Présentation des chapitres de mathématiques de MP

Ci-dessous sont introduits et motivés les différents chapitres que vous découvrirez l'an prochain. Mes explications sont parsemées de petits exercices normalement peu difficiles mais, je l'espère, assez éclairants. Cette lecture et ce travail vous donneront une vue d'ensemble des enjeux et vous permettront d'avoir les idées claires dès l'abord de ces chapitres en MP (mais il s'agit uniquement d'une introduction : il ne s'agit pas d'entrer dans le vif du sujet ni d'introduire de nouvelles notions, au-delà de quelques manipulations naïves).

Ils sont présentés dans le même ordre prévisionnel que celui dans lequel ils seront traités durant l'année. La table des matières est à la fin du document.

Lorsque l'énoncé d'un exercice est accompagné du symbole « ✪ » dans la marge, cela signifie qu'il ne peut pas être traité avec toute la rigueur mathématique attendue tant qu'on est cantonné aux outils de MPSI (j'indique alors clairement ce qu'il est nécessaire d'admettre). Vous pourrez essayer de les traiter à nouveau, rigoureusement, au moment adéquat l'année prochaine.

On peut se demander quel est l'intérêt de traiter un exercice où l'on passe à côté de difficultés essentielles : ces exercices concernent majoritairement les *problèmes d'interversion* (que j'introduis dans les sections 2 et 8). Ils sont le cœur d'exercices souvent épineux parce qu'ils nécessitent des compétences variées : 1° la mémorisation ou compréhension d'énoncés aux hypothèses et conclusions nombreuses, qui se ressemblent beaucoup d'un énoncé à l'autre sans être exactement les mêmes ; 2° la reconnaissance d'une situation où on doit les appliquer ; 3° la maîtrise des majorations fines en analyse, 4° la maîtrise dans l'étude des convergences de séries et intégrales, 5° la maîtrise dans le calcul brut. Une solution à cela est de *diluer* ces difficultés. Avec ces exercices, vous aurez déjà une bonne pratique des points 2° et 5°, et vous pourrez vous consacrer pleinement aux points 1°, 3° et 4° durant l'année. → pages 3 et 30

1 Convergence des suites numériques ou de fonctions

Dans ce chapitre préliminaire très court, nous donnerons quelques compléments sur les suites numériques, puis nous introduirons ce qu'est une suite de fonctions, avec l'outil de base nécessaire à l'étude d'une telle suite : la norme infinie.

Une difficulté en analyse est lorsqu'on veut étudier la convergence d'une suite complexe $(u_n)_{n\geqslant 0}$ pour laquelle on n'a pas la moindre idée de sa limite ℓ, ou lorsque son existence est déjà un problème en soi (c'est fréquemment le cas si ℓ est la somme d'une série, mais pas seulement). On contourne la difficulté grâce à une idée naturelle, et que nous devons formaliser : si $(u_{n+1}-u_n)_{n\geqslant 0}$ ou $(u_{n+p}-u_n)_{n\geqslant 0}$ est « suffisamment petit » pour n au voisinage de l'infini, alors il est tentant de penser que $(u_n)_{n\geqslant 0}$ converge. L'avantage est qu'étudier la « taille » de ces quantités ne nécessite aucunement de connaître *a priori* la limite ℓ.

Nous formalisons ce qu'on entend par être « suffisamment petit ». Dans le cas de $(u_{n+1}-u_n)_{n\geqslant 0}$, il s'agit d'être suffisamment petit pour que la série télescopique $\sum_{n\geqslant 0}(u_{n+1}-u_n)$ converge : c'est le lien suite-série, que vous connaissez déjà. C'est l'accent sur sa mise en œuvre qui est éventuellement inédit. En application, nous démontrons l'existence de la constante d'Euler-Mascheroni et le théorème du point fixe de Banach-Picard *via* l'utilisation de séries télescopiques adéquates.

Dans le cas de $(u_{n+p}-u_n)_{n\geqslant 0}$, c'est la notion de suite de Cauchy qui permet de formaliser le fait d'être suffisamment petit pour que la suite $(u_n)_{n\geqslant 0}$ converge. Une suite est de Cauchy si $u_{n+p}-u_n$ est arbitrairement petit (c'est-à-dire : plus petit en valeur absolue qu'un réel strictement positif $\varepsilon>0$ quelconque) pour n suffisamment grand (c'est-à-dire : au-delà d'un certain rang N), et uniformément : le rang N ne doit pas dépendre de p. Cela se quantifie ainsi :

$$\forall\varepsilon>0,\ \exists N\in\mathbb{N},\ \forall n\geqslant N,\ \forall p\in\mathbb{N},\quad |u_{n+p}-u_n|\leqslant\varepsilon.$$

Cela n'équivaut pas à la convergence de $(u_{n+p}-u_n)_{n\geqslant 0}$ pour tout $p\in\mathbb{N}$ (puisque dans ce cas, le rang N au-delà duquel $|u_{n+p}-u_n|$ est inférieur à $\varepsilon>0$ dépend éventuellement de p).

Exercice 1.

1. Montrer que la suite $(\ln(n))_{n\geqslant 1}$ n'est pas de Cauchy alors que pour tout $p\in\mathbb{N}$, on a : $\lim\limits_{n\to+\infty}(\ln(n+p)-\ln(n))=0$.
2. Soit $\alpha\in]0,1[$. Faire la même vérification avec la suite $(n^\alpha)_{n\geqslant 0}$.

J'insiste sur le côté *uniforme* de la majoration par ε (indépendance en p du rang N), puisque ce mot reviendra très souvent cette année et sera la clé pour avoir de bons théorèmes.

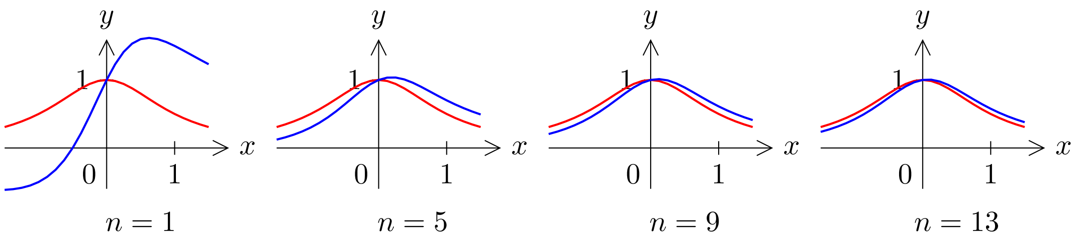
Nous démontrerons alors qu'une suite de Cauchy réelle ou complexe est nécessairement convergente, ce qui fournira une nouvelle approche pour démontrer la convergence de suites sans manipuler la limite candidate ℓ. En application, nous (re)démontrerons que la convergence absolue d'une suite numérique implique sa convergence.

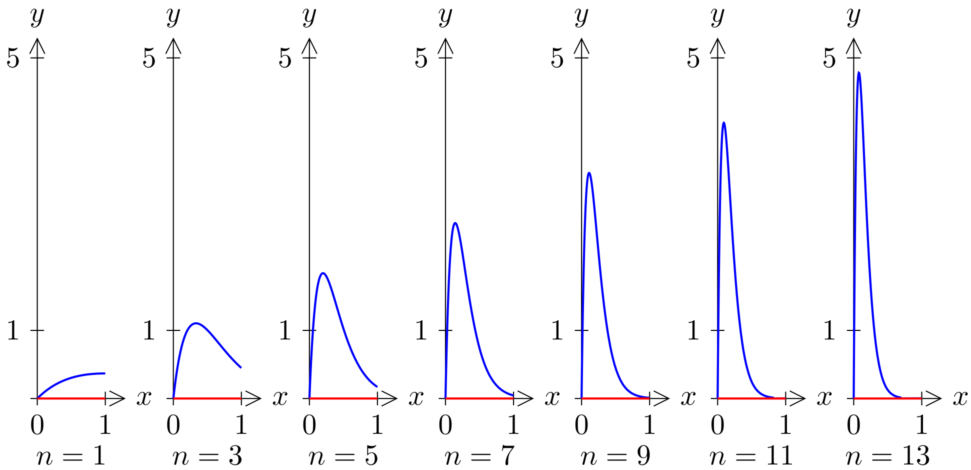
La seconde partie de ce chapitre préliminaire introduira les suites de fonctions qui, sans grande surprise, concernent des fonctions dépendant d'un paramètre entier n, et dont on étudie le comportement quand $n\to+\infty$. Nous définirons la convergence d'une telle suite ; plus précisément, il sera question de convergence *simple* ou *uniforme* : ici, la convergence uniforme de $(f_n:I\to K)_{n\geqslant 0}$ vers une fonction $f:I\to K$ signifie que la majoration de $|f_n(x)-f(x)|$ par ε est vraie au-delà d'un certain rang N indépendant de $x\in I$:

$$\forall\varepsilon>0,\ \exists N\in\mathbb{N},\ \forall n\geqslant N,\ \forall x\in I,\quad |f_n(x)-f(x)|\leqslant\varepsilon.$$

Nous nous efforcerons de donner une compréhension intuitive de ce qu'est la convergence uniforme, afin de la reconnaître à l'œil nu dans les cas simples ; on retiendra notamment qu'une suite de fonctions $(f_n)_{n\geqslant 0}$ converge uniformément vers f si le graphe de f_n se rapproche indéfiniment du graphe de f, tandis qu'elle ne converge pas uniformément (mais simplement) lorsqu'un maximum qui se déplace quand n varie ; on parle couramment de « bosse glissante » pour qualifier ce phénomène.

Ci-dessous nous présentons un exemple visuel de suite de fonctions uniformément convergente, et un autre où ce n'est pas le cas (le graphe rouge est celui de la fonction f qui serait la seule limite candidate, le bleu est celui de f_n) :





La convergence simple de $(f_n : I \to K)_{n\geqslant 0}$ vers $f : I \to K$ signifie simplement (hé oui) que l'on a : $\forall x \in I,\ f(x) = \lim\limits_{n\to+\infty} f_n(x)$. En termes epsilonesques, cela revient à prendre les quantificateurs ci-dessus, mais en mettant « $\forall x \in I$ » en début de ligne (en particulier, N peut dépendre de x).

Une problématique récurrente en analyse est : si $(f_n)_{n\geqslant 0}$ converge (uniformément ou simplement) vers une fonction f, est-ce que la fonction f hérite des propriétés de la fonction f_n ? Nous ne traiterons que très partiellement cette question dans ce chapitre, reléguant la plupart des problématiques et des théorèmes qui y répondent au chapitre *Suites et séries de fonctions* (page 30), néanmoins l'interprétation graphique ci-dessus, et ses deux exemples, rendent acceptable l'intuition suivante : c'est en cas de convergence uniforme que le graphe de f et le graphe de f_n semblent suffisamment proches pour partager des propriétés de régularité : par exemple, en cas de convergence uniforme, on imaginerait mal f_n être continue, c'est-à-dire avoir un graphe qui se trace sans lever le stylo, tandis que le graphe de f serait en deux tenants. Ainsi nous montrerons que la réponse à la question précédente est positive dans le cas de la convergence uniforme d'une suite de fonctions continues : la limite reste continue. Ce n'est pas le cas si l'on n'a pas convergence uniforme :

Exercice 2.

1. Soit $I = [0,1]$. On pose : $\forall n \in \mathbb{N} \setminus \{0\},\ \forall x \in I,\ f_n(x) = x^n$. Montrer que $(f_n)_{n\geqslant 1}$ converge simplement sur I vers une fonction f qui n'est pas continue sur I (et pourtant, f_n est continue pour tout entier $n \geqslant 1$).
2. Soit $a \in [0,1[$. On prend cette fois-ci : $I = [0, a]$. Montrer que $(f_n)_{n\geqslant 1}$ converge uniformément (et simplement) sur I vers une fonction f, et comparer les régularités de f_n pour tout $n \in \mathbb{N} \setminus \{0\}$ et de f.

On retiendra, aussi bien pour les suites de Cauchy que pour les suites de fonctions, le principe selon lequel ce sont les majorations uniformes qui permettent d'avoir des théorèmes d'analyse fins. Ce chapitre est surtout pour l'illustrer et vous y préparer mentalement.

2 Intégration

2.1 Intégrales impropres

Nous commencerons le chapitre d'intégration de 2$^{\text{e}}$ année par une généralisation des intégrales définies en 1$^{\text{re}}$ année : alors que vous n'intégriez que sur des segments jusqu'à présent, on s'autorisera désormais à intégrer sur tout type d'intervalle. Il peut y avoir plusieurs motivations à cela :

— un simple désir d'exhaustivité : en principe, il ne devrait pas y avoir de raison de se limiter à des segments : en 1$^{\text{re}}$ année vous définissiez l'intégrale d'une fonction continue comme limite d'intégrales de fonctions en escalier, et l'apport majeur du segment fut que la continuité uniforme (découlant du théorème de Heine) vous permit aisément de fabriquer des fonctions en escalier f_n « approchant bien » une fonction continue f (avec le vocabulaire de 2$^{\text{e}}$ année, vous diriez que $(f_n)_{n\geqslant 0}$ converge *uniformément* vers f) ; bref, la restriction à des segments était une question de confort et non un obstacle théorique majeur ;

— comme le physicien ou le probabiliste le sait bien, une intégrale peut être interprétée comme une moyenne, disons sur un ensemble de valeurs prises par une variable aléatoire X ; mais si X peut prendre n'importe quelle valeur réelle par exemple (imaginons par exemple que X représente le nombre de secondes avant la désintégration d'un atome : X peut prendre n'importe quelle valeur de $\mathbb{R}_+$), alors une moyenne impliquant X nécessite de savoir calculer une intégrale de la forme $\displaystyle\int_{-\infty}^{+\infty} \bigstar$: encore faut-il savoir définir un tel objet et le manipuler dans les calculs ;

— dans la même veine que l'item précédent : produire des objets invariants par une opération nécessite parfois de « moyenner » ; par exemple, pour fabriquer une fonction 1-périodique à partir d'une fonction quelconque f, il suffit de considérer $x \mapsto \displaystyle\sum_{n\in\mathbb{Z}} f(x+n)$ (à condition que cela ait un sens) ; mais si l'on veut produire une fonction invariante par une opération faisant intervenir un ensemble continu plutôt que discret ($\mathbb{Z}$ dans l'exemple ci-avant est discret), par exemple $x \mapsto ax$ pour tout $a \in \mathbb{R}_+^*$, alors la somme ne peut faire l'affaire, puisque seules les familles au plus dénombrables sont sommables : c'est l'intégration qui permet de remplacer la somme, mais dans ce cas l'intervalle d'intégration peut éventuellement être $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_+^*$, etc. ;

— même en intégrant sur des segments, des passages à la limite peuvent naturellement faire intervenir des intervalles qui n'en sont pas ; par exemple, il serait tendant de dire que : $\lim\limits_{n\to+\infty} \displaystyle\int_0^n \left(1-\frac{t}{n}\right)^n \mathrm{d}t = \int_0^{+\infty} e^{-t}\mathrm{d}t$, étant donné que : $\displaystyle\bigcup_{n=1}^{+\infty}[0,n] = \mathbb{R}_+$, et : $\forall t \in \mathbb{R}_+$, $\lim\limits_{n\to+\infty}\left(1-\dfrac{t}{n}\right)^n = e^{-t}$; mais pour cela, encore faut-il que l'intégrale $\displaystyle\int_0^{+\infty} e^{-t}\mathrm{d}t$ ait un sens.

L'ensemble des motivations ci-dessus n'est pas exhaustif.

Pour définir l'intégrale d'une fonction continue par morceaux sur un intervalle de la forme $[a,b[$ (où b est un réel ou $+\infty$), nous utiliserons un passage à la limite :

$$\int_a^b f = \lim_{x\to b^-}\int_a^x f,$$

sous réserve que la limite existe ; le cas échéant, nous parlerons d'intégrale *convergente*. Si l'intervalle est de la forme $]a,b]$ avec $a \in \mathbb{R}\cup\{-\infty\}$, on procède de même : $\displaystyle\int_a^b f = \lim_{x\to a^+}\int_x^b f$. La définition pour une intégrale sur $]a,b[$ est légèrement différente *.

Exercice 3. Soit $s \in \mathbb{R}$. Discuter la convergence des intégrales suivantes, en utilisant la définition ci-dessus :

$$\int_0^{+\infty} e^{-st}\mathrm{d}t,\ \int_0^1 \frac{\mathrm{d}t}{t^s},\ \int_1^{+\infty}\frac{\mathrm{d}t}{t^s},\ \int_0^1 \ln(t)\mathrm{d}t,\ \int_1^e \frac{\mathrm{d}t}{t\sqrt{1-\ln(t)}},\ \int_0^{\frac{\pi}{2}}\tan(t)\mathrm{d}t,\ \int_1^{+\infty}\frac{\sin(\sqrt{t})}{\sqrt{t}}e^{\cos(\sqrt{t})}\mathrm{d}t,$$

et donner leur valeur en cas de convergence.

L'interprétation visuelle comme aire algébrique reste valable. Dans l'idée, pour qu'une intégrale $\displaystyle\int_a^{+\infty} f$ converge, il suffit que f décroisse « suffisamment vite » au voisinage de $+\infty$ pour que l'aire sous la courbe y soit « négligeable ». Cette interprétation très informelle n'a bien sûr aucune valeur de démonstration, mais elle permet déjà de pressentir les intégrales qui convergent ou non.

Exercice 4.

1. Pour chacune de ces intégrales, se les représenter graphiquement (même très grossièrement), et en déduire pour lesquelles il est ÉVIDENT que l'intégrale ne converge pas (sans démonstration) :

$$\int_0^{+\infty}(\sin(t))^2\mathrm{d}t,\ \int_0^{+\infty}\sin(t)\mathrm{d}t,\ \int_1^{+\infty}\frac{(\sin(t))^3}{t}\mathrm{d}t,\ \int_0^{+\infty}(t^3-2t+6)\mathrm{d}t,\ \int_0^{+\infty}\frac{\ln(t)}{t}\mathrm{d}t,\ \int_0^{+\infty}\lambda\mathrm{d}t,$$

   où $\lambda \in \mathbb{R}$. Ne rien conclure si ce n'est pas « évident ».

2. Démontrer vos conjectures si possible.

De plus, on sent que la formalisation de cette intuition peut se faire *via* un théorème de comparaison, comme pour les séries numériques : c'est en effet le moyen rigoureux de définir ce que, dans la proposition : « être suffisamment petit pour que l'intégrale converge », l'on entend par « suffisamment petit ». Ce théorème de comparaison aura le même intérêt que pour les séries : en général on ne sait pas calculer explicitement une intégrale (pire encore : parfois, on ne *peut* pas, un exemple fameux étant $x \mapsto \displaystyle\int_0^x e^{-t^2}\mathrm{d}t$ qui ne peut pas s'exprimer grâce à des sommes, produits, exponentiations et compositions de fonctions usuelles), et donc le calcul direct et explicite de $\displaystyle\lim_{x\to b^-}\int_a^x f$ est hors de portée. Une comparaison permet donc d'avoir la convergence de l'intégrale de manière indirecte.

*. Cette définition de l'intégrale sur un intervalle quelconque est assez laborieuse, mais c'est celle au programme. Je donnerai aussi la « vraie » définition éventuellement selon mon avancée dans le traitement du chapitre.

Ainsi que ces intégrales généralisées soient simples d'emploi dans les calculs, nous généraliserons toutes les propriétés de base connues dans le cas d'un segment, ainsi que l'importante formule de l'intégration par parties ou du changement de variable.

2.2 Passages à la limite sous le signe intégrale, intégrales à paramètres

Si l'on s'intéresse à des calculs de limite de la forme : $\lim_{n\to+\infty}\int_I f_n(t)\mathrm{d}t$, dans le cas où l'intégrale n'est pas explicitement calculable, une envie très naïve serait de tout simplement passer à la limite dans l'intégrale, en calculant d'abord pour tout $t\in I$: $\lim_{n\to+\infty} f_n(t)$, puis en intégrant le résultat obtenu. Autrement dit, on aimerait écrire :

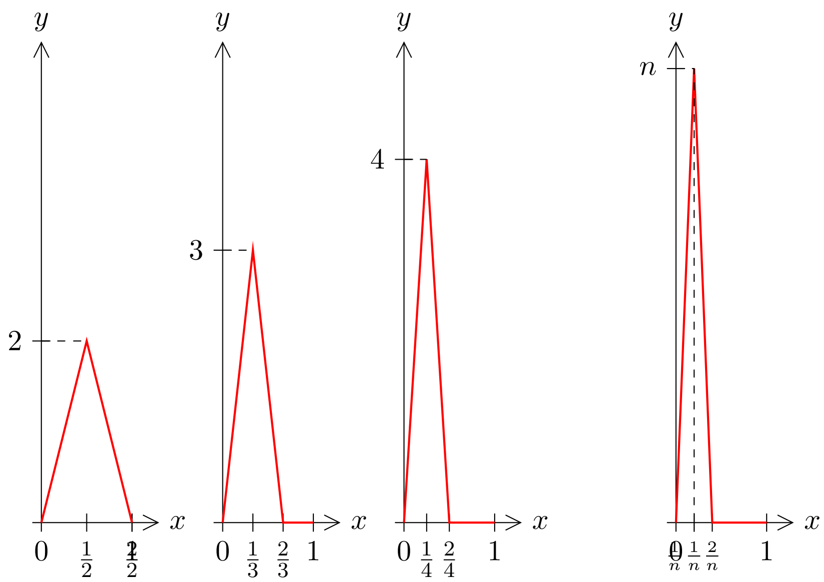
$$\lim_{n\to+\infty}\int_I f_n(t)\mathrm{d}t = \int_I \lim_{n\to+\infty} f_n(t)\mathrm{d}t. \tag{$*$}$$

C'est ce qu'on appelle un *problème d'interversion*. Oui, c'est un problème, parce qu'on se heurte très vite à des contre-exemples de cette identité, même quand les limites existent et sont finies :

Exercice 5. Montrer :

$$\lim_{n\to+\infty}\int_0^1 nt^n\mathrm{d}t \neq \int_0^1 \lim_{n\to+\infty} nt^n\mathrm{d}t, \quad \text{et :} \quad \lim_{n\to+\infty}\int_0^1 n^2e^{-nx}\mathrm{d}x \neq \int_0^1 \lim_{n\to+\infty} n^2e^{-nx}\mathrm{d}x.$$

Le problème à l'origine de ces deux contre-exemples, qui empêche l'égalité $(*)$ d'être vraie, est le même. Nous allons le mettre en valeur avec un exemple plus visuel. Pour tout $n\in\mathbb{N}\setminus\{0,1\}$, considérons la fonction f_n définie sur $[0,1]$ dont le graphe est le suivant (je représente d'abord les cas $n\in\{2,3,4\}$, puis n arbitraire) :



Alors : $\forall x\in[0,1]$, $\lim_{n\to+\infty} f_n(x) = 0$ (pour $x=0$ c'est immédiat, et pour $x\in]0,1]$ on note que pour n suffisamment grand, x est « à droite du pic » et donc $f_n(x)=0$), et donc : $\int_0^1 \lim_{n\to+\infty} f_n(x)\mathrm{d}x = \int_0^1 0\mathrm{d}x = 0$. Or, d'après la formule : Aire d'un triangle $= \frac{\text{Base}\times\text{Hauteur}}{2}$, on a : $\forall n\in\mathbb{N}\setminus\{0,1\}$, $\int_0^1 f_n(x)\mathrm{d}x = \frac{\frac{2}{n}\times n}{2} = 1$, donc : $\lim_{n\to+\infty}\int_0^1 f_n(x)\mathrm{d}x = 1$. Or : $1\neq 0$, donc on a bien démontré : $\lim_{n\to+\infty}\int_0^1 f_n(x)\mathrm{d}x \neq \int_0^1 \lim_{n\to+\infty} f_n(x)\mathrm{d}x$.

