu'elle n'a vraiment de sens qu'en travaillant sur des polynômes (il n'y a pas de notion de multiplicité de racine d'une simple fonction).

Pour la montrer, notons d la dimension de E_λ et prenons une base $(e_1,\dots,e_d)$ de E_λ. Par le théorème de la base incomplète, on peut prolonger $(e_1,\dots,e_d)$ en une base $(e_1,\dots,e_n)$ de E. Dans cette base, la matrice de u est de la forme :

$$A = \left(\begin{array}{ccccc|c} \lambda & 0 & \dots & \dots & 0 & \\ 0 & \lambda & \ddots & & \vdots & \\ \vdots & \ddots & \ddots & \ddots & \vdots & C \\ \vdots & & \ddots & \lambda & 0 & \\ 0 & \dots & \dots & 0 & \lambda & \\ \hline & & 0 & & & B \end{array}\right).$$

Et donc

$$\det(A-XI) = \left|\begin{array}{ccccc|c} \lambda-X & 0 & \dots & \dots & 0 & \\ 0 & \lambda-X & \ddots & & \vdots & \\ \vdots & \ddots & \ddots & \ddots & \vdots & C \\ \vdots & & \ddots & \lambda-X & 0 & \\ 0 & \dots & \dots & 0 & \lambda-X & \\ \hline & & 0 & & & B-XI \end{array}\right|.$$

En calculant par blocs ce déterminant, on obtient $\chi_u = (\lambda-X)^d \det(B-XI) = (-1)^d (X-\lambda)^d \det(B-XI)$.

On en déduit que $(X-\lambda)^d$ divise χ_u donc que $d \leq m_\lambda$.

•

Corollaire 28-2-9 : En dimension n, un endomorphisme possède au plus n valeurs propres.
Démonstration : Le polynôme χ_u, de degré n, possède au plus n racines.

•

Corollaire 28-2-10 : Les espaces propres sont en somme directe.
Démonstration : Notons F la somme des espaces propres. Remarquons que $u(F) \subset F$: en effet pour tout $x \in F$, par définition de la somme de sous-espaces, il existe des vecteurs $x_1,\dots,x_k$, propres pour les valeurs propres $\lambda_1,\dots,\lambda_k$ tels que $x = \lambda_1 x_1 + \cdots + \lambda_k x_k$; on a alors $u(x) = u(x_1 + \cdots + x_k) = u(x_1) + \cdots + u(x_k) = \lambda_1 x_1 + \cdots + \lambda_k x_k$ et donc $u(x) \in F$. Cela a donc un sens de parler de la restriction v de u à F.

Soit $\lambda_1,\dots,\lambda_r$ l'énumération complète des valeurs propres de u (ou de v). Pour chaque valeur propre λ_i, notons d_i la dimension de l'espace propre correspondant de u (qui est aussi l'espace propre correspondant de v) et m_i la multiplicité de λ_i comme racine de χ_v.

On a alors, en sommant les inégalités du théorème, appliquées à v :

$$d_1 + \cdots + d_r \leq m_1 + \cdots + m_r$$

et aussi, parce que chaque $(X - \lambda_i)^{m_i}$ divise χ_v :

$$m_1 + \cdots + m_r \leq \mathrm{d}^\circ \chi_v = \dim F$$

et enfin, parce que F est la somme des espaces propres :

$$\dim F \leq d_1 + \cdots + d_r.$$

On a donc égalité de ces trois quantités, et en particulier $\dim F = d_1 + \cdots + d_r$. Par le théorème de caractérisation des sommes directes (théorème 4-7-9) la somme est donc directe.

•

Nous en savons désormais assez pour caractériser les endomorphismes diagonalisables :

Proposition 28-2-147 : Soit u un endomorphisme d'un espace vectoriel E de dimension finie. L'endomorphisme u est diagonalisable si et seulement si les conditions suivantes sont toutes deux réalisées :

* Le polynôme caractéristique χ_u est scindé.

* Pour toute valeur propre λ, la multiplicité de λ comme racine de χ_u est égale à la dimension de l'espace propre E_λ.

Démonstration :

* Preuve de $\Rightarrow$. Supposons u diagonalisable. Soit $(e_1, \ldots, e_n)$ une base de E dans laquelle la matrice de u est diagonale, et notons la matrice de u dans E

$$A = \begin{pmatrix} \lambda_1 & 0 & \ldots & \ldots & 0 \\ 0 & \lambda_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \lambda_{n-1} & 0 \\ 0 & \ldots & \ldots & 0 & \lambda_n \end{pmatrix}.$$

À partir de la forme diagonale de A on calcule aussitôt son polynôme caractéristique qui est $\chi_u = \chi_A = (\lambda_1 - X) \cdots (\lambda_n - X)$ et qui est donc visiblement scindé.

Pour ce qui est de l'identité concernant les dimensions, soit λ une valeur propre de u ; notons m la multiplicité de λ dans χ_u et d la dimension de E_λ. Le théorème précédent fournit l'égalité toujours vraie $d \leq m$. Pour l'autre inégalité, notons $i_1, \ldots, i_m$ les indices i, en nombre m, en lesquels le coefficient diagonal λ_i est égal à λ. Chacun des vecteurs $e_{i_1}, \ldots, e_{i_m}$ est alors un vecteur propre pour la valeur propre λ : on sait donc trouver un système libre de m vecteurs propres pour λ ; la dimension de l'espace propre correspondant vaut donc au moins m d'où l'inégalité $m \leq d$.