Exercice 6. Donner une description analytique de f_n pour tout $n \in \mathbb{N}\backslash\{0,1\}$, et vérifier soigneusement les affirmations ci-dessus (à savoir : $\forall x \in [0,1]$, $\lim\limits_{n\to+\infty} f_n(x) = 0$, et : $\forall n \in \mathbb{N}\setminus\{0,1\}$, $\displaystyle\int_0^1 f_n(x)\mathrm{d}x = 1$).

Analysons pourquoi ce contre-exemple en est un : on note que toute la contribution à l'intégrale des f_n se concentre en un « pic » de plus en plus resserré. Or, quand on passe à la limite quand $n \to +\infty$, le pic est resserré en un point, et vous savez depuis le cours de MPSI que les intégrales ignorent la valeur des fonctions en des points isolés : ainsi toute la masse qui était concentrée en un pic s'évapore subitement en passant à la limite. C'est la source des contradictions ci-dessus.

Exercice 7. Se représenter les graphes des applications $t \mapsto nt^n$ et $x \mapsto n^2e^{-nx}$ sur $[0,1]$, et reconnaître la formation du fameux « pic ». Comment le mathématiser ?

On a deux moyens d'y remédier :

— en ajoutant l'hypothèse supplémentaire que $(f_n)_{n\geqslant 0}$ converge uniformément (nous avons en effet souligné que cette notion de convergence empêchait l'apparition de pics) ;

— en imposant une borne *simultanée* à toutes les fonctions f_n, pour les empêcher de prendre des valeurs arbitrairement grandes.

Ces deux moyens d'y remédier impliquent deux théorèmes d'interversion, c'est-à-dire deux théorèmes dont la conclusion permet d'écrire $(*)$. L'un d'eux est le *théorème de convergence dominée*, qui est l'un des théorèmes d'analyse les plus utiles et importants de classes préparatoires. Le terme « dominée » renvoie à l'idée d'une borne simultanée pour toutes les fonctions f_n.

Lorsque ces deux théorèmes ne seront pas applicables, un autre recours (qui est le même qu'en 1re année) sera un usage adéquat du théorème des gendarmes : on encadre l'intégrande, puis par croissance de l'intégrale on en déduit un encadrement de l'intégrale.

Exercice 8. Déterminer, avec le théorème des gendarmes, les limites des intégrales suivantes quand $n \to +\infty$:

(a) $\displaystyle\int_1^2 \frac{e^{nt}}{\sqrt{1+t^3}}\mathrm{d}t$, (b) $\displaystyle\int_1^2 \frac{\sin(t)e^{-nt}}{1+\sqrt{t}}\mathrm{d}t$, (c) $\displaystyle\int_1^e \frac{\ln(t)}{n+t}\mathrm{d}t$, (d) $\displaystyle\int_0^1 \frac{t^n}{\sqrt{1+t^n}}\mathrm{d}t$, (e) $\displaystyle\int_1^{\sqrt{n}} \frac{e^{-t/n}}{t^2}\mathrm{d}t$.

Exercice 9.

1. Montrer : $\displaystyle\lim_{n\to+\infty}\int_0^{\frac{\pi}{2}} (\sin(t))^n\mathrm{d}t = 0$.
2. Montrer : $\displaystyle\lim_{n\to+\infty}\int_0^1 \frac{x^n}{\sqrt{1-x}}\mathrm{d}x = 0$ (voir les commentaires de la section 2.1 pour la définition de cette intégrale malgré le problème de définition en 1). ← page 3

Néanmoins ces deux exercices (surtout le dernier) nécessitent des majorations techniques et parfois tortueuses. Cela échoue dès que l'intégrande est trop compliqué. Le théorème de convergence dominée apportera beaucoup de confort dans ces calculs de limites.

Le calcul de telles limites a déjà un intérêt en soi, mais il peut aussi servir en sens inverse : étant donnée une intégrale $\displaystyle\int_I f$ qu'on ne sait pas calculer, on peut essayer de l'obtenir comme une limite d'intégrales $\displaystyle\int_I f_n$ où les fonctions f_n sont « simples », de sorte que ces intégrales soient calculables, et donc la limite ℓ de cette suite d'intégrales aussi. Par unicité de la limite : $\ell = \displaystyle\int_I f$. On peut par exemple obtenir ainsi la valeur de l'intégrale de Gauß $\displaystyle\int_0^{+\infty} e^{-t^2}\mathrm{d}t$.

☆ **Exercice 10.** On ADMET : $\displaystyle\lim_{n\to+\infty}\int_0^n \left(1-\frac{t^2}{n}\right)^n \mathrm{d}t = \int_0^{+\infty} e^{-t^2}\mathrm{d}t$. En déduire la valeur de l'intégrale de Gauß (*indication : se ramener* via *changement de variable à une intégrale de Wallis, et utiliser ce que vous aviez démontré là-dessus dans vos exercices de 1re année*).

Ayant à disposition des théorèmes permettant d'écrire sous quelle condition suffisante l'égalité $(*)$ est vraie, et en utilisant la caractérisation séquentielle de la limite, nous pourrons obtenir des énoncés analogues où l'entier n tendant vers $+\infty$ est remplacé par une variable réelle x tendant vers n'importe quel $a \in \mathbb{R} \cup \{\pm\infty\}$. Puisque toutes les notions de régularité (continuité, dérivabilité) se définissent par un passage à la limite, nous en déduirons des énoncés de régularité des fonctions de la forme :

$$g : x \mapsto \int_J f(x,t)\mathrm{d}t$$

qu'on appelle des intégrales à paramètres, le réel x étant ci-dessus le *paramètre* de l'intégrale. La problématique sera essentiellement : est-ce que la fonction g hérite des propriétés de la fonction f ? Si g est dérivable, est-ce que sa dérivée s'exprime en fonction de la dérivée par rapport à x de f ? Nous en déduirons une nouvelle méthode de calculs des intégrales : même si $g(x) = \displaystyle\int_J f(x,t)\mathrm{d}t$ n'est pas calculable par les méthodes standards, on espère néanmoins que $g'(x) = \displaystyle\int_J \frac{\partial f}{\partial x}(x,t)\mathrm{d}t$ l'est (la dérivation de f par rapport à x pouvant faire disparaître un quotient gênant, ou remplacer un logarithme ou une arc tangente par une fraction rationnelle, etc.), de sorte que g' soit explicite, et donc g aussi après intégration.

Exercice 11. On admettra que l'égalité : $\forall x \in I,\ \dfrac{\partial}{\partial x}\displaystyle\int_J f(x,t)\mathrm{d}t = \int_J \frac{\partial f}{\partial x}(x,t)\mathrm{d}t$ est valable, *uniquement dans cet exercice* (ce n'est pas le cas en toute généralité !).

1. Montrer : $\forall x \in \mathbb{R}_+,\ \displaystyle\int_0^1 \frac{t-1}{\ln(t)} t^x \mathrm{d}t = \ln\left(\frac{x+2}{x+1}\right)$ (l'intégrande se prolonge par continuité en 0 et 1 et il n'y a donc pas de problème de définition).
2. Montrer : $\forall (a,b) \in]1,+\infty[^2,\ \displaystyle\int_0^\pi \ln\left(\frac{b-\cos(t)}{a-\cos(t)}\right)\mathrm{d}t = \pi \ln\left(\frac{b+\sqrt{b^2-1}}{a+\sqrt{a^2-1}}\right)$.

Exercice 12. Trouver un contre-exemple à l'égalité : $\dfrac{\partial}{\partial x}\displaystyle\int_J f(x,t)\mathrm{d}t = \int_J \frac{\partial f}{\partial x}(x,t)\mathrm{d}t$. Vous pouvez vous inspirer de l'un des contre-exemples donnés pour l'égalité $(*)$ de cette section.

Plusieurs valeurs d'intégrales remarquables (l'intégrale de Gauß $\displaystyle\int_0^{+\infty} e^{-t^2}\mathrm{d}t$ ou de Dirichlet $\displaystyle\int_0^{+\infty} \frac{\sin(t)}{t}\mathrm{d}t$) s'obtiennent par cette approche ! On ajoute intelligemment une variable x dans l'intégrande (idée qu'on reverra à la section 8.1 pour le calcul de sommes non triviales), on dérive pour se ramener à une intégrale calculable, et on intègre le résultat obtenu. → page 31

2.3 Résumé : objectif du chapitre

Nous définirons les intégrales sur des intervalles quelconques, en généralisant tous les résultats connus dans le segment des intégrales sur des segments. Un peu de prudence est nécessaire, du fait que les intégrales sur des intervalles quelconques n'existent pas toujours (puisque, on l'a vu, c'est défini comme des limites : encore faut-il qu'elles existent). Nous donnerons des moyens pratiques de montrer l'existence de ces intégrales, dont certains rappelleront l'étude des séries numériques (théorèmes de comparaison). Nous donnerons des exemples de référence (fonction Γ d'Euler, intégrale de Dirichlet).

Dans un second temps, nous apporterons des réponses au problème d'interversion : « sous quelles hypothèses l'égalité suivante est-elle valable : $\displaystyle\lim_{n\to+\infty}\int_I f_n(t)\mathrm{d}t = \int_I \lim_{n\to+\infty} f_n(t)\mathrm{d}t$? ». L'une de ces réponses est le fameux théorème de convergence dominée, qui admet également pour corollaires des théorèmes de régularité des intégrales à paramètres, c'est-à-dire : sous certaines hypothèses, l'application $g : x \mapsto \displaystyle\int_J f(x,t)\mathrm{d}t$ hérite de la régularité de f par rapport à x, et dans le cas dérivable on a : $\dfrac{\partial}{\partial x}\displaystyle\int_J f(x,t)\mathrm{d}t = \int_J \dfrac{\partial f}{\partial x}(x,t)\mathrm{d}t$.

Forts de tous ces théorèmes, nous pourrons en exercice étudier voire calculer plusieurs intégrales remarquables :

$$\Gamma : x \mapsto \int_0^{+\infty} t^{x-1}e^{-t}\mathrm{d}t, \qquad \int_0^{+\infty} e^{-t^2}\mathrm{d}t, \qquad \int_0^{+\infty} \frac{\sin(t)}{t}\mathrm{d}t,$$

mais aussi nous reviendrons sur la remarque introductive selon laquelle l'intégration sert à *moyenner*, ou à *régulariser* des fonctions. C'est ce qui fait le succès de plusieurs transformations : notamment la transformée de Laplace vue en SII, mais aussi la *transformée de Fourier* d'une fonction f :

$$\hat{f} : x \mapsto \int_{-\infty}^{+\infty} f(t)e^{-ixt}\mathrm{d}t,$$

ou le *produit de convolution* de deux fonctions f et g :

$$f * g : x \mapsto \int_{-\infty}^{+\infty} f(t)g(x-t)\mathrm{d}t,$$

qu'on peut interpréter comme une moyenne de f pondérée par g, en quelque sorte. Ce moyennage permet de lisser f, par exemple : si g est de classe C^∞, alors $f * g$ l'est aussi (même si f ne l'est pas). Grâce à cette opération de lissage, nous obtiendrons plusieurs résultats d'approximation intéressants, dont le théorème de Weierstraß dans un chapitre ultérieur (toute fonction continue sur un segment est limite uniforme de fonctions polynomiales).

3 Séries numériques et familles sommables

3.1 Séries numériques

Les séries numériques ayant déjà été conséquemment abordées en 1re année, nous n'apporterons que quelques compléments.

En 1re année, lorsque vous utilisiez le théorème de comparaison des séries à termes positifs pour obtenir la nature d'une série, vous faisiez en général la comparaison au terme général d'une série de Riemann. Cela marche effectivement dans la majeure partie des cas, mais pas pour des séries sophistiquées telles que $\displaystyle\sum_{n\geqslant 1} \frac{n!}{n^n}$ ou $\displaystyle\sum_{n\geqslant 0} \frac{(n!)^2}{(2n)!}$ (le problème est notamment la présence de la factorielle, dont l'étude asymptotique est compliquée). Nous donnons une règle plus adaptée pour l'étude de telles séries, et plus généralement de séries dont le terme général a une vitesse de croissance ou décroissance rapide (c'est-à-dire : géométrique) : la règle de D'Alembert.

C'est un théorème de comparaison déguisé (le déguisement tombe dans la démonstration), à des séries géométriques ; de comparaison *intelligente*, même. En effet, dans les deux exemples ci-dessus, on voit mal à quelles séries géométriques on les comparerait ; la règle de D'Alembert part de l'observation que si $\displaystyle\sum_{n\geqslant 0} u_n$ est une série à termes strictement positifs telle que :

$$\lim_{n\to+\infty} \frac{u_{n+1}}{u_n} = \ell \in \mathbb{R}_+,$$

alors : $u_{n+1} \approx \ell u_n$, ce qui est une relation *presque* géométrique de raison ℓ, donc on est tenté d'écrire : $u_n \approx c\ell^n$, avec $c \in \mathbb{R}_+$ (attention, cela n'a rien de correct, on ne donne qu'une heuristique). Or la série $\displaystyle\sum_{n\geqslant 0} \ell^n$ converge si et seulement si $\ell < 1$, donc on devrait s'attendre à ce que la série $\displaystyle\sum_{n\geqslant 0} u_n$ converge si et seulement si $\ell < 1$ également (nous omettons le cas où la limite est infinie : il est facile de montrer dans ce cas que la série diverge grossièrement). Il s'avère que ce raisonnement *pas rigoureux du tout*, après formalisation, nous montre que la conclusion est vraie !... Sauf dans le cas $\ell = 1$, où l'on ne peut pas conclure. En effet, comme 1 est le réel de démarcation entre la convergence et la divergence d'une série géométrique, l'approximation « presque géométrique de raison ℓ » est trop abusive pour permettre de conclure quand $\ell = 1$, selon que $(u_n)_{n\geqslant 0}$ soit « au-dessus ou en dessous » de 1.

Exercice 13. Utiliser la règle de D'Alembert pour obtenir la nature de $\sum_{n\geqslant 1}\frac{n!}{n^n}$, $\sum_{n\geqslant 0}\frac{(n!)^2}{(2n)!}$ et $\sum_{n\geqslant 0}\frac{n^2}{2^n}$.

Exercice 14. Vérifier grâce à plusieurs exemples que le cas $\ell = 1$ ne permet effectivement pas de conclure : la série peut aussi bien converger que diverger dans ce cas de figure.

Exercice 15. Rendre rigoureux le raisonnement informel ci-dessus, grâce à la définition de la limite, pour démontrer la règle de D'Alembert.

L'intérêt de la règle de D'Alembert réside aussi dans sa simplicité : le quotient $\frac{u_{n+1}}{u_n}$ se simplifie considérablement quand u_n dépend de factorielles et de puissances. Mais son défaut est le cas d'incertitude $\ell = 1$.

Exercice 16.

1. Réétudier la nature des deux séries ci-dessus dont le terme général dépend de factorielles, mais cette fois-ci avec la formule de Stirling. Que pensez-vous de l'efficacité de chaque approche ?
2. Étudier la nature de la série $\sum_{n\geqslant 1}\frac{4^n(n!)^2}{(2n)!}$.

Nous compléterons l'étude des séries numériques par des méthodes donnant des équivalents asymptotiques de sommes partielles ou de restes. L'une d'elles est la technique de comparaison série-intégrale (déjà vue en 1re année) :

Exercice 17.

1. Montrer, par une comparaison série-intégrale : $\sum_{k=1}^{n}\frac{1}{k}\underset{n\to+\infty}{\sim}\ln(n)$.
2. Montrer, par une comparaison série-intégrale : $\forall\alpha\in]-\infty,1[$, $\sum_{k=1}^{n}k^{-\alpha}\underset{n\to+\infty}{\sim}\frac{n^{1-\alpha}}{1-\alpha}$. Que donne cette méthode lorsque $\alpha > 1$?
3. Soit $\alpha\in]1,+\infty[$. Trouver un équivalent asymptotique quand $n\to+\infty$ de $\sum_{k=n+1}^{+\infty}k^{-\alpha}$. Que donne cette méthode lorsque $\alpha < 1$?
4. Soit $\alpha\in\mathbb{R}$. Étudier pour quelle valeur de α la méthode de comparaison série-intégrale permet d'obtenir un équivalent asymptotique quand $n\to+\infty$ de $\sum_{k=1}^{n}\frac{1}{k(\ln(k))^\alpha}$. Même question avec $\sum_{k=n+1}^{+\infty}\frac{1}{k(\ln(k))^\alpha}$ (sous réserve que la somme existe).

L'autre moyen proviendra d'un théorème de sommation des relations de comparaison : vous savez qu'en général, les équivalents ne sont pas compatibles avec les sommes. Néanmoins nous montrerons que pour des séries à termes POSITIFS et DIVERGENTES, il est possible de sommer des équivalents (et les autres relations de comparaison). En cas de séries convergentes, il en est de même avec les restes. Comme les suites sont des cas particuliers de séries *via* le lien suite-série, cette méthode permettra d'obtenir des équivalents de suites difficiles à étudier asymptotiquement par les méthodes habituelles.

3.2 Familles sommables

Dans la seconde partie de ce cours, nous chercherons à généraliser les sommes $\sum_{j\in J}x_j$ au cas où J n'est pas fini ni $\mathbb{N}$, le cas $J=\mathbb{N}$ correspondant au cas des séries (à quelques subtilités près, que j'aborde ci-bas). Une idée naïve est de « numéroter » les éléments de J au moyen d'une bijection

$\varphi : \mathbb{N} \to J$ (cela revient bien à numéroter les éléments de J, en décrétant que $\varphi(n)$ est le n^{e} élément de J), afin de se ramener à la situation classique d'une somme indexée par les entiers naturels. On poserait alors, lorsque cela a un sens : $\displaystyle\sum_{j\in J} x_j = \sum_{n=0}^{+\infty} x_{\varphi(n)}$, ce que l'on sait étudier. Mais encore faut-il qu'il existe une telle bijection de $\mathbb{N}$ dans J : cela nous conduira à l'étude des ensembles *dénombrables* (ce sont les ensembles pour lesquels une telle bijection existe). Nous donnerons des exemples de référence d'ensembles qui sont dénombrables et d'ensembles qui ne le sont pas. Puis nous verrons que notre idée naïve est finalement pertinente, puisqu'il n'est pas possible d'obtenir une somme correctement définie (et ayant une valeur finie) autrement qu'avec un ensemble dénombrable.

Exercice 18. Nous vous recommandons de réfléchir de manière imagée dans un premier temps, pour traiter cet exercice. Ou encore : d'abord énumérez *naturellement*, puis réfléchissez ensuite à une bijection $\mathbb{N} \to J$ qui respecterait votre énumération.

1. Montrer que $\mathbb{N}$, $2\mathbb{N}$, $2\mathbb{N}+1$ et $\mathbb{N}\setminus\{0\}$ sont dénombrables. Plus généralement, montrer que toute partie infinie de $\mathbb{N}$ est dénombrable.
2. Montrer que $\mathbb{Z}$ est dénombrable.
3. Montrer que $\mathbb{N}\times\mathbb{N}$ est dénombrable. C'est assez difficile d'écrire la bijection explicitement ici : il est déjà gratifiant de comprendre son principe sur la base d'une interprétation graphique des éléments de $\mathbb{N}\times\mathbb{N}$.
4. Montrer que $\mathcal{P}(\mathbb{N})$ n'est pas dénombrable (*indication : s'il existe une surjection f de $\mathbb{N}$ dans $\mathcal{P}(\mathbb{N})$, considérer un antécédent par f de l'ensemble :* $A = \{x \in \mathbb{N} \mid x \notin f(x)\}$).
5. En déduire que $\{0,1\}^{\mathbb{N}}$ n'est pas dénombrable non plus.

La solution trouvée n'est *a priori* pas totalement satisfaisante : comment s'assurer que la valeur de la somme $\displaystyle\sum_{j\in J} x_j$ ne dépend pas de la bijection choisie ? En fait, elle peut en dépendre, ce qui pose un gros problème de définition... Sauf si la série $\displaystyle\sum_{n\geqslant 0} x_{\varphi(n)}$ est absolument convergente (c'est ce qu'on appellera la *sommabilité* de la famille $(x_j)_{j\in J}$). Dans ce cas, peu importe la bijection $\varphi : \mathbb{N} \to J$ choisie, la somme $\displaystyle\sum_{n=0}^{+\infty} x_{\varphi(n)}$ a toujours la même valeur et donc $\displaystyle\sum_{j\in J} x_j$ ne pose plus de problème d'interprétation.

Ce que nous disons ici ne sera pas exactement l'axe de réflexion choisi pour définir la somme de la famille $(x_j)_{j\in J}$, mais c'est équivalent. Pour démontrer avec plus d'aisance les différentes propriétés de ces sommes (l'objectif étant de s'assurer qu'elles peuvent se manipuler aussi agréablement que les sommes que vous avez manipulées jusqu'à aujourd'hui), nous introduirons une définition plus savante en termes de borne supérieure.

Nous donnerons alors différentes formules de calcul qui permettront de calculer ces sommes avec beaucoup de souplesse, et notamment de se ramener aisément à des sommes indexées par des entiers naturels, même sans avoir à construire de bijection φ (et tant mieux, car c'est parfois difficile : l'exercice ci-dessus vous en convaincra peut-être) :

— le théorème de sommation par paquets, qui autorise à sommer $\displaystyle\sum_{j\in J} x_j$ en regroupant d'abord des x_j dans des paquets plus petits et qui nous arrangent (par exemple, si $J = \mathbb{Z}$, en sommant d'abord les termes positifs d'une part et négatifs d'autre part, ou bien en sommant d'abord deux à deux les termes dont les indices sont opposés) ;

— le théorème de Fubini, permettant de calculer des sommes indexées par des couples d'entiers (ce théorème dit essentiellement : que pour calculer une somme double, on peut sommer d'abord en fixant un entier, puis l'autre, dans l'ordre qu'on veut) ;

— le produit de Cauchy.

Ainsi, sommer sur $\mathbb{Z}$ ou sur $\mathbb{N}^2$ n'aura plus de secret pour vous.

La grande philosophie, pour retenir les résultats sur ces sommes, sera : quand il y a positivité du terme général ou convergence absolue, tous les théorèmes s'appliquent et on somme comme on veut !

4 Structures algébriques

Vous avez vu en 1re année différentes structures algébriques : groupes, anneaux, corps, espaces vectoriels, avec l'accent mis plus particulièrement sur cette dernière structure. Nous revenons sur les autres, surtout sur les groupes.

4.1 Théorie des groupes et des anneaux

La théorie des groupes est infiniment profonde et possède des motivations riches qui lui sont propres (même si l'algébriste finit par découvrir qu'un groupe n'est réellement intéressant que lorsqu'il *agit* sur un ensemble, de la même manière qu'une permutation σ dans S_n ne s'étudie pas en soi, mais à travers son action sur $[\![1,n]\!]$, et malheureusement le programme de MPSI/MP ne met pas en valeur cet aspect). Cependant, le point de départ de ma motivation, dans ce document, puise dans une analogie avec l'algèbre linéaire ; non pas parce que c'est le plus pertinent, mais parce que c'est le plus parlant pour vous, à ce stade de votre formation mathématique.

Vous avez pu juger de l'intérêt des bases en algèbre linéaire de deux manières :

— un certain nombre de résultats, valables pour tout vecteur d'un espace vectoriel de dimension finie, se démontrent en se contentant de vérifier ces résultats pour tout vecteur d'une base ;

— une base de E permet d'exprimer tout vecteur de E à l'aide d'un n-uplet de *coordonnées* qui le caractérise entièrement ; or raisonner avec n-uplet revient à raisonner dans K^n, qui est un espace vectoriel très simple d'étude (là où E ne l'était pas forcément).

Le dernier point se formule de façon plus savante en disant que le choix d'une base permet d'obtenir un isomorphisme d'espaces vectoriels $f : K^n \to E$: comme un isomorphisme préserve tout ce qui est relatif à la structure, on peut raisonner indifféremment avec E ou K^n (de même avec $\mathrm{L}(E)$: le choix d'une base permet de raisonner indifféremment avec $\mathrm{M}_n(K)$).

Exercice 19.

1. Démontrer cette affirmation : quel est l'isomorphisme $f : K^n \to E$ obtenu en fixant une base ?
2. Et si l'on remplace la base par une famille libre, que devient l'application linéaire f de la question précédente ? Et si l'on considère une famille génératrice ?

Étant donné le succès de la notion dans le cadre de l'algèbre linéaire, on peut se demander si on peut la transposer à d'autres structures, à commencer par les groupes. Même si la notion de base se transpose assez mal (et d'ailleurs nous n'en parlerons pas), l'exercice suivant expliquant le type de souci que l'on peut rencontrer :

Exercice 20. Pour comprendre ce qui coince lorsqu'on veut transposer la notion de base à des groupes, étudions le cas particulier de $\mathbb{Z}$. On dit qu'une famille de $\mathbb{Z}$ est libre, génératrice, une base si elle vérifie les propriétés qu'on attend d'une telle famille dans un K-espace vectoriel, mais en remplaçant K par $\mathbb{Z}$ (attention à ne pas vouloir généraliser cela à tout groupe, cela n'aurait en général aucun sens).

1. Montrer que les seules bases de $\mathbb{Z}$ sont (1) et (-1).
2. Montrer que la famille $(2,3)$ est une famille génératrice de $\mathbb{Z}$, mais qu'aucune de ses familles extraites n'est une base de $\mathbb{Z}$.
3. Montrer que le théorème de la base incomplète ne se généralise pas aux groupes, en montrant que pour tout $n \in \mathbb{Z} \setminus \{-1,0,1\}$ la famille (n) est libre (mais n'est pas une base, ce qui contredit un autre résultat classique d'algèbre linéaire : lequel ?) et ne peut pas être complétée en une base de $\mathbb{Z}$.

Mais ce n'est pas si grave : nous allons tout de même définir ce qu'est une partie génératrice d'un groupe. Dans l'idée : si S est une partie de G, alors nous dirons que S engendre G si tout élément $g \in G$ s'écrit comme produit d'éléments de S et de leurs inverses :

$$\forall g \in G,\ \exists r \in \mathbb{N},\ \exists (g_1, \ldots, g_r) \in S^r,\ \exists (\varepsilon_1, \ldots, \varepsilon_r) \in \{-1,1\}^r, \quad g = g_1^{\varepsilon_1} \cdots g_r^{\varepsilon_r}.$$

Si G est commutatif et sa loi notée additivement, alors la dernière égalité devient : $g = \sum_{i=1}^{r} \varepsilon_i g_i$. Mieux : si $S = \{s_1, \ldots, s_k\}$ est une partie finie de G (toujours supposé commutatif), alors le terme général de la somme précédente est de la forme $\pm s_i$ pour tout i et, quitte à regrouper les termes qui sont égaux, on peut réécrire de manière équivalente la propriété d'être une partie génératrice ainsi :

$$\forall g \in G,\ \exists (a_1, \ldots, a_k) \in \mathbb{Z}^k,\ g = \sum_{i=1}^{k} a_i s_i.$$

C'est le point de vue de l'exercice 20.

Exercice 21. Soit $n \in \mathbb{N} \setminus \{0\}$.

1. Justifier que $\mathbb{Z}$ admet $\{1\}$ et $\{-1\}$ comme parties génératrices.
2. Montrer que $\mathbb{U}_n$ admet $\left\{e^{\frac{2i\pi}{n}}\right\}$ comme partie génératrice. En proposer d'autres à un seul élément.
3. Soit C l'ensemble des cycles dans S_n. Justifier que S_n admet C comme partie génératrice.
4. Soit T l'ensemble des transpositions dans S_n. Justifier que S_n admet T comme partie génératrice.

Comme en algèbre linéaire, un certain nombre de résultats peuvent se démontrer en raisonnant uniquement sur une partie génératrice :

Exercice 22. Soit G un groupe dont on note S une partie génératrice.

1. Soit H un sous-groupe de G. Montrer : $H = G \iff S \subseteq H$.
2. Soit $f : K \to G$ un morphisme de groupes. Montrer que f est surjective si et seulement si : $S \subseteq \mathrm{im}(f)$.
3. Soient $f : G \to K$ et $g : G \to K$ deux morphismes de groupes. Montrer : $f = g \iff f_{|S} = g_{|S}$.

En vérité, la définition de partie génératrice donnée ci-dessus n'est pas la plus commode pour traiter cet exercice. Pour la théorie, on privilégiera une autre définition : on définira le sous-groupe engendré par S, noté $\langle S \rangle$, comme le plus petit sous-groupe de G contenant les éléments de S (plus petit au sens de l'inclusion), et on dira que S engendre G si : $\langle S \rangle = G$. Vérifiez que cela simplifie grandement les raisonnements de cet exercice en les affranchissant de calculs.

Nous donnerons des exemples de parties génératrices : l'exercice 21 en donne quelques-uns, mais nous en donnerons davantage, surtout dans le cas particulier du groupe symétrique S_n. Une partie génératrice est en effet d'autant plus intéressante qu'elle a peu d'éléments, et nous pouvons faire mieux que les parties génératrices C et T proposées. Le cas le plus favorable est bien sûr celui où un groupe admet une partie génératrice avec **un seul** élément : dans ce cas, il suffit la plupart du temps de raisonner sur cet unique générateur pour obtenir un résultat valable pour tout le groupe ! Cette situation est si appréciable qu'elle mérite qu'on lui donne un nom : on parle dans ce cas de groupe *monogène*, et plus particulièrement de groupe *cyclique* s'il est en plus de cardinal fini.

Le groupe monogène infini par excellence est $\mathbb{Z}$ (voir l'exercice 21). Nous montrerons en effet que tout autre groupe monogène infini est isomorphe à $\mathbb{Z}$; comme un isomorphisme de groupes conserve tout ce qui est relatif à la structure de groupe, étudier $\mathbb{Z}$ permet d'étudier tous les groupes monogènes infinis. Mais il n'est d'aucune utilité pour les groupes cycliques, qui sont pourtant bien plus nombreux dans la nature (à commencer par les groupes $\mathbb{U}_n$ des racines n^{es} de l'unité). Pour leur étude spécifique, nous aurons besoin d'introduire :

- la notion très importante *d'ordre* d'un élément : si g est un élément d'un groupe G, alors on dit que g est d'ordre fini si le groupe engendré par g est fini, et dans ce cas son cardinal est appelé *l'ordre de* g (de manière équivalente, l'ordre de g est le plus petit entier naturel non nul k tel que : $g^k = 1_G$, s'il en existe) ;
- les groupes $\mathbb{Z}/n\mathbb{Z}$, qui sont informellement « le groupe $\mathbb{Z}$ où l'on a “posé $n = 0$” ».

Au vu de tout ce que l'on a raconté, un groupe G est cyclique s'il admet une partie génératrice de la forme $\{g\}$ avec g d'ordre fini : $\exists g \in G,\ G = \langle g \rangle$. De la même manière que $\mathbb{Z}$ décrit tous les groupes monogènes infinis à isomorphisme près, nous montrerons que $\mathbb{Z}/n\mathbb{Z}$ décrit tous les groupes cycliques (toujours à isomorphisme près). Par conséquent, en étudiant $\mathbb{Z}/n\mathbb{Z}$, nous connaîtrons tous les groupes cycliques. C'est très important, parce qu'il y en a partout ! Pour tout groupe fini G et tout $g \in G$, le groupe engendré par g est nécessairement cyclique, par définition.

Exercice 23. Vérifier l'équivalence entre les deux définitions données de l'ordre d'un élément $g \in G$.

Exercice 24. Sans chercher à être rigoureux : avec la description qu'on a donnée de $\mathbb{Z}/n\mathbb{Z}$, justifier que c'est un groupe pour l'addition à n éléments, et en donner un générateur.

Même dans le cas d'un groupe non monogène, les éléments d'ordre fini sont utiles parce qu'ils donnent des informations sur la structure du groupe (le cardinal, notamment), et permettent d'avoir des isomorphismes avec des groupes plus simples d'étude. En effet, pour expliciter un groupe fini compliqué G, une première piste est éventuellement de considérer un élément non trivial $g_1 \in G$ et ses puissances : 1_G, g_1, g_1^2, etc., jusqu'à ce que cela boucle (et ce sera le cas à partir du moment où l'on obtient $g_1^k = 1_G$: c'est là qu'intervient l'ordre de g) ; puis passe à un second élément g_2, dont on considère toutes les puissances ainsi que les produits par les puissances de g_1 (cela revient à considérer le groupe engendré par g_1 et g_2), et ainsi de suite, espérant ainsi finir par obtenir tout le groupe pour des raisons de cardinalité. Plus les éléments ainsi considérés sont d'ordre élevé, plus on va obtenir d'éléments par ce procédé, et donc plus on est susceptible de reconstituer une grande partie de G. Cette observation est naïve, mais on peut même faire mieux, puisqu'il s'avère qu'il y a des relations de divisibilité très fortes entre les ordres des éléments et le cardinal de G. Imaginons par exemple, pour simplifier, que G soit commutatif, non cyclique, mais admette tout de même une partie génératrice relativement petite $\{g_1, g_2\}$, avec g_1 et g_2 distincts, respectivement d'ordres d_1 et d_2. Il est alors facile de montrer que l'application :

$$f : \begin{cases} \mathbb{Z}/d_1\mathbb{Z} \times \mathbb{Z}/d_2\mathbb{Z} & \to & G \\ (k, \ell) & \mapsto & g_1^k g_2^\ell \end{cases}$$

est bijective ; elle conserve donc les cardinaux, et on a : $\operatorname{card}(G) = d_1 d_2$, donc les ordres de g_1 et g_2 divisent G ; nous verrons en fait que dans un groupe fini, même sans hypothèse de commutativité ni sur les parties génératrices, l'ordre d'un élément divise l'ordre de G (cas particulier du théorème de Lagrange) : c'est un résultat fort, qui impose beaucoup de contraintes sur les ordres de G (quand on connaît le cardinal de G) ou sur G (quand on connaît l'ordre de certains de ses éléments) ! Il nous aidera énormément à expliciter des groupes par de simples raisonnements arithmétiques : voir l'exercice 26 pour une démonstration de cette efficacité, et du lien que cela implique entre le cardinal d'un groupe et sa structure.

Exercice 25. Vérifier que l'application ci-dessus est bien bijective, sous les hypothèses effectuées.

Exercice 26. Soit p un nombre premier. Utiliser le résultat énoncé ci-dessus pour en déduire que tout groupe de cardinal p est cyclique.

Exercice 27. Soit G un groupe fini et commutatif de cardinal n. Soit $g \in G$ un élément d'ordre d. En exprimant le produit $\prod_{x \in G} x$ de deux manières différentes (*utiliser le fait que l'application $x \mapsto gx$ soit bijective*), montrer : $g^n = 1_G$. En déduire que d divise n. C'est un cas particulier du résultat énoncé ci-dessus.

De nombreux exercices seront consacrés aux implications de ce lien entre ordre des éléments, cardinal du groupe, et structure du groupe.

Certaines problématiques de la théorie des groupes ont leur analogue dans la théorie des anneaux. La notion qui leur est spécifique, et que nous introduirons cette année, est celle *d'idéal* d'un anneau : les idéaux sont aux anneaux ce que les sous-espaces vectoriels sont aux espaces vectoriels, et en cela ils sont plus instructifs sur la structure d'un anneau que ses sous-anneaux (ce qui peut paraître contre-intuitif). Nous ne donnerons que les définitions et propriétés de base : c'est le chapitre suivant qui en parlera plus amplement.

4.2 Résumé : objectif du chapitre

Nous allons d'abord donner des énoncés généraux sur les relations d'équivalence et les ensembles-quotients, qui seront pour nous d'un grand confort pour fabriquer de nouveaux groupes et anneaux, à commencer par l'anneau $\mathbb{Z}/n\mathbb{Z}$ dont nous parlions ci-dessus (qui revient à « poser $n = 0$ » dans $\mathbb{Z}$). Si le temps nous le permet, nous montrerons comment ils permettent de donner une définition rigoureuse de $\mathbb{Z}$ et $\mathbb{Q}$ (cette construction étant hors programme).

Ensuite, nous introduirons les notions de partie génératrice d'un groupe, d'ordre d'un groupe ou d'un élément d'un groupe, de groupe monogène ou cyclique, suivant ce que nous disions ci-dessus.

Les groupes sont aussi plus agréables à étudier si l'on considère les morphismes qui les relient ; les meilleurs d'entre eux étant les isomorphismes, puisqu'ils préservent tout ce qui est relatif à la structure : ainsi, ayant un isomorphisme entre un groupe compliqué (inconnu) et un groupe simple (connu), nous pourrons résoudre tout problème dans le groupe inconnu en étudiant le même problème dans le groupe connu. Cette stratégie est très fructueuse et nous ferons donc des rappels sur les morphismes.

Après avoir justifié que tout groupe monogène est isomorphe à $\mathbb{Z}$ ou $\mathbb{Z}/n\mathbb{Z}$, nous étudierons de près le groupe $\mathbb{Z}/n\mathbb{Z}$: en le connaissant, nous connaîtrons tous les groupes monogènes finis, d'après le principe formulé ci-dessus.

L'étude du nombre de générateurs de $\mathbb{Z}/n\mathbb{Z}$ (et donc, par extension, de tout groupe cyclique), et plus généralement du nombre d'éléments d'ordre d dans un groupe cyclique, nous conduira à introduire l'indicatrice d'Euler qui est, par définition, le nombre d'éléments inversibles de $\mathbb{Z}/n\mathbb{Z}$, mais aussi le nombre d'éléments de $[\![1, n]\!]$ qui sont premiers avec n. C'est une fonction arithmétique centrale.

L'étude du groupe symétrique S_n est plus fine parce que c'est un groupe extrêmement riche. Nous donnerons entre autres des exemples de parties génératrices avec des applications du même ordre que celles données plus haut : pour démontrer certains résultats, il suffit parfois de les montrer sur tout élément d'une partie génératrice. Comme les transpositions engendrent S_n, et que de plus elles sont toutes reliées les unes aux autres assez simplement par conjugaison (nous dirons ce qu'on entend par là), cela permet à peu de frais de démontrer certains résultats sur S_n en raisonnant sur une seule transposition !

La partie consacrée aux anneaux sera brève, une étude plus approfondie étant faite au chapitre suivant. Nous introduirons simplement la notion d'idéal, en donnant quelques propriétés de base sur ces analogues annelés des sous-espaces vectoriels, ainsi que la notion d'algèbres (une K-algèbre étant un ensemble muni à la fois d'une structure d'anneau et de K-espace vectoriel).

5 Arithmétique des entiers et des polynômes

5.1 Généraliser le vocabulaire arithmétique : motivation

Ce chapitre sur la reine des mathématiques commence à reformuler les principales notions de l'arithmétique (divisibilité, irréductibilité ou primalité) en des termes plus généraux. L'intérêt ? Pouvoir imiter les raisonnements d'arithmétique sur les entiers dans d'autres circonstances. Cette idée vient après un grand succès qu'eurent Euler puis Kummer dans l'étude d'équations diophantiennes. Leur idée est qu'en faisant de l'arithmétique contenant davantage de nombres que $\mathbb{Z}$, on obtient davantage de factorisations et donc de relations de divisibilité possibles ; ayant plus d'informations, cela permet de résoudre des équations diophantiennes plus facilement.

Un exemple pour illustrer cette idée, où l'on admettra que « tout se passe comme dans $\mathbb{Z}$ » : si l'on cherche à résoudre l'équation : $y^2 + 2 = x^3$, d'inconnue $(x, y) \in \mathbb{Z}^2$, de l'arithmétique élémentaire permet dans un premier temps de montrer que x et y doivent tous les deux être impairs. De plus, dans l'anneau $\mathbb{Z}[i\sqrt{2}] = \left\{ z \in \mathbb{C} \mid \exists (a, b) \in \mathbb{Z}^2,\ z = a + i\sqrt{2}b \right\}$ (qui contient $\mathbb{Z}$), on a la relation :

$$(y + i\sqrt{2})(y - i\sqrt{2}) = x^3$$

et, si on s'autorise à faire de l'arithmétique dans $\mathbb{Z}[i\sqrt{2}]$ comme dans $\mathbb{Z}$: à ce stade, on démontre que les deux facteurs du membre de gauche sont premiers entre eux, c'est-à-dire que tout diviseur commun est inversible. En effet, si $d \in \mathbb{Z}[i\sqrt{2}]$ divise $y + i\sqrt{2}$ et $y - i\sqrt{2}$, alors il divise la différence $-2i\sqrt{2} = (i\sqrt{2})^3$. Or $i\sqrt{2}$ est un nombre premier dans $\mathbb{Z}[i\sqrt{2}]$ (voir l'exercice 29 pour comprendre ce qu'on entend par là), donc l'unicité de la décomposition en facteurs premiers dans $\mathbb{Z}[i\sqrt{2}]$ implique qu'il existe $u \in \mathbb{Z}[i\sqrt{2}]^\times$ et $k \in [\![0,3]\!]$ tels que : $d = u(i\sqrt{2})^k$. Justifions que $k = 0$: si $k \geqslant 1$ alors, du fait que d divise $y + i\sqrt{2}$ et $y - i\sqrt{2}$, son carré d^2 divise leur produit, c'est-à-dire x^3 ; or : $d^2 = u^2(i\sqrt{2})^{2k} = (-1)^k u^2 2^k$: par conséquent, si $k \geqslant 1$, alors 2 divise x^3, et donc x est pair. C'est impossible, on a affirmé tantôt que x est impair ! Par l'absurde : $k = 0$, donc $d = u$ est inversible.

Ainsi $y + i\sqrt{2}$ et $y - i\sqrt{2}$ sont premiers entre eux, et leur produit est un cube, donc ils sont eux-mêmes des cubes. Soit $(a, b) \in \mathbb{Z}^2$ tel que : $y + i\sqrt{2} = (a + i\sqrt{2}b)^3$. En développant cette puissance et en identifiant parties réelles et imaginaires, on obtient :

$$y = a^3 - 6ab^2 = a(a^2 - 6b^2), \quad 1 = 3a^2b - 2b^3 = b(3a^2 - 2b^2).$$

Partant de là, on déduit : $b = 1$, $a^2 = \frac{1+2b^2}{3} = 1$, et il n'est plus difficile d'en déduire que les solutions $(x, y) \in \mathbb{Z}^2$ de l'équation initiale sont $(3,5)$ et $(3, -5)$ (la réciproque étant triviale à vérifier).