* Preuve de $\Leftarrow$. Supposons que u vérifie les deux hypothèses de la proposition.Soit $\lambda_1, \ldots, \lambda_r$ la liste des valeurs propres de E ; on notera (pour $1 \leq i \leq r$) E_i en abréviation de E_{λ_i}, m_i la multiplicité de λ_i dans χ_u, égale par hypothèse à la dimension de E_i.

Comme le polynôme χ_u est scindé, son degré doit être égal au nombre total de facteurs du premier degré qui apparaissent dans sa factorisation, donc à la somme des multiplicités de toutes ses racines : on a donc

$$m_1 + \cdots + m_r = \mathrm{d}^\circ \chi_u = n.$$

On sait que les espaces propres sont en somme directe ; l'entier $m_1 + \cdots + m_r$ qui est la somme des dimensions des espaces propres est donc aussi la dimension de leur somme. La somme des espaces propres est donc égale à l'espace E tout entier et on peut écrire (sans se tromper) :

$$E = E_1 \oplus \cdots \oplus E_r.$$

Prenons alors pour chaque i $(1 \leq i \leq r)$ une base $(e_1^{(i)}, \ldots, e_{m_i}^{(i)})$ de E_i et considérons le système obtenu par juxtaposition de toutes ces bases, je veux dire le système $(e_1^{(1)}, \ldots, e_{m_1}^{(1)}, e_1^{(2)}, \ldots, e_{m_2}^{(2)}, \ldots, e_1^{(r)}, \ldots, e_{m_r}^{(r)})$.

Avec un peu d'habitude des sommes directes, on ne doute pas un instant que ce système est une base de E. Si besoin, en voici l'explication : c'est évidemment un système générateur de $E = E_1 + \cdots + E_r$ et comme il a $n = m_1 + \cdots + m_r$ vecteurs, c'en est une base.

Tous les vecteurs de cette base de E sont alors propres ; la matrice de u dans cette base est donc diagonale.

•

Corollaire 28-2-11 : Soit u un endomorphisme d'un espace vectoriel E de dimension finie. Si χ_u est scindé et toutes ses racines sont simples, u est diagonalisable.

Démonstration : La première des conditions de la proposition précédente est vérifiée, pour la seconde, notons que pour une racine simple λ de χ_u, les inégalités du théorème ci-dessus se résolvent à $1 \leq \dim E_\lambda \leq 1$ et garantissent donc l'égalité de la multiplicité (à savoir 1) de λ dans χ_u et de la dimension de l'espace propre correspondant.

•

3 - Diagonalisation des matrices

Il s'agit simplement de réécrire quelques définitions données pour les endomorphismes dans le cadre des matrices.

Définition 28-3-225 : Soit A une matrice carrée à coefficients dans un corps commutatif. On dit que A est **diagonalisable** lorsqu'il existe une matrice inversible P telle que $P^{-1}AP$ soit diagonale.

Le lien entre les deux concepts sera fait par la proposition (évidente) suivante :

Proposition 28-3-148 : Soit u un endomorphisme d'un espace vectoriel E de dimension finie, $\underline{e}$ une base de E et A la matrice de u dans cette base.

Alors u est diagonalisable si et seulement si A est diagonalisable.

Démonstration : Si u est diagonalisable, Soit $\underline{f}$ une base diagonalisant u et notons P la matrice de passage de $\underline{e}$ à $\underline{f}$. La matrice de u dans la nouvelle base est $P^{-1}AP$; ainsi $P^{-1}AP$ est diagonale et A est donc diagonalisable.

Réciproquement, si A est diagonalisable, diagonalisée par la matrice P, appelons f_i le vecteur de E dont la matrice dans $\underline{e}$ est la i-ème colonne de P (pour $1 \leq i \leq n$, où n est la dimension de E) : il est clair que $\underline{f} = (f_1, \ldots, f_n)$ est une base de E et que P est la matrice de passage de $\underline{e}$ à $\underline{f}$. Dès lors la matrice de u dans $\underline{f}$ est $P^{-1}AP$ et est donc diagonale.

•

Remarque : Pour transférer les résultats démontrés au sujet des endomorphismes à une matrice A à coefficients dans le corps commutatif $\mathbf{K}$, il faut préalablement disposer d'un espace vectoriel, d'une base de celui-ci et enfin d'un endomorphisme dont la matrice dans cette base est A. En construire est très facile : on peut prendre $\mathbf{K}^n$ pour espace vectoriel, puis sa base canonique, et enfin l'endomorphisme u dont la matrice dans la base canonique est A. Une fois construit cet endomorphisme, tout se transportera à la matrice A.

Remarque : Il n'est pas très clair de savoir ce qu'on doit appeler "vecteur propre" d'une matrice. De nombreuses sources appellent un peu par abus de langage "vecteurs propres" d'une matrice carrée A tout vecteur colonne X non nul tel que AX soit proportionnel à X, mais je n'aime pas trop ça et préfère personnellement éviter ce terme. En revanche, les énoncés seront trop alambiqués si on s'interdit la

Définition 28-3-226 : Soit A une matrice carrée. On appelle **valeur propre** de A toute racine du polynôme caractéristique de A.

TABLE DES MATIÈRES