Exercice 28. Compléter les éléments non détaillés suivants :

1. Vérifier que $\mathbb{Z}[i\sqrt{2}]$ est un anneau. Pourquoi veut-on un anneau, d'ailleurs ?
2. Soit $(x, y) \in \mathbb{Z}^2$. Vérifier effectivement que si : $y^2 + 2 = x^3$, alors x et y sont impairs.
3. Vérifier que la propriété utilisée dans $\mathbb{Z}[i\sqrt{2}]$ est vraie au moins dans $\mathbb{Z}$: si a et b sont deux entiers relatifs non nuls et premiers entre eux, et si $c \in \mathbb{Z}$ vérifie : $c^3 = ab$, alors il existe $(u, v) \in \mathbb{Z}^2$ tel que : $a = \pm u^3$, $b = \pm v^3$.
4. Soit $x \in \mathbb{Z}$. Montrer que si 2 divise x^3, alors 2 divise x.
5. *A priori*, d^2 divise x^3 dans $\mathbb{Z}[i\sqrt{2}]$, et après j'en déduis que 2 divise x^3 au sens classique, c'est-à-dire dans $\mathbb{Z}$ (puisque je raisonne sur la parité, qui a été définie et établie dans $\mathbb{Z}$). C'est trop rapide : comment suis-je passé de $\mathbb{Z}[i\sqrt{2}]$ à $\mathbb{Z}$? Détailler cette étape.
   On dit que a divise b dans $\mathbb{Z}[i\sqrt{2}]$ s'il existe $c \in \mathbb{Z}[i\sqrt{2}]$ tel que : $b = ac$.
6. Justifier les expressions de a, b, puis de (x, y) proposées (pourquoi n'a-t-on pas $b = -1$?).

Exercice 29.

1. Montrer que s'il existe $(z, z') \in \mathbb{Z}[i\sqrt{2}]^2$ tel que : $i\sqrt{2} = zz'$, alors soit z, soit z' est égal à ± 1.
   Indication : prendre le module au carré pour se ramener à une relation dans $\mathbb{Z}$.
2. Montrer qu'il en est de même avec 2, 3, $1 + i\sqrt{5}$ et $1 - i\sqrt{5}$, mais en remplaçant $(z, z') \in \mathbb{Z}[i\sqrt{2}]^2$ par $(z, z') \in \mathbb{Z}[i\sqrt{5}]^2$.
   Par définition, on a : $\mathbb{Z}[i\sqrt{5}] = \left\{z \in \mathbb{C} \mid \exists (a, b) \in \mathbb{Z}^2,\ z = a + i\sqrt{5}b\right\}$.

Exercice 30.

1. ☆ S'inspirer de cette approche pour démontrer que les triplets pythagoriciens (c'est-à-dire : les triplets $(x, y, z) \in \mathbb{Z}^3$ tels que : $x^2 + y^2 = z^2$) sont tous de la forme :
   $$\left(d(u^2 - v^2), 2duv, d(u^2 + v^2)\right),$$
   avec $(d, u, v) \in \mathbb{Z}^3$. On commencera par se ramener, par de l'arithmétique classique dans $\mathbb{Z}$, au cas où x, y et z sont premiers entre eux dans leur ensemble, puis on justifiera que sous cette hypothèse, exactement un de ces trois entiers est pairs, et que ce ne peut pas être z.
   On admettra qu'on peut faire de l'arithmétique dans $\mathbb{Z}[i]$ comme dans $\mathbb{Z}$.
2. Redémontrer ce résultat par la géométrie en notant que cela revient à expliciter les points $(x, y) \in \mathbb{Q}^2$ tels que : $x^2 + y^2 = 1$.
   Indication : considérer les droites de pente rationnelle et passant par $(-1,0)$. Montrer que les points d'intersection de ces droites avec le cercle unité donnent tous les points que l'on cherche.
3. Redémontrer ce résultat uniquement en recourant à l'arithmétique traditionnelle (dans $\mathbb{Z}$).

Néanmoins attention car l'hypothèse qu'on peut faire de l'arithmétique dans $\mathbb{Z}[i\sqrt{2}]$ comme dans $\mathbb{Z}$ est une très, très grosse hypothèse, loin d'être évidente. Si l'on se plaçait dans $\mathbb{Z}[i\sqrt{5}] = \left\{z \in \mathbb{C} \mid \exists(a,b) \in \mathbb{Z}^2,\ z = a + i\sqrt{5}b\right\}$, les égalités $6 = 2 \times 3 = \left(1 + i\sqrt{5}\right)\left(1 - i\sqrt{5}\right)$ empêchent l'unicité de la décomposition en facteurs premiers, puisque tous ces facteurs sont premiers au sens donné dans l'exercice 29 (en vérité on parle plutôt d'élément *irréductible*, je parle de primalité seulement pour mieux faire le parallèle avec les nombres premiers de $\mathbb{Z}$). Sans cette unicité, on se convainc qu'il n'y a pas non plus de lemme d'Euclide, de théorème de Gauß, de notion claire de pgcd ou ppcm, etc.

L'exemple ci-dessus permet néanmoins de voir que l'arithmétique ne se borne pas à $\mathbb{Z}$. Il s'agit d'étendre les définitions de l'arithmétique à d'autres anneaux, et il s'avère que les idéaux d'un anneau (dont nous parlions déjà dans la section précédente) donnent le bon langage pour cela. Outre la traduction de la divisibilité en termes d'idéaux, c'est pour la définition du pgcd et du ppcm, ainsi que pour les relations de Bezout, que les idéaux apportent davantage de confort, à condition d'être dans un anneau dit « principal » (c'est-à-dire : tout idéal de l'anneau est de la forme $aA = \{ab \mid b \in A\}$ avec $a \in A$). Au-delà de cette définition savante, il faut retenir qu'un anneau principal est un anneau « où on peut faire de l'arithmétique comme dans $\mathbb{Z}$ », et un anneau où il existe un théorème de division euclidienne est toujours principal.

5.2 Arithmétique des polynômes

Or il s'avère que $K[X]$ possède une division euclidienne ! Ainsi, en raisonnant par analogie avec $\mathbb{Z}$, nous montrerons que tous les théorèmes connus de l'arithmétique dans $\mathbb{Z}$ sont valables avec les polynômes : les polynômes se décomposent de manière unique en produit d'irréductibles (qu'il est donc intéressant expliciter, au moins dans $\mathbb{R}[X]$ ou $\mathbb{C}[X]$), le théorème de Bezout reste valable, ainsi que le lemme d'Euclide, etc. Cela aura des conséquences dans plusieurs domaines, en particulier :

— l'algèbre linéaire, puisque nous nous en servirons pour définir et étudier le *polynôme minimal* d'un endomorphisme ou d'une matrice (outil essentiel pour étudier leur « réduction », chose qu'on va motiver dans la section suivante), et pour démontrer le lemme des noyaux (qui permet de décomposer des espaces vectoriels en somme de sous-espaces plus petits, toujours suivant la discussion de la section suivante) ;

— l'étude des nombres algébriques, c'est-à-dire des nombres complexes annulés par un polynôme non nul à coefficients rationnels.

Deux choses rendent l'arithmétique des polynômes parfois plus simple que celle des entiers : il y a deux autres moyens d'obtenir des relations de divisibilité entre polynômes :

— la dérivation ;

— la caractérisation des racines :
$$P(\alpha) = 0 \iff X - \alpha \mid P,$$
que l'on peut généraliser à l'aide de la notion de *polynôme minimal* (soit K un sous-corps de $\mathbb{C}$; pour tout $\alpha \in \mathbb{C}$, on peut définir son polynôme minimal sur K comme le plus petit polynôme unitaire de $K[X]$, au sens du degré, qui annule α, du moins s'il en existe ; mais la meilleure définition est avec en termes d'idéaux) : si $\alpha \in \mathbb{C}$ admet $\pi_\alpha \in K[X]$ pour polynôme minimal sur K, et si $P \in K[X]$, alors :
$$P(\alpha) = 0 \iff \pi_\alpha \mid P.$$
Cela donne en général un facteur de plus haut degré que $X - \alpha$. Ainsi on peut démontrer des relations de divisibilité par une banale évaluation d'un polynôme (pourvu qu'elle soit nulle).

Exercice 31. Soit $\alpha \in \mathbb{C}$. Soit K un sous-corps de $\mathbb{C}$. On suppose qu'il existe des polynômes non nuls de $K[X]$ qui s'annulent en α, de sorte que $\pi_\alpha \in K[X]$ existe.

1. Justifier que π_α est irréductible (*montrer que le cas contraire contredirait sa minimalité*).
2. Montrer l'équivalence ci-dessus avec un bon usage de la division euclidienne.
3. Montrer que si $P \in K[X]$ est irréductible, unitaire, et annule α, alors : $P = \pi_\alpha$. C'est un moyen pratique de déterminer le polynôme minimal.
4. Soit $\theta \in \mathbb{R} \setminus \pi\mathbb{Z}$. Montrer que $X^2 - 2\cos(\theta)X + 1$ est le polynôme minimal sur $\mathbb{R}$ de $e^{i\theta}$.

5. Soit $\theta \in \mathbb{R} \setminus \pi\mathbb{Z}$. Montrer que pour tout $n \in \mathbb{N} \setminus \{0,1\}$, le polynôme $X^2 - 2\cos(\theta)X + 1$ divise le polynôme $\sin(\theta)X^n - \sin(n\theta)X + \sin((n-1)\theta)$.

Cette dernière question montre l'apport de cette équivalence qui n'a aucun... équivalent, dans $\mathbb{Z}$: comment obtiendrait-on cette relation de divisibilité sans s'en servir ?

5.3 Arithmétique modulo n

Enfin, une composante essentielle de l'arithmétique est ce qu'on appelle l'arithmétique modulaire (c'est-à-dire modulo un entier). À cet effet, nous étudierons $\mathbb{Z}/n\mathbb{Z}$ plus longuement que dans le chapitre précédent (qui ne s'intéressait à cet ensemble qu'en tant que groupe cyclique et non pour des questions arithmétiques). Il est aussi muni d'une structure d'anneau provenant de la structure d'anneau de $\mathbb{Z}$ (où l'on « pose $n = 0$ »). Cela permet de montrer des relations de divisibilité du type : « n divise $\bigstar$ », en montrant que $\bigstar$ est nul dans $\mathbb{Z}/n\mathbb{Z}$. Passer par cet anneau a de nombreux avantages : d'abord c'est un ensemble fini, ce qui est plus agréable à manipuler que $\mathbb{Z}$ pour de bêtes raisons combinatoires (dites-vous qu'au moins *en principe*, dans un ensemble fini, on peut par exemple résoudre des équations par recensement exhaustif) ; ensuite, les relations sur les ordres dont nous parlions dans l'introduction du chapitre précédent permettent de calculer très facilement des puissances modulo n. On retrouvera notamment le petit théorème de Fermat (si p est un nombre premier, alors pour tout $a \in \mathbb{Z}$ on a : $a^p \equiv a \bmod p$) et on le généralisera au cas où p n'est pas premier.

Le cas où n est premier est remarquable à bien des égards : dans ce cas $\mathbb{Z}/n\mathbb{Z}$ est un corps. Le fait que tout élément non nul soit inversible rend l'arithmétique modulo un nombre premier très agréable. Pour le cas où n n'est pas premier, un important résultat, appelé le *lemme chinois*, permettra de diluer encore la difficulté des études dans $\mathbb{Z}/n\mathbb{Z}$ en regardant ce qu'il se passe modulo les diviseurs premiers de n : ainsi on pourra (partiellement) se ramener au cas précédent.

5.4 Résumé : objectif du chapitre

Après avoir réinterprété les notions de base de l'arithmétique en termes d'idéaux, nous définirons ce qu'est un anneau principal, et montrerons que dans un anneau principal on « peut faire de l'arithmétique comme dans $\mathbb{Z}$ ». Parmi les propriétés au programme, on insistera plus spécifiquement sur l'existence du ppcm, du pgcd, et le théorème de Bezout. Nous retrouverons les résultats connus dans $\mathbb{Z}$ et vérifierons que l'anneau $K[X]$ est aussi principal : on peut faire de l'arithmétique dans $K[X]$.

Puis nous consacrerons le reste du chapitre à l'arithmétique dans $\mathbb{Z}/n\mathbb{Z}$. Nous aurons besoin pour cela de clarifier sa structure d'anneau, et nous verrons que les diviseurs premiers de n jouent un rôle essentiel pour la comprendre : d'abord, $\mathbb{Z}/n\mathbb{Z}$ est un corps si et seulement si n est premier (ce qui simplifie les calculs dans $\mathbb{Z}/n\mathbb{Z}$, du fait que tout élément non nul soit inversible et que $\mathbb{Z}/n\mathbb{Z} \setminus \{0\}$ soit un groupe) et, même dans le cas où n n'est pas premier, le lemme chinois nous permettra de ramener l'étude de $\mathbb{Z}/n\mathbb{Z}$ à l'étude de $\mathbb{Z}/p^{v_p(n)}\mathbb{Z}$ où p divise n.

Ces résultats profonds de structure nous permettront de démontrer à peu de frais le petit théorème de Fermat et de le généraliser (son intérêt étant de simplifier l'étude des relations de divisibilité vérifiées par les puissances d'entiers relatifs), grâce à l'indicatrice d'Euler que nous avons brièvement mentionnée dans la section précédente. Cette fonction indicatrice éclaire à la fois la structure de groupe de $\mathbb{Z}/n\mathbb{Z}$ en comptant ses générateurs, mais aussi la structure d'anneau en comptant le nombre d'inversibles de $\mathbb{Z}/n\mathbb{Z}$ pour la multiplication, comme nous le verrons. Le lemme chinois nous donnera un moyen de la calculer autrement qu'en revenant à sa définition.

6 Réduction des endomorphismes

Le cours de 1re année a posé les fondements de l'algèbre linéaire moderne. Celui de 2e année aborde son problème majeur, que nous formulons dans un premier temps en des termes vagues : étant donné un espace vectoriel E, sur lequel nous étudions l'action d'une application linéaire f (la plupart du temps il s'agit d'un endomorphisme), on se demande s'il est possible de décomposer E en somme de sous-espaces vectoriels* :

$$E = F_1 \oplus F_2 \oplus \cdots \oplus F_k,$$

où F_i est un sous-espace vectoriel de E tel que la restriction de f à F_i soit « simple » à étudier. Intuitivement, l'étude sur F_i est nécessairement plus simple que sur E pour de bêtes raisons dimensionnelles (si l'on raisonne matriciellement : le nombre de coefficients de la matrice représentative de $f_{|F_i}$ dans une base donnée est d'autant plus petit que la dimension de F_i est petite, le cas le plus favorable étant celui où F_i est une droite vectorielle ; en tous les cas, il y a moins de coefficients à déterminer que si l'on étudie la matrice de f relativement à une base de E), mais nous verrons que cela ne suffit pas et que les F_i devront être obtenus intelligemment.

Une telle décomposition a deux objectifs selon que l'objet d'étude soit E ou f :

— **si E est un espace vectoriel mal connu dont on veut expliciter les éléments :** on veut alors le décomposer comme somme de sous-espaces vectoriels F_i plus « simples », suffisamment simples pour qu'on sache les expliciter ; connaissant les éléments des F_i, on en déduit les éléments de E en sommant ;

— **si f est une application linéaire sur E dont la description est « compliquée », ce qui se traduit souvent par une matrice représentative A qui est elle-même « compliquée » :** dans ce cas il peut être difficile d'obtenir son image (et donc son rang), son noyau, ses puissances, etc. ; mais en se ramenant à des sous-espaces F_i plus « simples », la restriction de f à ces sous-espaces devrait être elle-même plus simple à étudier ; or, si $f_{|F_i}$ est connu pour tout i, par linéarité on en déduit f.

Le programme de MP aborde surtout le second point (d'où le nom du chapitre : on « réduit » des endomorphismes en exprimant leurs matrices sous la forme la plus simple possible), bien que le premier point soit (de mon point de vue) bien plus riche, avec encore des ramifications dans les mathématiques contemporaines.

Le propos est pour l'instant assez abstrait. Le premier item concerne généralement les espaces vectoriels E qui sont les solutions d'une certaine équation linéaire (pour les exemples les plus banals : équation différentielle, espace des suites vérifiant une certaine relation de récurrence, etc.). Le second item sera majoritairement étudié du point de vue matriciel, et reviendra en général à ramener l'étude d'une matrice quelconque à celle d'une matrice diagonale (« la plus simple possible »), ou diagonale « par blocs » ou triangulaire faute de mieux. Nous allons illustrer ceci concrètement dans les deux sous-sections suivantes.

6.1 Pour expliciter des espaces vectoriels mal connus

Pour comprendre cette problématique, oublions un petit moment l'algèbre linéaire et intéressons-nous à la relation de récurrence :

$$\forall n \in \mathbb{N}, \quad u_{n+2} + au_{n+1} + bu_n = 0,$$

où $(a, b) \in \mathbb{R}^2$ est un couple fixé et $(u_n)_{n \geqslant 0}$ une suite réelle. On cherche à expliciter les suites vérifiant une telle relation, sans utiliser les techniques que vous avez vues en 1re année, mais en ramenant cette relation de récurrence d'ordre 2 à des relations d'ordre 1 : on veut *abaisser l'ordre*.

Pour y parvenir : de la même manière que les racines d'un polynôme permettent de le « casser » en plusieurs facteurs de degré plus petit (vu que α est racine d'un polynôme P si et seulement s'il existe un polynôme Q tel que $P = (X - \alpha)Q$), nous espérons que les racines de $X^2 + aX + b$ permettront de « casser » cette relation d'ordre 2 en relations d'ordre 1. À cet effet, nous supposons désormais que le polynôme $X^2 + aX + b$ admet deux racines réelles distinctes r_1 et r_2. Grâce aux relations

∗. Vous n'avez pas encore vu la somme directe de k sous-espaces vectoriels lorsque $k \geqslant 3$. C'est essentiellement la même chose que dans le cas où $k = 2$: l'égalité $E = F_1 \oplus F_2 \oplus \cdots \oplus F_k$ signifie que tout vecteur de E s'écrit de manière unique comme somme de vecteurs de F_1, F_2, etc., F_k.

coefficients-racines, on a : $b = r_1r_2$, et : $a = -(r_1 + r_2)$. La relation d'ordre 2 ci-dessus peut alors se réécrire :

$$\forall n \in \mathbb{N},\ u_{n+2} - r_1u_{n+1} - r_2u_{n+1} + r_1r_2u_n = 0, \quad \text{soit donc :} \quad \forall n \in \mathbb{N},\ u_{n+2} - r_1u_{n+1} = r_2\left(u_{n+1} - r_1u_n\right),$$

si bien qu'en posant :

$$\forall n \in \mathbb{N}, \quad v_n = u_{n+1} - r_1u_n, \quad w_n = u_{n+1} - r_2u_n,$$

on observe que l'on a :

$$\forall n \in \mathbb{N}, \quad v_{n+1} = r_2v_n, \quad \text{et de même :} \quad \forall n \in \mathbb{N}, \quad w_{n+1} = r_1w_n.$$

On a *abaissé l'ordre* : on reconnaît deux suites géométriques, de raisons respectives r_2 et r_1. On en déduit : $\forall n \in \mathbb{N}$, $v_n = r_2^n v_0$, $w_n = r_1^n w_0$. On a explicité v_n et w_n pour tout $n \in \mathbb{N}$, et on en déduit une expression de la suite originelle $(u_n)_{n\geqslant 0}$ en remarquant que l'on a :

$$\forall n \in \mathbb{N}, \quad u_n = \frac{v_n - w_n}{r_2 - r_1} = \frac{u_1 - r_1u_0}{r_2 - r_1}r_2^n + \frac{u_1 - r_2u_0}{r_1 - r_2}r_1^n.$$

On a explicité toute suite vérifiant la relation de récurrence d'ordre 2 ci-dessus, en démontrant qu'elle est nécessairement combinaison linéaire de $(r_1^n)_{n\geqslant 0}$ et $(r_2^n)_{n\geqslant 0}$ (la réciproque est facile à vérifier).

Fort du succès rencontré dans le cas des suites récurrentes linéaires d'ordre 2, on peut tenter de la transposer à un autre cadre proche : celui des équations différentielles linéaires d'ordre 2. Si l'on veut déterminer les applications $y : \mathbb{R} \to \mathbb{R}$ de classe C^2 sur $\mathbb{R}$ et vérifiant :

$$y'' + ay' + by = 0,$$

alors en posant : $f = y' - r_1y$, et : $g = y' - r_2y$, un calcul analogue à celui fait avec les suites permet de démontrer que l'on a : $f' = r_2f$, et : $g' = r_1g$ (faites-le). On a à nouveau *abaissé l'ordre* : f et g vérifient des équations différentielles linéaires du premier ordre. On sait les résoudre. On en déduit : $\forall t \in \mathbb{R}$, $f(t) = f(0)e^{r_2t}$, $g(t) = g(0)e^{r_1t}$, et on revient à la fonction inconnue d'origine en écrivant :

$$\forall t \in \mathbb{R}, \quad y(t) = \frac{f(t) - g(t)}{r_2 - r_1} = \frac{y'(0) - r_1y(0)}{r_2 - r_1}e^{r_2t} + \frac{y'(0) - r_2y(0)}{r_1 - r_2}e^{r_1t}.$$

On a démontré que toute solution de l'équation différentielle linéaire d'ordre 2 ci-dessus est combinaison linéaire des applications $t \mapsto e^{r_1t}$ et $t \mapsto e^{r_2t}$ (la réciproque est facile à vérifier).

Tout cela est très bien, mais pas entièrement satisfaisant :

— on n'est pas très efficace : va-t-on devoir recommencer cette stratégie chaque fois que l'on rencontre une équation linéaire d'ordre 2 ? (d'ailleurs, comment reconnaître une équation linéaire qui se résoudrait par cette approche ? est-ce applicable aux suites vérifiant $(n+2)u_{n+2}+n^2u_n = 0$ pour tout $n \in \mathbb{N}$? et à celles vérifiant $(n + 2)(n + 1)u_{n+2} + (n + 1)u_{n+1} + u_n = 0$ pour tout $n \in \mathbb{N}$? comment savoir ?) ne peut-on pas donner un énoncé général, dont les résolutions ci-dessus seraient des applications à des cas particuliers ?

— l'approche a tout l'air du bricolage : et si l'on rencontre une relation de récurrence ou une équation différentielle d'ordre 3, 4, etc., peut-on se ramener à des équations d'ordre inférieur ? si oui, comment ?

Je ne détaillerai pas le second point dans cette introduction, me contentant de dire qu'il y a bien un moyen de se ramener à des équations d'ordre inférieur, sans expliciter le « comment » (qui proviendra d'un résultat appelé le *lemme des noyaux*). Pour le premier point, nous allons expliquer dans les grandes lignes comment produire un énoncé général. Le bon cadre est celui de l'algèbre linéaire. De manière générale, lorsqu'on cherche à déterminer l'ensemble E des $\clubsuit$ vérifiant : $\bigstar = 0$, alors en posant g l'application $\clubsuit \mapsto \bigstar$, on note que l'on a :

$$E = \{\clubsuit \mid g(\clubsuit) = 0\} = \ker(g),$$

du moins si g est un morphisme (d'espaces vectoriels, de groupes, d'anneaux, etc., à ceci près que dans le cas d'un groupe le membre de droite doit éventuellement être 1, ou id, etc., selon le groupe). Cette réécriture a l'intérêt de reconnaître de la structure, et de pouvoir utiliser les énoncés généraux de vos cours sur les structures pour étudier E. Dans cette section, g est une application **linéaire**.

Dans le cas des deux exemples ci-dessus, on pourrait ainsi introduire les applications $(u_n)_{n\geqslant 0} \mapsto (u_{n+2}+au_{n+1}+bu_n)_{n\geqslant 0}$ et $y \mapsto y''+ay'+by$, mais pour correctement définir notre problématique (se ramener à des équations d'ordre plus petit : comment définir l'*ordre*, en toute généralité ?), il est plus pertinent d'introduire les deux applications linéaires suivantes, très fréquemment rencontrées dans $K^{\mathbb{N}}$ et $\mathrm{C}^\infty(I,K)$:

$$T : \begin{cases} \mathbb{R}^{\mathbb{N}} & \to & \mathbb{R}^{\mathbb{N}} \\ (u_n)_{n\geqslant 0} & \mapsto & (u_{n+1})_{n\geqslant 0} \end{cases}, \quad \text{et :} \quad D : \begin{cases} \mathrm{C}^\infty(\mathbb{R},\mathbb{R}) & \to & \mathrm{C}^\infty(\mathbb{R},\mathbb{R}) \\ y & \mapsto & y' \end{cases}$$

(je les appelle T comme « translation » et D comme « dérivation »). Avec ces notations, je vous laisse vérifier que :

$$\left\{(u_n)_{n\geqslant 0} \in \mathbb{R}^{\mathbb{N}} \mid \forall n \in \mathbb{N},\ u_{n+2}+au_{n+1}+bu_n = 0\right\} = \ker\left(T^2+aT+b\mathrm{Id}_{\mathbb{R}^{\mathbb{N}}}\right)$$

(pour rappel : $T^2 = T\circ T$; plus généralement, T^k se définit par récurrence *via* la relation : $T^{k+1} = T^k \circ T$, et de même pour tout endomorphisme), et :

$$\{y \in \mathrm{C}^\infty(\mathbb{R},\mathbb{R}) \mid y''+ay'+by = 0\} = \ker\left(D^2+aD+b\mathrm{Id}_{\mathrm{C}^\infty(\mathbb{R},\mathbb{R})}\right).$$

Les deux ensembles que nous cherchions ont donc pour point commun de s'écrire sous la forme :

$$\ker\left(f^2+af+b\mathrm{Id}_E\right),$$

où f est un endomorphisme d'un espace vectoriel E. On peut parler dans chaque cas d'une équation *linéaire* (c'est le noyau d'une application linéaire), d'ordre 2 puisque c'est le noyau d'un « polynôme d'endomorphisme » (nous ferons un jour tomber les guillemets) de degré 2.

Dans ce contexte, la problématique de cette section est alors d'écrire ce noyau « mal connu » (on veut expliciter ses éléments) comme somme de noyaux de la forme $\ker(f-\alpha\mathrm{Id}_E)$ (qui donnent des solutions d'une équation d'ordre inférieur, et donc plus « simple »). Le lien avec l'introduction est fait dans ce cas particulier (avec $E = \ker\left(f^2+af+b\mathrm{Id}_E\right)$ et les F_i de la forme $\ker(f-\alpha\mathrm{Id}_E)$).

Pour y parvenir : en s'inspirant des deux cas particuliers ci-dessus, on démontre que l'on a : $\ker\left(f^2+af+b\mathrm{Id}_E\right) = \ker\left(f-r_1\mathrm{Id}_E\right)\oplus\ker\left(f-r_2\mathrm{Id}_E\right)$, ce qui abaisse effectivement l'ordre, puisque les deux noyaux du membre de droite n'ont plus de f^2 ! C'est valable pour tout endomorphisme, donc c'est en particulier valable pour T et D définis ci-dessus. Or :

$$\forall r \in \mathbb{R}, \quad \ker\left(T-r\mathrm{Id}_{\mathbb{R}^{\mathbb{N}}}\right) = \left\{(u_n)_{n\geqslant 0} \in \mathbb{R}^{\mathbb{N}} \mid \forall n \in \mathbb{N},\ u_{n+1}-ru_n = 0\right\} = \mathrm{Vect}\left((r^n)_{n\geqslant 0}\right),$$

et :

$$\forall r \in \mathbb{R}, \quad \ker\left(D-r\mathrm{Id}_{\mathrm{C}^\infty(\mathbb{R},\mathbb{R})}\right) = \{y \in \mathrm{C}^\infty(\mathbb{R},\mathbb{R}) \mid y'-ry = 0\} = \mathrm{Vect}\left(x \mapsto e^{rx}\right),$$

donc :

$$\ker\left(T^2+aT+b\mathrm{Id}_{\mathbb{R}^{\mathbb{N}}}\right) = \ker\left(T-r_1\mathrm{Id}_{\mathbb{R}^{\mathbb{N}}}\right)\oplus\ker\left(T-r_2\mathrm{Id}_{\mathbb{R}^{\mathbb{N}}}\right) = \mathrm{Vect}\left((r_1^n)_{n\geqslant 0}, (r_2^n)_{n\geqslant 0}\right),$$

et de même : $\ker\left(D^2+aD+b\mathrm{Id}_{\mathrm{C}^\infty(\mathbb{R},\mathbb{R})}\right) = \mathrm{Vect}\left(x\mapsto e^{r_1x}, x\mapsto e^{r_2x}\right)$, ce qui permet de retrouver la forme des suites récurrentes linéaires et des solutions d'équations différentielles obtenue ci-dessus.

Mission accomplie ! On a bien posé le problème (ramener une équation linéaire *quelconque* d'ordre 2 à des équations d'ordre 1), on l'a résolu en ajoutant une hypothèse sur des racines, et il ne restait plus qu'à l'appliquer directement à des endomorphismes bien choisis pour obtenir la forme des suites récurrentes d'ordre 2 et des solutions d'une équation différentielle linéaire d'ordre 2, *en même temps* * ! Et le fait d'avoir proposé une approche générale permet de l'appliquer à toutes les équations du même type que vous rencontrerez : voir l'exercice 32 ci-dessous.

Exercice 32. Soit $(a, b) \in \mathbb{R}^2$.

1. Soit f un endomorphisme d'un espace vectoriel réel E. On suppose : $f^2 + af + b = 0_{\mathrm{L}(E)}$ (cette hypothèse équivaut à : $E = \ker\left(f^2 + af + b\mathrm{Id}_E\right)$), et que le polynôme $X^2 + aX + b$ admet deux racines réelles distinctes r_1 et r_2.
   Démontrer effectivement ce qui a été omis ci-dessus : $E = \ker(f - r_1\mathrm{Id}_E) \oplus \ker(f - r_2\mathrm{Id}_E)$.
2. Soient E l'espace vectoriel des suites réelles vérifiant : $\forall n \in \mathbb{N}$, $u_{n+2} + au_{n+1} + bu_n = 0$, et f l'endomorphisme $(u_n)_{n\geqslant 0} \mapsto (u_{n+1})_{n\geqslant 0}$. On suppose uniquement dans cette question que le polynôme $X^2 + aX + b$ admet une unique racine réelle r. Montrer que l'égalité : $E = \ker(f - r\mathrm{Id}_E)$ est impossible (*si possible*, se passer de votre connaissance *a priori* des relations de récurrence d'ordre 2).

On utilisera le résultat de la question 1 pour traiter les suivantes.

3. Déterminer les applications $y : \left]0, \frac{\pi}{2}\right[\to \mathbb{R}$ de classe C^∞ sur $\left]0, \frac{\pi}{2}\right[$ et vérifiant :
$$(\tan\cdot(\tan\cdot y'))' = y.$$
4. On suppose que le polynôme $X^2 + aX + b$ admet deux racines réelles distinctes r_1 et r_2. Déterminer les applications $y : \mathbb{R} \to \mathbb{R}$ de classe C^∞ sur $\mathbb{R}$ vérifiant :
$$\forall x \in \mathbb{R}, \quad y''(x) + (2x + a)y'(x) + \left(x^2 + 1 + ax + b\right) y(x) = 0.$$

Exercice 33. Soit $(a, b) \in \mathbb{R}^2$. On se place dans le cas où $X^2 + aX + b$ admet une unique racine réelle.

1. Toujours par des techniques d'algèbre linéaire, déterminer les suites réelles vérifiant : $\forall n \in \mathbb{N}$, $u_{n+2} + au_{n+1} + bu_n = 0$. On n'utilisera pas notre connaissance *a priori* des suites récurrentes linéaires d'ordre 2 pour traiter cette question.
2. De même pour les applications $y : \mathbb{R} \to \mathbb{R}$ de classe C^∞ sur $\mathbb{R}$ telles que : $y'' + ay' + by = 0$.

Exercice 34. Soit $(a, b, c) \in \mathbb{R}^3$. On suppose que le polynôme $X^3 + aX^2 + bX + c$ admet une racine réelle simple μ et une racine réelle double λ.

1. Soit f un endomorphisme d'un espace vectoriel réel E, vérifiant : $f^3 + af^2 + bf + c\mathrm{Id}_E = 0_{\mathrm{L}(E)}$. Montrer :
$$E = \ker(f - \mu\mathrm{Id}_E) \oplus \ker\left((f - \lambda\mathrm{Id}_E)^2\right).$$
2. En déduire la forme explicite des suites réelles $(u_n)_{n\geqslant 0}$ vérifiant :
$$\forall n \in \mathbb{N}, \quad u_{n+3} + au_{n+2} + bu_{n+1} + cu_n = 0,$$
   et des applications $y : \mathbb{R} \to \mathbb{R}$ de classe C^∞ sur $\mathbb{R}$ vérifiant : $y^{(3)} + ay'' + by' + cy = 0$.
3. Traiter tous les autres cas possibles, selon les ordres de multiplicité des racines du polynôme $X^3 + aX^2 + bX + c$ (ne pas oublier le cas où ce polynôme n'est pas scindé).

Le dernier exercice semble indiquer que la méthode semble s'adapter aux ordres supérieurs strictement à 2. En fait, nous montrerons un résultat important, *le lemme des noyaux*, qui permet d'écrire tout noyau de « polynôme en f » comme une somme directe de noyaux de polynômes de degré inférieur (sauf si le polynôme en f initial est irréductible), qui proviennent de sa décomposition en facteurs irréductibles. Ainsi la méthode se généralise effectivement et permettra de donner les solutions de toute équation différentielle à coefficients constants, ainsi que la forme de toutes les suites récurrentes linéaires à coefficients constants. Et comme on le voit dans l'exercice 32, changer l'endomorphisme permet de résoudre bien d'autres équations !

*. On n'expose là qu'un seul des très nombreux charmes et apports des *structures*.

6.2 Pour simplifier les calculs impliquant un endomorphisme

La théorie de la réduction se justifie naturellement aussi bien par la géométrie que par l'algèbre matricielle.

Il n'est pas nécessaire de préciser l'apport gigantesque qui a été fait aux mathématiques en introduisant la notion de repère (ou de base), afin de raisonner avec des coordonnées. Seulement, les calculs peuvent s'avérer difficiles si le repère est mal choisi. Par exemple, quand on fait tourner des objets autour d'un axe (et qu'il y a donc une rotation en jeu), n'importe quelle personne sensée préférerait que l'axe de la rotation soit aussi l'un des axes du repère choisi. De la même manière, une réflexion en dimension 3 est plus facile à étudier si son plan de réflexion est engendré par des axes du repère.

À l'inverse, si l'on a affaire à une transformation linéaire non identifiée, on la reconnaîtra mieux si son action sur des vecteurs directeurs des axes est simple. Prenons un exemple concret. Munissons $\mathbb{R}^2$ de la base canonique $\mathcal{B}_c = (\vec{i}, \vec{j})$, et soit $f : \mathbb{R}^2 \to \mathbb{R}^2$ l'application *linéaire* définie par :

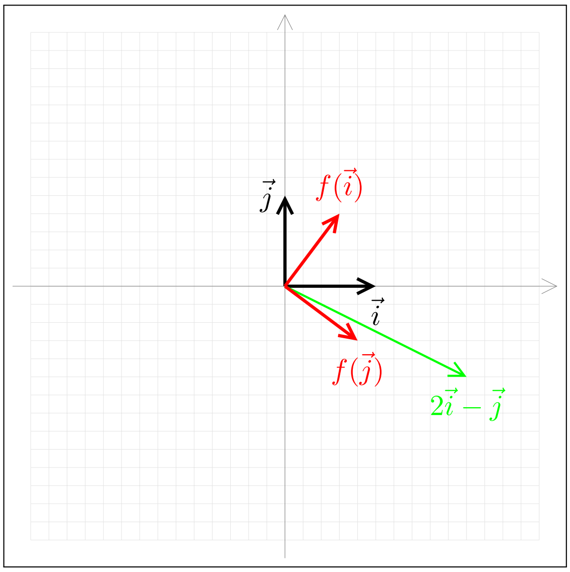
$$f(\vec{i}) = \frac{3}{5}\vec{i} + \frac{4}{5}\vec{j}, \quad \text{et :} \quad f(\vec{j}) = \frac{4}{5}\vec{i} - \frac{3}{5}\vec{j}.$$

Définir une application linéaire sur une base suffit à la déterminer en tout vecteur. En effet, pour tout $(x, y) \in \mathbb{R}^2$, on a :

$$f((x,y)) = f(x\vec{i}+y\vec{j}) = xf(\vec{i})+yf(\vec{j}) = x\left(\frac{3}{5}\vec{i}+\frac{4}{5}\vec{j}\right)+y\left(\frac{4}{5}\vec{i}-\frac{3}{5}\vec{j}\right) = \left(\frac{3}{5}x+\frac{4}{5}y\right)\vec{i}+\left(\frac{4}{5}x-\frac{3}{5}y\right)\vec{j}.$$

Cela permet de calculer l'image par f de tout vecteur (d'ailleurs, même sans cette expression analytique, vous y parviendriez : le faire dans l'exercice suivant).

Exercice 35. Sur la figure ci-dessous, dessiner l'image par f de $2\vec{i} - \vec{j}$. Vous pouvez aussi de vous-mêmes dessiner d'autres vecteurs $\vec{u}$ et leurs images par f.



Même si l'on sait en théorie dessiner l'image par f de tout vecteur, le procédé est laborieux. Nous proposons une autre base où f est plus facile à calculer. Posons :

$$\vec{e}_1 = 2\vec{i} + \vec{j} \quad \text{et :} \quad \vec{e}_2 = \vec{i} - 2\vec{j}.$$

On admet que $\mathcal{B} = (\vec{e}_1, \vec{e}_2)$ est une base. Alors :

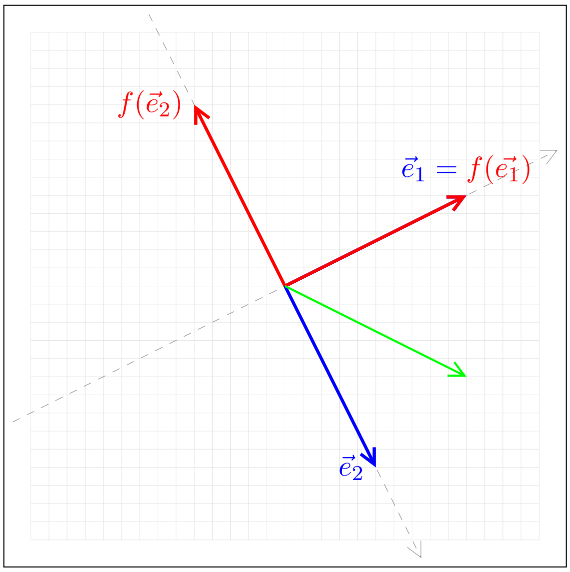
$$f(\vec{e}_1) = 2\vec{i} + \vec{j} = \vec{e}_1, \quad \text{et :} \quad f(\vec{e}_2) = -\vec{i} + 2\vec{j} = -\vec{e}_2,$$

ce qui facilite le calcul de l'image par f de tout vecteur du plan. Si $\vec{u} = x\vec{e}_1 + y\vec{e}_2$, alors :

$$f(\vec{u}) = f(x\vec{e}_1 + y\vec{e}_2) = x\vec{e}_1 - y\vec{e}_2.$$

Je vous laisse apprécier la simplicité des représentations dans cette nouvelle base, avec l'exercice suivant :

Exercice 36. Sur la figure ci-dessous, dessiner à nouveau l'image par f du vecteur $2\vec{\imath}-\vec{\jmath}$ (le vecteur vert, dont les coordonnées dans la base $(\vec{e}_1, \vec{e}_2)$ sont : $\left(\frac{3}{5}, \frac{4}{5}\right)$, mais vous n'en avez normalement pas besoin pour votre dessin). Dessinez ensuite de vous-mêmes d'autres vecteurs $\vec{u}$ et leurs images par f. Qu'en pensez-vous ?



Si vous vous y prenez bien, vous reconnaîtrez en f la réflexion par rapport à la droite dirigée par $\vec{e}_1$, tout simplement ! Voyez comment tout peut dépendre du bon choix de base.

Ainsi une application linéaire décrite en termes géométriques simples peut avoir une expression obscure dans la base canonique. Ce problème se lit aussi sur les matrices de f dans les bases $\mathcal{B}_c$ et $\mathcal{B}$; on a en effet :

$$\mathrm{M}_{\mathcal{B}_c}(f) = \begin{pmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{pmatrix}, \text{ et } \mathrm{M}_{\mathcal{B}}(f) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

et il est manifeste que $\mathrm{M}_{\mathcal{B}}(f)$ est plus simple à étudier que $\mathrm{M}_{\mathcal{B}_c}(f)$, du fait qu'elle soit *diagonale* : ses puissances, son déterminant, son noyau, etc., se calculeraient plus rapidement. On passe de l'une à l'autre *via* la formule du changement de base appliquée à f, entre les bases $\mathcal{B}_c$ et $\mathcal{B}$:

$$\begin{pmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{pmatrix} = P \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} P^{-1}, \quad \text{où :} \quad P = \mathrm{M}_{\mathcal{B}_c}(\mathcal{B}) = \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}.$$

Cette relation complètement explicite montre comment, après avoir achevé une étude dans la base $\mathcal{B}$, il n'est pas difficile d'en déduire ce qu'il se passe dans la base $\mathcal{B}_c$. Ainsi on ne perd aucune information en changeant de base.

La question qu'on se pose, à la lumière de cet exemple, est : étant donné un endomorphisme quelconque en dimension finie, est-il possible de déterminer une base où sa matrice est simple (« réduite »), en particulier *diagonale* ? Et si oui, comment déterminer cette base ? C'est cela qu'on appelle la *diagonalisation* d'un endomorphisme.

Nous avons formulé la question pour les endomorphismes. La question de diagonalisation pour les matrices est : étant donnée une matrice carrée, existe-t-il une matrice *diagonale* qui lui soit *semblable* ? Si oui, comment la déterminer ?

Il s'avère que ce n'est pas toujours possible (et nous aurons des critères simples pour trancher la question), et dans le cas contraire nous essaierons, faute de mieux, d'avoir d'autres réductions : la réduction à des matrices triangulaires supérieures est la seule au programme (en dehors de la réduction aux matrices diagonales), mais il y en a d'autres que nous croiserons en exercices (réduction de Dunford, Jordan, Frobenius, etc.).

Les exercices suivants permettent d'avoir une idée de ce qui permet ou non de *diagonaliser*.

Exercice 37.

1. Montrer que la matrice $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ n'est pas semblable à une matrice diagonale (*indication : montrer, grâce à la trace et au déterminant par exemple, ou par un bon usage du rang, que si c'était le cas, les coefficients diagonaux d'une matrice diagonale semblable seraient* 1 *et* 1 *; en déduire une contradiction*).
2. Généraliser cet exemple : donner pour tout $n \in \mathbb{N} \setminus \{0,1\}$ une famille infinie de matrices de $\mathrm{M}_n(\mathbb{C})$ qui ne sont pas semblables à une matrice diagonale.

Exercice 38. Soit $(\lambda_1, \ldots, \lambda_n) \in \mathbb{C}^n$ un n-uplet de scalaires DISTINCTS. Soit f un endomorphisme de $\mathbb{C}^n$. On suppose que pour tout $i \in [\![1, n]\!]$, il existe un vecteur non nul $\vec{e}_i \in \mathbb{C}^n$ tel que : $f(\vec{e}_i) = \lambda_i \vec{e}_i$. On veut montrer que la famille $(\vec{e}_i)_{1\leqslant i\leqslant n}$ est libre. Soit $(\alpha_i)_{1\leqslant i\leqslant n} \in \mathbb{C}^n$; on suppose : $\displaystyle\sum_{i=1}^{n} \alpha_i \vec{e}_i = \vec{0}$.

1. Montrer : $\forall k \in \mathbb{N}$, $\displaystyle\sum_{i=1}^{n} \alpha_i \lambda_i^k \vec{e}_i = \vec{0}$.
2. En déduire : $\forall P \in \mathbb{C}[X]$, $\displaystyle\sum_{i=1}^{n} \alpha_i P(\lambda_i) \vec{e}_i = \vec{0}$.
3. Conclure grâce à des polynômes interpolateurs bien choisis.
4. Écrire la matrice représentative de f dans la base $(\vec{e}_i)_{1\leqslant i\leqslant n}$. Qu'avons-nous démontré dans cet exercice ?

L'exercice précédent a l'intérêt de vous faire jouer avec des notions majeures : les scalaires λ_i et les vecteurs non nuls $\vec{e}_i$ vérifiant : $f(\vec{e}_i) = \lambda_i \vec{e}_i$, sont respectivement appelés des *valeurs propres* et des *vecteurs propres* de f. Ce sont des objets cruciaux dans l'étude de n'importe quel endomorphisme et ils sont très riches en informations. Il n'est pas exagéré de dire que, peu importe votre cursus scientifique à venir, vous en aurez un usage abondant.

Les situations où il est plus intéressant de se ramener à une matrice diagonale ou, faute de mieux, triangulaire, sont nombreuses (en fait, on pourrait presque dire que *toute* situation s'y prête mieux). Les applications de la réduction ne peuvent donc pas être données exhaustivement. En voici quelques-unes néanmoins :

— l'étude des suites récurrentes linéaires et des équations différentielles (suivant un point de vue complémentaire à celui de la section 6.1) ; ← page 18

— la résolution d'équations matricielles ;

— l'étude du comportement asymptotique d'une suite ou d'une série de matrices.

L'exercice suivant permet de diagonaliser une matrice d'ordre 3 et d'illustrer deux de ces trois items :

Exercice 39. On pose : $A = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}$. Soit f l'endomorphisme de $\mathbb{R}^3$ canoniquement associé à A.

1. Trouver un vecteur non nul $\vec{x}_1 \in \mathbb{R}^3$ tel que : $f(\vec{x}_1) = 5\vec{x}_1$, et deux vecteurs de $\mathbb{R}^3$ *linéairement indépendants* $\vec{x}_2$ et $\vec{x}_3$ tels que : $f(\vec{x}_2) = -\vec{x}_2$, $f(\vec{x}_3) = -\vec{x}_3$.
2. Montrer que $\mathcal{B} = (\vec{x}_1, \vec{x}_2, \vec{x}_3)$ est une base de $\mathbb{R}^3$, et écrire la matrice représentative de f relativement à $\mathcal{B}$.
3. En déduire qu'il existe $P \in \mathrm{GL}_3(\mathbb{R})$ telle que : $A = PDP^{-1}$, avec : $D = \begin{pmatrix} 5 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$. *Utiliser la formule du changement de base.*

4. Calculer D^n pour tout $n \in \mathbb{N}$, et en déduire une expression de A^n pour tout $n \in \mathbb{N}$.
5. Déterminer les suites réelles $(u_n)_{n\geqslant 0}$, $(v_n)_{n\geqslant 0}$ et $(w_n)_{n\geqslant 0}$ vérifiant :

$$\forall n \in \mathbb{N}, \quad \begin{cases} u_{n+1} &= u_n + 2v_n + 2w_n, \\ v_{n+1} &= 2u_n + v_n + 2w_n, \\ w_{n+1} &= 2u_n + 2v_n + w_n. \end{cases}$$

Indication : on introduira le vecteur colonne X_n défini pour tout $n \in \mathbb{N}$ par : $X_n = \begin{pmatrix} u_n \\ v_n \\ w_n \end{pmatrix}$, puis on commencera par expliciter les coordonnées de $P^{-1}X_n$ pour tout $n \in \mathbb{N}$.

6.3 Résumé : objectif du chapitre

Nous allons dans un premier temps étendre les résultats sur les sommes de sous-espaces vectoriels, afin de donner un sens à la somme directe : $E = \bigoplus_{i=1}^{k} F_i$, quand $k \geqslant 3$. En effet, selon la complexité de l'espace vectoriel E, il ne suffira pas de le décomposer en une somme directe de seulement deux sous-espaces vectoriels. Puis nous formulerons l'important lemme des noyaux, qui donne un moyen très efficace d'obtenir de telles décompositions lorsque E est un noyau (et c'est le cas si E est l'espace des solutions d'une équation linéaire homogène).

Puisque ces noyaux sont souvent des « polynômes en un endomorphisme f », nous devrons introduire la notion, ainsi que celle de polynôme *annulateur* (ce sont des polynômes en f dont le noyau donne tout l'espace) et donner des résultats de base sur ces polynômes. Un exemple important de polynôme annulateur est le polynôme minimal, que nous définirons.

Outre la motivation ci-dessus, les polynômes en f ont le bon goût de commuter avec f : la commutation est rare et précieuse avec les endomorphismes et matrices, donc la rencontrer a un intérêt particulier.

Viendra ensuite l'importante problématique de la réduction, qui survolera toute l'algèbre linéaire de cette année (et s'invitera même épisodiquement dans les autres chapitres). Nous introduirons les importantes notions de *valeurs propres*, *vecteurs propres*, *sous-espaces propres* et *polynômes caractéristiques* qui donnent un moyen pratique voire algorithmique de réduction des endomorphismes et matrices. De nombreux critères seront donnés pour déterminer si un endomorphisme ou une matrice est diagonalisable ou non ; lorsque ces critères ne sont pas vérifiés, nous devrons nous contenter d'exprimer un endomorphisme comme somme d'un endomorphisme diagonalisable et d'un endomorphisme nilpotent (décomposition de Dunford) : ainsi, moralement, quand on comprend les endomorphismes diagonalisables et les nilpotents, on comprend tout endomorphisme. C'est pourquoi ces derniers seront aussi l'objet de notre attention.

Nous donnerons de nombreuses applications, dont celles citées en fin de section précédente.

7 Topologie des espaces vectoriels normés

7.1 Faire de l'analyse en dimension supérieure

L'objectif du chapitre est de fournir un bon cadre théorique pour faire de l'analyse dans d'autres espaces que $\mathbb{R}$ ou $\mathbb{C}$ (sachant que même l'analyse dans $\mathbb{C}$ est très restreinte en MPSI : elle n'apparaît que dans les ensembles *d'arrivée* des fonctions que vous étudiiez jusqu'à présent).

Le point de départ de l'analyse est la notion de taille (je parlerai aussi, indifféremment, de longueur) ou de distance : d'abord pour parler de suite ou fonction bornée (ce qui n'est pas la plus subtile des notions de l'analyse, certes), mais surtout pour parler de limite : pour dire qu'on se « rapproche indéfiniment » d'un point, ou qu'on devient « arbitrairement grand » dans le cas d'une limite infinie, encore faut-il définir ce qu'est la *proximité* à un point, ce qu'est être arbitrairement grand (ou petit). Une fois qu'on a défini la notion de limite, toutes les autres notions de l'analyse s'ensuivent puisqu'elles sont souvent définies à l'aide d'une limite : continuité, dérivabilité, etc.

Ainsi, pour faire de l'analyse dans d'autres espaces vectoriels que $\mathbb{R}$ ou $\mathbb{C}$, il faut pouvoir y définir la taille d'un vecteur. Vous avez déjà quelques objets qui le permettent : la notion de norme euclidienne d'un vecteur, d'abord (et très manipulée en Physique, plus particulièrement dans $\mathbb{R}^2$ ou $\mathbb{R}^3$), mais aussi la norme infinie d'une fonction, justement introduite dans le chapitre préliminaire pour définir la proximité entre deux fonctions, puis le fait qu'une suite de fonctions $(f_n)_{n\geqslant 0}$ converge vers une fonction f. Plus généralement, on appelle *norme* sur un espace vectoriel E toute notion de taille qui permet de faire de l'analyse dans E.

Pour comprendre ce que doit vérifier une norme sur E pour accomplir ce que l'on souhaite (faire de l'analyse dans E), réfléchissons par analogie avec la valeur absolue sur $\mathbb{R}$, puisque c'est elle qui définit la taille d'un réel ou la proximité entre deux réels (la distance entre a et b est $|a-b|$). Elle vérifie les propriétés suivantes :

— elle est positive (chose naturelle pour une notion de longueur ou de distance !) ;

— seul zéro est de valeur absolue nulle, ce qui implique en particulier qu'il y a toujours une distance strictement positive entre deux réels *distincts* (vu que : $|a-b|=0 \iff a=b$) ; en cela, on dit que la valeur absolue « sépare les points » ;

— elle vérifie l'inégalité triangulaire, ce qui implique concrètement une sorte de transitivité de la notion de proximité : si a est arbitrairement proche de b et b arbitrairement proche de c, alors a est arbitrairement proche de c (grâce à : $|a-c| \leqslant |a-b|+|b-c|$) ;

— elle est multiplicative : $\forall (a,b)\in\mathbb{R}^2$, $|ab|=|a|\cdot|b|$, ce qui assure que si $f(x)$ est « infiniment petit », alors $\lambda f(x)$ l'est aussi pour toute constante réelle λ (cette propriété est *utile*, mais pas *indispensable* pour faire de l'analyse).

Quand on regarde de près les démonstrations des résultats fondamentaux de l'analyse, seules ces propriétés de la valeur absolue interviennent :

Exercice 40. Revoir les démonstrations des résultats suivants (c'est encore mieux si vous savez les reproduire), et observer lesquelles des propriétés ci-dessus sont utilisées :

1. Si une suite converge, sa limite est unique.
2. Si $(u_n)_{n\geqslant 0}$ et $(v_n)_{n\geqslant 0}$ convergent vers ℓ_1 et ℓ_2 respectivement, alors pour tout $(a,b)\in\mathbb{R}^2$ la suite $(au_n+bv_n)_{n\geqslant 0}$ converge vers $a\ell_1+b\ell_2$.
3. Si une suite converge, alors elle est bornée.

Ainsi il est raisonnable de définir une norme $\|\cdot\| : E\to\mathbb{R}$ comme une fonction sur le K-espace vectoriel E vérifiant des propriétés analogues :

— la *positivité* : $\forall \vec{x}\in E$, $\|\vec{x}\|\geqslant 0$;

— la propriété de *séparation* des points : $\forall \vec{x}\in E$, $\|\vec{x}\|=0 \iff \vec{x}=\vec{0}$ (qui implique aisément : $\forall(\vec{x},\vec{y})\in E^2$, $\|\vec{x}-\vec{y}\|>0 \iff \vec{x}\neq\vec{y}$, d'où la terminologie) ;

— l'*inégalité triangulaire* : $\forall(\vec{x},\vec{y})\in E^2$, $\|\vec{x}+\vec{y}\|\leqslant\|\vec{x}\|+\|\vec{y}\|$;

— la propriété d'*homogénéité* : $\forall(\lambda,\vec{x})\in K\times E$, $\|\lambda\vec{x}\|=|\lambda|\cdot\|\vec{x}\|$.

Il y a en vérité quelques redondances dans cette définition (voir l'exercice ci-dessous) qui sont faites par confort, pour simplifier la manipulation d'une fonction qu'on nous présente déjà comme une norme.

Exercice 41. Soit $\|\cdot\| : E\to\mathbb{R}$ une application définie sur un K-espace vectoriel E.

1. Montrer que si $\|\cdot\|$ vérifie la propriété d'homogénéité, alors : « $\vec{x}=\vec{0}\implies\|\vec{x}\|=0$ » est vrai.
2. Montrer que si $\|\cdot\|$ vérifie l'inégalité triangulaire et la propriété d'homogénéité, alors elle est positive. On commencera par montrer :
   — $\forall\vec{x}\in E$, $\|\vec{x}\|=\|-\vec{x}\|$;
   — $\forall(\vec{x},\vec{y})\in E^2$, $\big|\|\vec{x}\|-\|\vec{y}\|\big|\leqslant\|\vec{x}-\vec{y}\|$ (inégalité triangulaire renversée).

Une fois qu'un espace vectoriel E est muni d'une norme $\|\cdot\|$, on reprend toutes les définitions de l'analyse réelle en remplaçant les valeurs absolues de réels par des normes lorsqu'il y a des vecteurs, et le tour est joué. Par exemple, une suite $(\vec{u}_n)_{n\geqslant 0}$ à valeurs dans E converge vers $\vec{\ell}\in E$ si :

$$\forall \varepsilon\in\mathbb{R}_+^*,\ \exists N\in\mathbb{N},\ \forall n\in\mathbb{N},\quad \left(n\geqslant N\Longrightarrow \left\|\vec{u}_n-\vec{\ell}\right\|\leqslant\varepsilon\right).$$

Exercice 42.

1. Avec ces définitions de norme et de convergence, vérifier que tous les résultats de l'exercice 40 restent valables dans un K-espace vectoriel E muni d'une norme $\|\cdot\|$ (on remplacera simplement $(a,b)\in\mathbb{R}^2$ par $(a,b)\in K^2$, et on dira qu'une suite $(\vec{u}_n)_{n\geqslant 0}$ à valeurs dans E est bornée s'il existe $r\in\mathbb{R}_+$ tel que : $\forall n\in\mathbb{N}$, $\|\vec{u}_n\|\leqslant r$).
2. Montrer que si $(\alpha_n)_{n\geqslant 0}\in K^{\mathbb{N}}$ et $(\vec{u}_n)_{n\geqslant 0}\in E^{\mathbb{N}}$ convergent, alors $(\alpha_n\vec{u}_n)_{n\geqslant 0}$ également.
3. Démontrer que si une suite converge, alors toutes ses suites extraites convergent la même limite.
4. Montrer que si $(\vec{u}_n)_{n\geqslant 0}$ converge vers $\vec{\ell}$, alors $(\|\vec{u}_n\|)_{n\geqslant 0}$ converge vers $\|\vec{\ell}\|$.

On définira ensuite la limite d'une fonction, la continuité en un point, le caractère lipschitzien, etc. La dérivabilité attendra (chapitres *Dérivation et intégration des fonctions vectorielles* et *Calcul différentiel*). Puisque nous manipulerons des fonctions qui ne sont pas de la variable réelle, nous aurons besoin de nouvelles fonctions continues de référence, desquelles découleront toutes les fonctions continues que nous manipulerons (par somme, composition, etc.). En particulier, nous donnerons des conditions suffisantes pour qu'une application linéaire ou multilinéaire soit continue.

→ page 49
→ page 59

Bien entendu, nous donnerons des exemples explicites de normes (et donc d'espaces vectoriels sur lesquels on peut faire de l'analyse). Cela inclut tous les espaces vectoriels de dimension finie (ainsi l'analyse matricielle reviendra en plusieurs endroits : voir par exemple l'exponentielle matricielle introduite dans la section 11.2), mais aussi des espaces de fonctions, de polynômes, etc. Sur un même espace vectoriel E nous serons parfois amenés à définir plusieurs normes, certaines étant plus pratiques que d'autres selon les situations.

→ page 50

Exercice 43.

1. Soit $n\in\mathbb{N}\setminus\{0\}$. Pour tout $\vec{x}=(x_i)_{1\leqslant i\leqslant n}\in K^n$, on pose :

$$\|\vec{x}\|_1=\sum_{k=1}^{n}|x_i|,\quad \|\vec{x}\|_2=\sqrt{\sum_{k=1}^{n}|x_i|^2},\quad \|\vec{x}\|_\infty=\max_{1\leqslant i\leqslant n}|x_i|.$$

   Vérifier que $\|\cdot\|_1$, $\|\cdot\|_2$ et $\|\cdot\|_\infty$ sont des normes sur K^n.
2. Soit $(a,b)\in\mathbb{R}^2$ tel que : $a<b$. Soit $E=\mathrm{C}^0([a,b],K)$. Pour toute fonction $f\in E$, on pose :

$$\|f\|_1=\int_a^b|f|,\quad \|f\|_2=\sqrt{\int_a^b|f|^2},\quad \|f\|_\infty=\sup_{[a,b]}|f|.$$

   Vérifier que $\|\cdot\|_1$, $\|\cdot\|_2$ et $\|\cdot\|_\infty$ sont des normes sur E.

La norme $\|\cdot\|_2$ sur K^n est essentielle quand on veut faire de la géométrie, mais sa racine carrée peut être embêtante dans les calculs : la norme $\|\cdot\|_1$ peut alors être préférable en cela, voire la norme $\|\cdot\|_\infty$ pour ne travailler qu'avec une seule coordonnée à la fois. On changera donc de choix de norme selon nos besoins !

7.2 Généraliser les intervalles, les segments, etc.

Pour les théorèmes de 1re année portant sur les fonctions de la variable réelle, le domaine de définition, de continuité, de dérivabilité, etc., joue souvent un rôle décisif :

— le théorème des bornes atteintes est faux si l'on remplace un segment par un intervalle non borné ou ouvert en une extrémité, de même pour le théorème de Heine ;

— les fonctions dérivables de dérivée nulle ne sont plus nécessairement constantes si l'on remplace un intervalle par une réunion disjointe d'intervalles (ou des parties de $\mathbb{R}$ encore moins régulières) ; de même pour le théorème des valeurs intermédiaires ;

— les extremums ne sont pas toujours en des points critiques si l'on remplace un intervalle ouvert par un segment ;

et ainsi de suite.

Exercice 44. Vérifier la justesse de ces affirmations à l'aide de contre-exemples dans chaque cas.

Si l'on veut généraliser les théorèmes de 1re année au cas des fonctions définies sur un espace vectoriel muni d'une norme, il faut donc aussi généraliser les notions de segment, d'intervalle ouvert, d'intervalle tout simplement, etc. C'est en comprenant quelles propriétés essentielles de ces ensembles sont utilisées dans les théorèmes de 1re année, que nous saurons trouver leurs analogues en dimension supérieure. Ainsi :

— pour généraliser la notion d'intervalle, nous proposerons la notion de parties *convexes* ou *connexes par arcs* ;

— pour généraliser la notion d'intervalle fermé, d'intervalle ouvert, nous définirons les parties *ouvertes* et *fermées* ;

— pour généraliser la notion de segment, nous définirons les parties *compactes* ;

— enfin, nous définirons les parties *denses* par analogie avec les parties denses de $\mathbb{R}$.

La plupart des théorèmes de 1re année auront alors leur analogue en dimension supérieure : le théorème des bornes atteintes est vrai en remplaçant les segments par des compacts ; le théorèmes des valeurs intermédiaires se généralise en remplaçant les intervalles par les parties connexes par arcs ; et ainsi de suite.

Si la notion de densité fut assez peu utilisée dans $\mathbb{R}$, elle fait preuve d'une efficacité remarquable dans d'autres contextes : c'est d'ailleurs par densité des fonctions en escalier que l'on définit l'intégrale d'une fonction continue par morceaux en 1re année. C'est-à-dire : 1° on définit l'intégrale d'une fonction en escalier, ce qui est facile puisque c'est simplement une somme d'aires de rectangles, 2° on montre que toute fonction f continue par morceaux sur un segment $[a, b]$ est limite uniforme d'une suite de fonctions en escalier $(f_n)_{n\geqslant 0}$ (même si vous n'utilisiez pas le terme de « limite uniforme » à l'époque), 3° on définit l'intégrale de f comme la limite de la suite d'intégrales $\left(\int_a^b f_n\right)_{n\geqslant 0}$ (en s'assurant que cette limite existe). La clé du raisonnement est la densité des fonctions en escalier.

De même, nous pourrons montrer des identités en commençant par des cas particuliers « simples » et étendre ces identités par passage à la limite, à condition que ces objets « simples » forment une partie dense. Une situation typique est en algèbre linéaire, où nous verrons que les matrices inversibles forment une partie dense de l'ensemble des matrices : un certain nombre de propriétés peuvent alors se démontrer en supposant que les matrices sont inversibles, et en passant à la limite pour avoir le résultat pour toute matrice.

7.3 Cas particulier de la dimension finie

Suite aux explications de la section précédente, il y a deux questions légitimes qu'on peut se poser :

— est-on vraiment obligé de se compliquer la vie à ce point pour calculer des limites dans un espace vectoriel, alors qu'il semble y avoir bien plus intuitif (en tout cas en dimension finie) ? En effet, si l'on introduisait les suites suivantes :

$$\forall n \in \mathbb{N}, \quad \vec{u}_n = \left(\arctan(n), \frac{n+1}{n+2}, e^{-2n}\right), \quad A_n = \begin{pmatrix} \frac{2n+1}{n+3} & \frac{(-1)^n}{n} \\ e^{-n} & 5 \end{pmatrix},$$

beaucoup diraient sans hésiter que $(\vec{u}_n)_{n\geqslant 0}$ converge vers $\left(\frac{\pi}{2},1,0\right)$ et $(A_n)_{n\geqslant 0}$ vers $\begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix}$, sans se casser la tête en introduisant une norme, etc. ;

— l'exercice 43 montre qu'il y a plusieurs normes possibles sur un espace vectoriel : mais si les différents résultats de l'analyse dépendent de la norme choisie, cela semble alourdir considérablement les raisonnements : imaginez que chaque étudiant ait sa propre norme préférée, et que chacun d'entre eux trouve des résultats analytiques différents à un exercice donné ! comment éviter ce problème ? faut-il fixer un choix de norme, quitte à perdre les avantages que peut avoir une norme ou l'autre selon les situations ?

Pour la première interrogation, il semble en effet que le calcul de limite « coordonnée par coordonnée » fonctionne :

Exercice 45. Vérifier, avec la norme de votre choix sur $\mathbb{R}^3$ ou $\mathrm{M}_2(\mathbb{R})$, que les suites $(\vec{u}_n)_{n\geqslant 0}$ et $(A_n)_{n\geqslant 0}$ convergent effectivement vers les limites annoncées.

La deuxième question est légitime : on peut avoir des résultats différents selon la norme choisie :

Exercice 46. On pose : $E = \mathrm{C}^0([0,1],\mathbb{R})$. Pour tout $n \in \mathbb{N}$, soit $f_n : [0,1] \to \mathbb{R}$ l'application définie par $x \mapsto x^n$. La suite $(f_n)_{n\geqslant 0}$ est à valeurs dans E. Montrer qu'elle converge vers l'application identiquement nulle si E est muni de la norme $\|\cdot\|_1$, mais que ce n'est pas le cas si E est muni de la norme $\|\cdot\|_\infty$ (ces normes sont définies dans l'exercice 43).

Exercice 47. Pour tout $P \in \mathbb{R}[X]$, on pose : $\|P\| = \sup\limits_{[0,1]} |P|$, et : $\|P\|' = \sup\limits_{[0,2]} |P|$.

1. Montrer que $\|\cdot\|$ et $\|\cdot\|'$ sont des normes sur $\mathbb{R}[X]$.
2. Montrer que la suite $(X^n)_{n\geqslant 0}$ est bornée pour $\|\cdot\|$ et ne l'est pas pour $\|\cdot\|'$ (nous avons défini ce qu'est une suite bornée dans l'exercice 42).

En fait, nous démontrerons un résultat remarquable : lorsque E est un espace vectoriel **réel ou complexe**, de dimension **FINIE**, tous les résultats analytiques sont les mêmes peu importe la norme choisie ! C'est un des théorèmes les plus importants de l'analyse. Il a de nombreuses conséquences, que nous ne citons pas exhaustivement ici :

— dans un exercice, on choisit la norme avec laquelle on est le plus à l'aise ;

— la plupart des résultats analytiques sont vrais si et seulement s'ils sont vrais coordonnée par coordonnée : avoir une limite, être borné, être continu, etc. ; et dans le cas où une suite ou une fonction admet une limite, elle s'obtient en prenant la limite de chaque coordonnée ;

— toutes les applications linéaires ou multilinéaires dont l'espace de départ est de dimension finie sont lipschitziennes donc continues.

En revanche, si l'on est dans un espace de fonctions ou de polynômes, il n'y a pas le choix. Nous devrons préciser avec quelle norme on travaille, et les résultats seront différents selon la norme choisie.

7.4 Résumé : objectif du chapitre

Nous définirons les normes sur des espaces vectoriels et les utiliserons pour généraliser toutes les définitions et propriétés de base de l'analyse, au cas des espaces vectoriels de dimension quelconque.

Pour discuter de continuité, nous aurons besoin de nouvelles fonctions continues de référence, et donnerons des critères pour obtenir la continuité d'applications linéaires ou multilinéaires.

Enfin, si l'on souhaite avoir des *théorèmes*, en particulier généraliser ceux de 1re année, nous devrons définir les analogues en toute dimension des intervalles, des segments, etc. Ce sera l'occasion d'introduire les parties fermées, ouvertes, compactes, connexes par arcs, etc.

En dimension finie, grâce au théorème des bornes atteintes, nous démontrerons l'important théorème selon lequel tous les résultats analytiques (continuité, caractère borné, etc.) ne dépendent pas de la norme choisie. Nous en déduirons que raisonner « coordonnée par coordonnée » suffit souvent à étudier les suites ou fonctions en dimension supérieure (mais finie). C'est très confortable puisque finalement, raisonner sur les coordonnées nous ramène à l'étude de suites ou fonctions réelles. Nous en déduirons aussi que toutes les applications linéaires et multilinéaires définies sur un espace vectoriel de dimension finie sont continues.

Les applications seront disséminées dans tous les chapitres de l'année : étude asymptotique des puissances d'une matrice, prolongement d'identités par densité, résolution d'équations différentielles mises sous forme matricielle, étude de séries à valeurs dans un espace vectoriel de dimension finie (et définition de l'exponentielle matricielle), etc. Le cas de la dérivabilité sera abordé plus en profondeur dans les chapitres *Dérivation et intégration des fonctions vectorielles* et *Calcul différentiel.*

→ page 49

→ page 59

8 Suites et séries de fonctions

Pour résoudre de très nombreux problèmes mathématiques et physiques, on peut être amené à écrire des fonctions f sous la forme : $f(x) = \sum_{n=0}^{+\infty} f_n(x)$, où f_n est, pour tout $n \in \mathbb{N}$, une « fonction simple » (en général : une fonction puissance, une fraction rationnelle, une fonction trigonométrique...). Des exemples de telles décompositions sont :

— les développements en série de Fourier, qui ne sont pas au programme des mathématiques de MP malgré leur importance notable dans toutes les sciences dures (vous en croiserez en sciences physiques) :

$$\forall x \in \mathbb{R}, \quad |\sin(x)| = \frac{2}{\pi} - \frac{4}{\pi}\sum_{n=1}^{+\infty} \frac{\cos(2nx)}{4n^2-1}, \quad \text{etc.}$$

— les développements en série entière, qui font l'objet du chapitre suivant (j'en reparlerai donc plus en détails en temps voulu) :

$$\forall x \in]-1{,}1[,\ \frac{1}{1-x} = \sum_{n=0}^{+\infty} x^n, \quad \forall x \in]-1{,}1],\ \ln(1+x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{n+1}}{n+1}, \quad \text{etc.}$$

— les développements eulériens et tant d'autres, que vous croiserez en devoir ou en exercice :

$$\forall x \in \mathbb{R} \setminus \pi\mathbb{Z}, \quad \frac{1}{\sin(x)} - \frac{1}{x} = \sum_{n=1}^{+\infty} (-1)^{n+1} \frac{2x}{(\pi n)^2 - x^2}, \quad \frac{1}{x} - \frac{1}{\tan(x)} = \sum_{n=1}^{+\infty} \frac{2x}{(\pi n)^2 - x^2}, \quad \text{etc.}$$

(On peut aussi écrire des fonctions comme *produit* de fonctions simples, mais nous ne parlerons pas ici des subtilités que cela implique, et n'en rencontrerons qu'à titre exceptionnel.)

Exercice 48.

1. Montrer l'identité : $\forall x \in]-1{,}1[,\ \ln(1+x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{n+1}}{n+1}$. Deux pistes potentielles :
   — utiliser une formule de Taylor bien choisie ;
   — intégrer la relation : $\forall x \in]-1{,}1[,\ \frac{1-(-x)^{N+1}}{1+x} = \sum_{n=0}^{N} (-x)^n$, et passer à la limite en faisant attention aux subtilités que cela implique.
2. Trouver une identité analogue vérifiée par l'arc tangente.

Plus généralement, un certain nombre de fonctions sont définies à l'aide d'une suite de fonctions supposément plus simples et par passage à la limite : $f(x) = \lim_{n\to+\infty} f_n(x)$. Cette idée est très lointaine (ainsi l'idée d'Archimède d'approcher un cercle par des polygones réguliers, dont le calcul d'aire est approché plus facilement par dichotomie, pour obtenir un bon encadrement de π), et elle apparaît aussi lorsque vous obtenez une intégrale comme limite d'une fonction de Riemann (cela peut être une définition, certes pas la plus souple, de l'intégrale d'une fonction continue). Les sommes ci-dessus rentrent dans ce cadre, puisqu'une somme à support infini est obtenue par passage à la limite à partir des sommes partielles.

On parle alors de *suite de fonctions* lorsqu'on étudie $(f_n)_{n\geqslant 0}$ et de *série de fonctions* lorsqu'on étudie $\sum_{n\geqslant 0} f_n$. Nous allons illustrer dans les trois sections suivantes l'intérêt de les étudier (je vais surtout puiser mes exemples dans les séries de fonctions, autant par sensibilité naturelle que parce que les applications me semblent plus directes et riches, si l'on accepte l'idée de taire provisoirement les subtilités). Nous verrons cependant qu'à chaque fois, il y a un vrai problème à savoir justifier rigoureusement, sans quoi toutes nos illustrations tombent à l'eau.

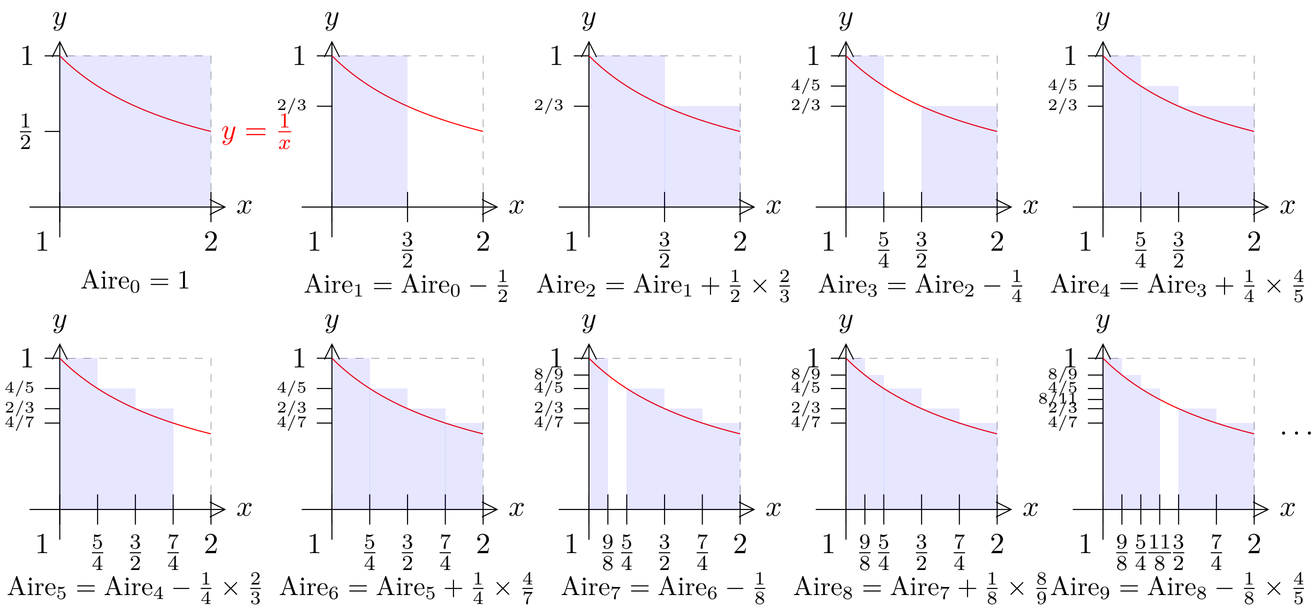
8.1 Pour le calcul de sommes non triviales

Plusieurs sommes difficiles d'accès ont été obtenues en partant de sommes dont la valeur est connue, en utilisant ensuite les opérations standards de l'analyse. Parfois, cela nécessite d'introduire une variable x là où il n'y en a pas. Supposons pour l'exemple que nous voulons obtenir les valeurs des sommes suivantes :

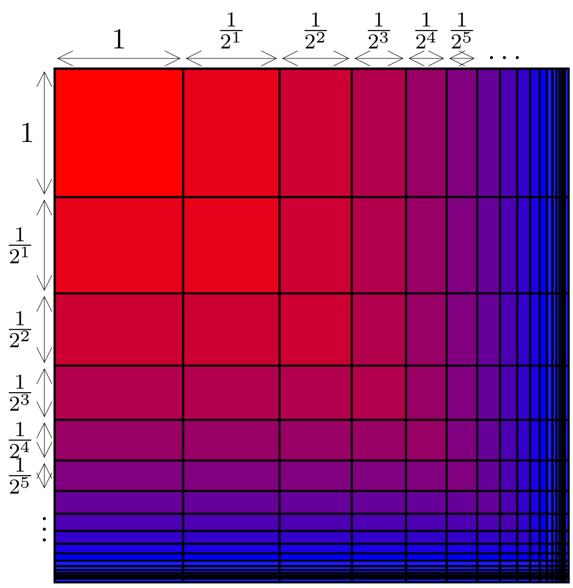
$$\sum_{n=0}^{+\infty} \frac{(-1)^n}{n+1}, \quad \text{et :} \quad \sum_{n=0}^{+\infty} \frac{n+1}{2^n}.$$

Un peu de flair géométrique permet de conjecturer (voire démontrer, si l'on s'y prend bien) la valeur de ces deux sommes :

Exercice 49. Utiliser la figure ci-dessous pour conjecturer la valeur de la somme $\displaystyle\sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{n}$.



Exercice 50. À partir de la figure ci-dessous, conjecturer la valeur de : $\displaystyle\sum_{n=0}^{+\infty} \frac{n+1}{2^n}$.



Pour obtenir ces deux sommes, nous allons compliquer pour simplifier, c'est-à-dire : nous allons introduire une variable x et plutôt calculer :

$$\sum_{n=0}^{+\infty}\frac{(-1)^n x^{n+1}}{n+1},\quad \text{et :}\quad \sum_{n=0}^{+\infty}\frac{(n+1)x^n}{2^n}.$$

(D'autres choix pertinents seraient possibles, par exemple : $\sum\limits_{n=0}^{+\infty}(n+1)x^n$). L'intérêt de compliquer *en apparence* la somme à calculer est qu'à présent, les opérations classiques de l'analyse (dériver, intégrer) nous permettent d'exprimer ces sommes inconnues à l'aide de sommes connues (essentiellement, à votre stade : une somme géométrique) :

$$\sum_{n=0}^{+\infty}\frac{(-1)^n x^{n+1}}{n+1}=\sum_{n=0}^{+\infty}(-1)^n\int_0^x t^n\mathrm{d}t \overset{(*)}{=} \int_0^x\sum_{n=0}^{+\infty}(-1)^n t^n\mathrm{d}t=\int_0^x\frac{1}{1+t}\mathrm{d}t=\ln(1+x),$$

et :

$$\sum_{n=0}^{+\infty}\frac{(n+1)x^n}{2^n}=\sum_{n=0}^{+\infty}\frac{1}{2^n}\frac{\mathrm{d}}{\mathrm{d}x}x^{n+1} \overset{(\dagger)}{=} 2\frac{\mathrm{d}}{\mathrm{d}x}\sum_{n=0}^{+\infty}\left(\frac{x}{2}\right)^{n+1}=2\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{1-\frac{x}{2}}-1\right)=\frac{1}{(1-\frac{x}{2})^2}.$$

En posant $x=1$ dans chaque égalité, sous réserve que le raisonnement soit valable, on obtient les valeurs attendues avec beaucoup de facilité :

$$\sum_{n=0}^{+\infty}\frac{(-1)^n}{n+1}=\ln(2),\quad \text{et :}\quad \sum_{n=0}^{+\infty}\frac{n+1}{2^n}=4.$$

Cette idée peut être exploitée à l'infini : pourquoi se contenter d'intégrer ou dériver une seule fois une somme connue ? Il semble possible de le faire autant de fois qu'on le souhaite, et d'ainsi obtenir autant de nouvelles valeurs de sommes : à la fin de la MPSI, vous ne connaissiez que les sommes géométriques, télescopiques et exponentielles, mais ce procédé permet de gonfler indéfiniment le répertoire !

Exercice 51. Donner les valeurs des sommes suivantes en vous inspirant du raisonnement ci-dessus :

$$\text{(a)}\ \sum_{n=0}^{+\infty}\frac{n^2}{3^n},\quad \text{(b)}\ \sum_{n=0}^{+\infty}\binom{n+3}{n}\frac{1}{2^n},\quad \text{(c)}\ \sum_{n=0}^{+\infty}\binom{n+k}{n}\frac{1}{2^n}\ (k\in\mathbb{N}),\quad \text{(d)}\ \sum_{n=0}^{+\infty}\frac{1}{2n+1}\left(\frac{2}{3}\right)^{2n}.$$

On traitera cet exercice *sans se soucier de justifier la validité des opérations classiques de l'analyse.*

Exercice 52. On traitera cet exercice *sans se soucier de justifier la validité des opérations classiques de l'analyse.* Vous utiliserez à chaque fois une ou plusieurs des identités apparues dans le reste de cette section (vous les admettrez donc, en attendant de savoir les démontrer l'an prochain).

1. Calculer $\sum\limits_{n=1}^{+\infty}\frac{1}{4n^2-1}$ et $\sum\limits_{n=1}^{+\infty}\frac{(-1)^n}{4n^2-1}$. On vérifiera que la valeur de $\sum\limits_{n=1}^{+\infty}\frac{1}{4n^2-1}$ trouvée est correcte en calculant cette somme par une autre méthode (*passer par une décomposition en éléments simples*).
2. Calculer $\sum\limits_{n=0}^{+\infty}\frac{(-1)^{n+1}}{n+1}$.
3. Calculer $\sum\limits_{n=1}^{+\infty}\frac{(-1)^{n+1}}{n^2}$ et $\sum\limits_{n=1}^{+\infty}\frac{(-1)^{n+1}}{n^4}$. En déduire la valeur de $\sum\limits_{n=1}^{+\infty}\frac{1}{n^2}$ et $\sum\limits_{n=1}^{+\infty}\frac{1}{n^4}$ (vous pouvez essayer de changer l'exposant de n en n'importe quel exposant pair strictement positif).
4. Obtenir le « produit de Weierstraß » du sinus, dû à Euler (notez l'anachronisme) :

$$\forall x\in\mathbb{R},\quad \sin(x)=x\prod_{n=1}^{+\infty}\left(1-\frac{x^2}{(\pi n)^2}\right).$$

L'utiliser pour en déduire à nouveau la valeur de $\sum_{n=1}^{+\infty} \frac{1}{n^2}$ (c'est historiquement la première résolution du problème de Mengoli, par Euler).

5. Déduire de la question précédente une identité analogue vérifiée par le sinus hyperbolique. Puis, en combinant les écritures de $\sin(x)$ et $\mathrm{sh}(x)$ sous forme de produit, obtenir la valeur de $\sum_{n=1}^{+\infty} \frac{1}{n^4}$ en suivant une inspiration analogue à celle de la question précédente. Si vous y parvenez : se demander ensuite pourquoi cette méthode ne permet pas d'obtenir la valeur de $\sum_{n=1}^{+\infty} \frac{1}{n^3}$.

Il y eut cependant un gros manque de prudence dans ces raisonnements, outre l'existence des sommes en présence : est-ce que les *interversions de symboles* avec les égalités $(*)$ et $(\dagger)$ sont dûment justifiées ? Si les sommes n'avaient qu'un nombre fini de termes, bien sûr que oui. Mais les choses peuvent se gâter dans le cas contraire et nous l'illustrerons dans la section 8.4. → page 36

8.2 Pour résoudre un problème en le discrétisant

Plusieurs résolutions d'équations fonctionnelles (différentielles, intégrales, et d'autres encore) peuvent se faire par « réitération » de l'équation vérifiée par une fonction solution. Ce procédé fait naturellement apparaître une suite ou série de fonctions.

Prenons un exemple que vous connaissez bien. Supposons qu'on veuille démontrer l'existence d'une solution au problème de Cauchy suivant :

$$\begin{cases} y' &= y, \\ y(0) &= 1, \end{cases}$$

en n'ayant jamais entendu parler de la fonction exponentielle (et si possible expliciter une telle solution). On peut la déterminer pas à pas en remarquant d'abord que si l'on intègre la première égalité de 0 à un réel x, elle équivaut à :

$$\begin{cases} \forall x \in \mathbb{R}, \quad y(x) - y(0) &= \displaystyle\int_0^x y(t)\mathrm{d}t, \\ y(0) &= 1, \end{cases}$$

soit donc :

$$\forall x \in \mathbb{R}, \quad y(x) = 1 + \int_0^x y(t)\mathrm{d}t. \tag{$*$}$$

Pour abréger, posons Y_0 l'application $x \mapsto \int_0^x y(t)\mathrm{d}t$, de sorte que l'on ait : $y = 1 + Y_0$. De même, dans ce qui suit, pour tout entier naturel n, nous définirons Y_{n+1} par récurrence comme l'unique primitive de Y_n s'annulant en 0. En injectant cette expression de y dans « elle-même », on obtient dans un premier temps :

$$\forall x \in \mathbb{R}, \quad y(x) = 1 + \int_0^x (1 + Y_0(t))\,\mathrm{d}t = 1 + x + \int_0^x Y_0(t)\mathrm{d}t = 1 + x + Y_1(x).$$

On peut réinjecter cette expression dans $(*)$ pour obtenir :

$$\forall x \in \mathbb{R}, \quad y(x) = 1 + \int_0^x (1 + t + Y_1(t))\,\mathrm{d}t = 1 + x + \frac{x^2}{2} + \int_0^x Y_1(t)\mathrm{d}t = 1 + x + \frac{x^2}{2} + Y_2(x).$$

Vous l'avez bien compris : je compte recommencer, autant de fois que nécessaire. On obtient successivement, comme vous le vérifierez aisément :

$$\begin{aligned}\forall N\in\mathbb{N},\ \forall x\in\mathbb{R},\quad y(x) &= 1+x+\frac{x^2}{2}+\frac{x^3}{2\cdot 3}+Y_3(x)\\ &= 1+x+\frac{x^2}{2}+\frac{x^3}{2\cdot 3}+\frac{x^4}{2\cdot 3\cdot 4}+Y_4(x)\\ &= 1+x+\frac{x^2}{2}+\frac{x^3}{2\cdot 3}+\frac{x^4}{2\cdot 3\cdot 4}+\frac{x^5}{2\cdot 3\cdot 4\cdot 5}+Y_5(x)\\ &\ \vdots\\ &= \sum_{n=0}^{N}\frac{x^n}{n!}+Y_N(x).\end{aligned}$$

Il semble donc que, si l'on itérait ce raisonnement indéfiniment, on trouverait comme solution candidate : $\forall x\in\mathbb{R},\ y(x)=\sum_{n=0}^{+\infty}\frac{x^n}{n!}$. Que reconnaît-on ? !

Bien entendu l'approche ci-dessus souffre d'un grand manque de rigueur (les Y_n s'évaporent par passage à la limite ? au nom de quoi ?), mais elle permet tout de même d'illustrer pourquoi, lors de la résolution d'équations fonctionnelles, il apparaît naturellement des sommes de fonctions dont le support est infini.

Exercice 53. S'inspirer de ce raisonnement très informel pour conjecturer des solutions éventuelles des équations fonctionnelles suivantes :

$$\text{(a) } \forall x\in\mathbb{R}_+,\ y(x)=1+\int_0^x y\left(\frac{t}{2}\right)\mathrm{d}t,\qquad \text{(b) } \begin{cases}\forall x\in\mathbb{R}_+^*,\ y(x+1)+y(x) &= \dfrac{1}{x^3},\\ \lim\limits_{x\to+\infty} y(x) &= 0.\end{cases}$$

À ce stade-là, pour rendre rigoureuse la résolution ci-dessus, les possibilités sont nombreuses. L'une d'elles est de réciproquement vérifier que la fonction définie par $x\mapsto\sum_{n=0}^{+\infty}\frac{x^n}{n!}$ est dérivable et solution du problème de Cauchy :

$$\frac{\mathrm{d}}{\mathrm{d}x}\sum_{n=0}^{+\infty}\frac{x^n}{n!}\overset{(\dagger)}{=}\sum_{n=0}^{+\infty}\frac{1}{n!}\frac{\mathrm{d}x^n}{\mathrm{d}x}=\sum_{n=1}^{+\infty}\frac{nx^{n-1}}{n!}=\sum_{n=1}^{+\infty}\frac{x^{n-1}}{(n-1)!}=\sum_{n=0}^{+\infty}\frac{x^n}{n!}.$$

Cela marche ! Seulement la justesse de l'égalité (†) ne va pas du tout de soi, bien qu'on sache dériver une somme de fonctions dérivables aisément *tant qu'elle est à support fini*, et nous en parlerons amplement dans la section 8.4. Une autre est de plutôt étudier la suite définie par récurrence par : → page 36

$$\forall x\in\mathbb{R},\ y_0(x)=1,\qquad \text{et :}\qquad \forall n\in\mathbb{N},\ \forall x\in\mathbb{R},\ y_{n+1}(x)=1+\int_0^x y_n(t)\mathrm{d}t.$$

Si vous y réfléchissez bien, c'est presque l'approche que j'ai suivie plus haut, mais bien mieux posée. Ainsi on est amené à étudier la suite $(y_n)_{n\geqslant 0}$ et son comportement quand $n\to+\infty$. Si elle converge vers une fonction y (en un sens à préciser), on a envie de penser que lorsqu'on prend la limite quand $n\to+\infty$ dans l'égalité ci-dessus, on en déduit que y vérifie $(*)$ comme attendu. Il se pose alors la question : est-ce bien convergent ? Si oui, peut-on passer à la limite dans l'intégrale ci-dessus ? (Question déjà abordée dans le chapitre *Intégration.*) Nous retrouvons des questionnements analogues à ceux suivant la première approche, et commençons à entrevoir les problématiques du chapitre, que nous synthétiserons plus bas.

8.3 Pour se ramener à des fonctions plus simples

Il est bien des situations où nous aimons décomposer un objet mathématique comme somme (dans un contexte de linéarité, sinon ce peut être autre chose) d'objets plus « simples », afin de faciliter l'étude du problème. C'est ainsi qu'en algèbre linéaire, on démontre un certain nombre de choses sur une base ou une partie génératrice, puis on étend le résultat à tout vecteur par linéarité.

On peut faire de même en analyse : écrire une fonction compliquée comme somme de fonctions plus simples (monômes, fractions rationnelles...) pour faciliter les calculs qui l'impliquent.

Exemple concret : si l'on souhaite calculer l'intégrale $\displaystyle\int_0^1 \frac{\ln(t)}{1-t}\mathrm{d}t$ (on admet que cette intégrale a un sens, bien que le problème de définition en 0 et 1 doive vous interpeller : ce sera expliqué dans le cadre du chapitre d'intégration), on est bien embêté en se contentant de recourir aux techniques classiques (détermination d'une primitive, intégration par parties, changement de variable...). Ce qui rend cette intégrale pénible à évaluer est aussi bien la présence du logarithme (parce que la présence d'une fraction rationnelle) que du quotient (parce qu'un logarithme « seul », ou multiplié par un facteur polynomial à la rigueur, on saurait l'intégrer). Bref, nous avons là une fonction « compliquée ». Cependant, grâce à la somme géométrique :

$$\forall t \in]-1,1[, \quad \frac{1}{1-t} = \sum_{n=0}^{+\infty} t^n,$$

nous pouvons éliminer le quotient ci-dessus et faire apparaître une somme de fonctions « simples » à intégrer :

$$\int_0^1 \frac{\ln(t)}{1-t}\mathrm{d}t = \int_0^1 \ln(t)\sum_{n=0}^{+\infty} t^n \mathrm{d}t \overset{(*)}{=} \sum_{n=0}^{+\infty}\int_0^1 \ln(t)t^n\mathrm{d}t \overset{(\dagger)}{=} \sum_{n=0}^{+\infty}\frac{1}{(n+1)^2},$$

et on sait démontrer que cette dernière somme est égale à $\dfrac{\pi^2}{6}$ (j'affirme que même si l'on ne connaissait pas la valeur exacte de cette somme, ce serait un résultat malgré tout satisfaisant : pourquoi ?). On comprend là l'intérêt de s'être ramené à des fonctions plus simples, au sens où on sait les intégrer.

Exercice 54.

1. Justifier l'égalité (†).
   Il y a un problème de définition de l'intégrande en 0 *: on admettra que dans l'usage de vos techniques de calcul intégral, il suffit ici de remplacer les évaluations par des passages à la limite pour y remédier, ce qu'on justifiera proprement dans le chapitre d'intégration.*
2. Obtenir le même résultat final, mais en utilisant plutôt l'identité démontrée dans l'exercice 48 (page 30) et qu'on revoit sous une forme légèrement différente dans l'exercice 55.

En fait, en faisant cela nous suivons de près la façon de faire de nos anciens, pour qui il était très concret et naturel d'écrire ainsi les fonctions usuelles (se demander pourquoi c'était « concret »). D'autres exemples :

Exercice 55. On traitera cet exercice *sans se soucier de justifier la validité des opérations classiques de l'analyse.*

1. Utiliser l'identité : $\forall x \in]0,2[$, $\ln(x) = \displaystyle\sum_{n=0}^{+\infty}(-1)^n\frac{(x-1)^{n+1}}{n+1}$, pour démontrer « naturellement » qu'une primitive du logarithme sur $]0,2[$ est $x \mapsto x\ln(x) - x$, et que sa dérivée est $x \mapsto \dfrac{1}{x}$.
2. Utiliser l'identité : $\forall x \in]-1,1[$, $\arctan(x) = \displaystyle\sum_{n=0}^{+\infty}(-1)^n\frac{x^{2n+1}}{2n+1}$, pour démontrer « naturellement » qu'une primitive de l'arc tangente sur $]-1,1[$ est $x \mapsto x\arctan(x) - \dfrac{1}{2}\ln(1+x^2)$.
3. Soit $k \in \mathbb{N}$. S'inspirer des deux questions précédentes pour déterminer une application F de classe C^k sur $]0,2[$, s'exprimant uniquement à l'aide du logarithme et de fonctions puissances, sans le signe intégrale, et telle que : $F^{(k)} = \ln$.

8.4 Un (très) gros problème se pose avec les sommes à support infini

Dans nos exemples et exercices des trois sections précédentes, nous avons à plusieurs reprises utilisé des identités du type :

$$\sum_{n=0}^{+\infty} \frac{\mathrm{d}f_n}{\mathrm{d}x}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\sum_{n=0}^{+\infty} f_n(x), \quad \sum_{n=0}^{+\infty}\int_I f_n(x)\mathrm{d}x = \int_I \sum_{n=0}^{+\infty} f_n(x)\mathrm{d}x, \quad \sum_{n=0}^{+\infty}\lim_{x\to a} f_n(x) = \lim_{x\to a}\sum_{n=0}^{+\infty} f_n(x).$$

Puisqu'il s'agit à chaque fois d'une permutation de deux symboles ($\sum_{n=0}^{+\infty}$ et $\frac{\mathrm{d}}{\mathrm{d}x}$, $\int_I$ ou $\lim_{x\to a}$), la justification rigoureuse de ces égalités est ce qu'on appelle un *problème d'interversion.* On parle de PROBLÈME parce que oui, il y en a un qui se pose ! Les contre-exemples sont même très faciles à produire, et tout peut arriver. Par exemple la fonction :

$$x \mapsto \sum_{n=1}^{+\infty} \frac{\sin(n^2 x)}{n^2},$$

bien que définie sur $\mathbb{R}$ et de terme général dérivable pour tout entier n, n'a certainement pas pour dérivée $x \mapsto \sum_{n=1}^{+\infty} \cos(n^2 x)$ puisque c'est la somme d'une série qui diverge grossièrement en presque tout réel.

Exercice 56. Démontrer les deux affirmations ci-dessus (domaine de définition de la fonction, divergence grossière de la série).

Ci-dessus, nous n'avons pas contredit pour autant que la fonction est dérivable sur $\mathbb{R}$. On peut trouver des exemples de sommes de fonctions dérivables qui ne sont dérivables en aucun point. Un exemple historiquement dû à Weierstraß est :

$$x \mapsto \sum_{n=0}^{+\infty} \frac{\cos(2^n \pi x)}{2^n}.$$

Démontrer que la fonction est continue partout est raisonnablement facile avec les outils de MP, mais la dérivabilité nulle part est un problème ardu (posé au concours des ENS en 2007).

Les autres interversions posent également problème. Nous en avons déjà parlé dans la section 2 ← page 3 (*Intégration*) pour la seconde interversion (on va alors recycler un contre-exemple qui y était fourni), et vous vous en convaincrez aisément pour la troisième :

Exercice 57.

1. Utiliser la série $\sum_{n\geqslant 0}\Big((n+1)x^{n+1} - nx^n\Big)$ pour montrer que l'interversion $\sum_{n=0}^{+\infty}\int_I \bigstar = \int_I \sum_{n=0}^{+\infty} \bigstar$ est fausse en générale.
2. Utiliser la série géométrique $\sum_{n\geqslant 0} x^n$ pour montrer que l'interversion $\sum_{n=0}^{+\infty}\lim_{x\to a} \bigstar = \lim_{x\to a}\sum_{n=0}^{+\infty} \bigstar$ est fausse en générale.

La question est de savoir : sous quelles hypothèses est-il possible d'intervertir les symboles $\sum_{n=0}^{+\infty}$ et $\frac{\mathrm{d}}{\mathrm{d}x}$, $\int_I$ ou $\lim_{x\to a}$? La question se pose également avec les suites de fonctions, auquel cas ce sont les interversions suivantes pour lesquelles on a besoin de théorèmes :

$$\lim_{n\to+\infty} \frac{\mathrm{d}f_n}{\mathrm{d}x}(x) = \frac{\mathrm{d}}{\mathrm{d}x}\lim_{n\to+\infty} f_n(x), \quad \lim_{n\to+\infty}\lim_{x\to a} f_n(x) = \lim_{x\to a}\lim_{n\to+\infty} f_n(x).$$

J'ai mis de côté le cas de l'intégration, puisque la question a déjà été abordée dans le chapitre associé.

Une réponse à cette question s'appelle un *théorème d'interversion*. On verra que le moyen d'éviter les anomalies à l'origine des problèmes d'interversion reste le même que dans le cas de l'intégration : avoir des majorations uniformes. Dans le cas des séries de fonctions : pour s'assurer qu'une interversion est possible, il s'agira de faire des majorations uniformes du *reste*, puisque c'est la fonction qui mesure l'écart entre les sommes $\sum_{n=0}^{N} f_n$ (pour lesquelles toutes les opérations de l'analyse se passent bien) et les sommes $\sum_{n=0}^{+\infty} f_n$ (pour lesquelles on *aimerait* que tout se passe bien). C'est en s'assurant que le reste $\sum_{n=N+1}^{+\infty} f_n$ (ou la dérivée du reste, ou... cela dépend du contexte) est « petit » qu'on évite les aberrations en passant des sommes finies aux sommes infinies.

Nous serons plus précis en temps voulu.

8.5 Résumé : objectif du chapitre

Nous donnerons des théorèmes d'interversion sur les suites de fonctions (théorème de la double limite, théorème de passage à la limite sous la dérivée, extension aux fonctions de classe C^k : le cas des passages à la limite sous le signe intégrale est déjà traité), en insistant sur l'importance des hypothèses de convergence uniforme, et sur le rôle très singulier que jouent parfois les segments dans ces hypothèses (pour éviter les aberrations pouvant se produire aux extrémités de l'intervalle d'étude).

Nous introduirons ensuite la définition d'une *série* de fonctions, et étendrons la définition de convergence uniforme ou simple à ces séries (en disant simplement que cela équivaut au même mode de convergence mais pour les suites de leurs sommes partielles). Il sera alors possible d'obtenir des théorèmes d'interversion pour les séries de fonctions en appliquant les théorèmes sur les suites de fonctions à leurs sommes partielles.

Forts de tous ces résultats d'interversion, nous pourrons en application répondre à des questions telles que celles formulées dans les sections précédentes, à savoir : nous résoudrons des problèmes par approximation par des fonctions plus simples, ou par construction de sommes de séries de fonctions qui répondent naturellement aux problèmes posés.

Parmi les résultats généraux d'approximation, nous verrons que toute fonction continue sur un segment et à valeurs complexes est limite uniforme de fonctions polynomiales (théorème de Weierstraß), ce qui permet en substance le raisonnement suivant : si l'on sait démontrer un résultat pour toute application polynomiale, alors sous de bonnes hypothèses un passage à la limite permet d'obtenir le résultat pour toute fonction continue.

9 Séries entières

9.1 Définitions et propriétés spécifiques de ces séries

On appelle fonction *développable en série entière en* 0 une fonction qui peut s'écrire sous la forme $x \mapsto \sum_{n=0}^{+\infty} a_n x^n$ dans un voisinage de 0 *, où $(a_n)_{n\geqslant 0}$ est une suite de nombres complexes (la variable x est réelle ou complexe). Vous avez déjà croisé de telles fonctions. Vous savez en effet que l'on a :

$$\forall z \in \mathbb{C},\ \exp(z) = \sum_{n=0}^{+\infty} \frac{z^n}{n!}, \quad \forall x \in]-1,1[,\ \frac{1}{1-z} = \sum_{n=0}^{+\infty} z^n.$$

Partant de la fonction exponentielle, vous en déduisez aisément d'autres exemples parmi les fonctions usuelles :

*. Si on veut définir une fonction développable en série entière en un point ω, on considère à la place l'application $x \mapsto \sum_{n=0}^{+\infty} a_n (x-\omega)^n$. Mais quitte à remplacer une telle fonction f par $x \mapsto f(x+\omega)$, on revient à un développement en 0. C'est pourquoi le programme nous cantonne à ce cas-là sans que ce soit contraignant.

Exercice 58. Montrer que les fonctions cosinus et sinus, aussi bien circulaires qu'hyperboliques, sont développables en série entière en 0. Voir aussi l'exercice 80 (page 52) si l'on ne voit pas quel développement on est supposé obtenir.

Il s'agit d'un cas particulier de série de fonctions (voir chapitre précédent). Les motivations données pour les séries de fonctions restent donc valables ici : les séries entières servent (entre autres) au calcul de sommes non triviales, à la résolution de problèmes par discrétisation, et à se ramener à des fonctions plus simples (des fonctions puissances ici). C'est illustré abondamment dans la section 8.

Mais d'autres raisons motivent leur étude spécifique : tout d'abord, tout bêtement, il y a « beaucoup » de fonctions développables en séries entières parmi celles que nous connaissons (l'exponentielle, le logarithme, les fonctions trigonométriques directes et réciproques, circulaires et hyperboliques, etc.), donc étudier les séries entières permet d'en savoir plus sur les fonctions que nous manipulons quotidiennement. Ensuite : les séries entières s'avèrent vérifier d'autres propriétés mirifiques :

— leur domaine de définition a une symétrie remarquable : une série entière converge absolument en tout point d'un certain disque ouvert (ou un intervalle ouvert si l'on se restreint à une variable réelle) centré en 0, et cela implique notamment un comportement « tout ou rien » : s'il y a convergence en un réel positif r, il y a aussi convergence en tout z tel que $|z| < r$ (avec une condition inverse en cas de divergence) ;

— nous disions dans la section précédente que la linéarité des opérations classiques de l'analyse (dérivation, intégration, passage à la limite...) ne se généralise pas aux sommes à support infini ; mais avec les séries entières, c'est toujours vrai ! (... en gros).

Attardons-nous sur deux autres propriétés, qui permettent de traiter de bâtir un pont entre le monde des **fonctions** et le monde des **suites** :

— si $S_a : x \mapsto \sum_{n=0}^{+\infty} a_n x^n$ et $S_b : x \mapsto \sum_{n=0}^{+\infty} b_n x^n$ sont deux fonctions développables en série entière en 0 (on les appelle *séries génératrices* de $(a_n)_{n\geqslant 0}$ et $(b_n)_{n\geqslant 0}$), alors :

$$S_a = S_b \overset{(\spadesuit)}{\iff} \exists r > 0,\ \forall x \in]-r, r[,\ S_a(x) = S_b(x) \overset{(\clubsuit)}{\iff} \forall n \in \mathbb{N},\ a_n = b_n\,;$$

— sous de bonnes hypothèses, on a :

$$\left(\sum_{n=0}^{+\infty} a_n x^n\right) \times \left(\sum_{n=0}^{+\infty} b_n x^n\right) = \sum_{n=0}^{+\infty}\left(\sum_{k=0}^{n} a_k b_{n-k}\right) x^n. \qquad \text{(produit de Cauchy)}$$

L'équivalence ($\spadesuit$) est remarquable : elle nous enseigne que, pour que deux fonctions développables en série entière soient égales PARTOUT, il suffit qu'elles soient égales AU VOISINAGE DE ZÉRO. C'est un moyen puissant de démontrer des identités sur $\mathbb{C}$ en les montrant sur $\mathbb{R}$ (ou sur un intervalle de $\mathbb{R}$ suffisamment petit pour nos besoins : $]-1,1[$ par exemple, si l'on a besoin d'avoir une suite ou série géométrique convergente). Ainsi, *sans le moindre effort*, elle nous assure que des identités telles que $\cos^2 + \sin^2 = 1$, $e^{x+y} = e^x e^y$, etc., sont valables sur $\mathbb{C}$ dès lors qu'on les a démontrées sur $\mathbb{R}$. On parle de *prolongement analytique* ou de *principe des zéros isolés* (selon notre façon d'utiliser cette équivalence).

Exercice 59. Déduire de ce paragraphe que les fonctions $x \mapsto \mathrm{Re}(x)$, $x \mapsto \mathrm{Im}(x)$ et $x \mapsto |x|^2$ ne sont pas développables en série entière en 0.

Toutefois, c'est bien sûr l'équivalence ($\clubsuit$) qui permet de faire un pont entre le monde des fonctions et celui des suites : deux fonctions développables en série entière sont égales si et seulement si leurs termes généraux sont égaux. Cela donne une nouvelle piste pour démontrer une égalité du type : $\forall n \in \mathbb{N}$, $a_n = b_n$, où a_n est une quantité qu'on cherche à simplifier ou expliciter et b_n une quantité tout à fait explicite, dépendant de suites usuelles. On introduit à la place l'application $S_a : x \mapsto \sum_{n=0}^{+\infty} a_n x^n$ et on montre qu'elle est égale à $S_b : x \mapsto \sum_{n=0}^{+\infty} b_n x^n$ (en fait, dans la plupart des exercices, on ne sait pas sur quelle suite $(b_n)_{n\geqslant 0}$ ou quelle fonction S_b on est supposé tomber ; pas grave, les différentes formules de calcul du chapitre permettront de la trouver). Si on y parvient, c'est gagné grâce à (♣).

On peut se demander en quoi cette idée saugrenue serait pertinente : pourquoi une somme de série serait plus facile à étudier que son terme général ? Pour y répondre, regardons ce que deviennent certaines opérations élémentaires sur les suites quand on considère leurs séries génératrices :

— **décaler l'indice d'une suite** $(a_n)_{n\geqslant 0}$ revient à **changer** S_a **en** $x \mapsto \dfrac{S_a(x)-a_0}{x}$ (ce que j'entends par là : si $c_n = a_{n+1}$ pour tout $n \in \mathbb{N}$, alors la série génératrice de $(c_n)_{n\geqslant 0}$ vérifie : $S_c(x) = \displaystyle\sum_{n=0}^{+\infty} a_{n+1}x^n = \sum_{n=1}^{+\infty} a_n x^{n-1} = \frac{S_a(x)-a_0}{x}$) ;

— **multiplier par** n le terme général d'une suite $(a_n)_{n\geqslant 0}$ revient essentiellement à **dériver** S_a (ce que j'entends par là : si $c_n = na_n$ pour tout $n \in \mathbb{N}$, alors : $S_c(x) = x\displaystyle\sum_{n=1}^{+\infty} a_n n x^{n-1} = xS_a'(x)$;

— **sommer les** $n+1$ **premiers termes** d'une suite $(a_n)_{n\geqslant 0}$ revient à **multiplier** S_a **par** $\dfrac{1}{1-x}$, grâce à la formule du produit de Cauchy : si $c_n = \displaystyle\sum_{k=0}^{n} a_k$ pour tout $n \in \mathbb{N}$, alors : $S_c(x) = \displaystyle\sum_{n=0}^{+\infty}\left(\sum_{k=0}^{n} 1\cdot a_k\right)x^n = \left(\sum_{n=0}^{+\infty} x^n\right)\left(\sum_{n=0}^{+\infty} a_n x^n\right) = \frac{S_a(x)}{1-x}$, et on peut trouver une traduction analogue pour toute relation du type $c_n = \displaystyle\sum_{k=0}^{n} a_k b_{n-k}$;

— et ainsi de suite : nous n'avons pas épuisé les traductions possibles !

En conséquence : une relation vérifiée par la suite $(a_n)_{n\geqslant 0}$ en implique une autre vérifiée par sa série génératrice S_a, en vertu des identités ci-dessus. Si l'on tombe sur une équation fonctionnelle ou différentielle plus simple que la relation initiale (espoir raisonnable comme on va le voir), alors on en déduit une expression explicite de S_a, puis une relation du type : $S_a = S_b$, où S_b est développable en série entière, puis une expression explicite de a_n pour tout $n \in \mathbb{N}$ grâce à (♣).

Imaginons par exemple qu'on veuille expliciter une suite vérifiant :

$$\forall n \in \mathbb{N}, \quad u_{n+2} = u_{n+1} + u_n,$$

avec pour premiers termes $u_0 = 0$ et $u_1 = 1$ (c'est la suite de Fibonacci).

On admet que la série $\displaystyle\sum_{n\geqslant 0} u_n x^n$ converge pour tout $x \in \mathbb{R}$ tel que : $|x| \leqslant \dfrac{1}{2}$ (on peut même faire un peu mieux, mais ce sera suffisant). Notons S sa somme. Si l'on multiplie par x^{n+2} l'égalité $u_{n+2} = u_{n+1} + u_n$, et qu'on somme de $n=0$ à $+\infty$, on obtient pour tout $x \in \left]-\frac{1}{2}, \frac{1}{2}\right[$:

$$\sum_{n=0}^{+\infty} u_{n+2}x^{n+2} = x\sum_{n=0}^{+\infty} u_{n+1}x^{n+1} + x^2 S(x), \qquad \text{soit donc :} \qquad S(x) - u_0 - u_1 x = x(S(x) - u_0) + x^2 S(x).$$

Comme $u_0 = 0$ et $u_1 = 1$ par hypothèse, on en déduit facilement après calculs :

$$\forall x \in \left]-\frac{1}{2}, \frac{1}{2}\right[, \quad S(x) = -\frac{x}{x^2+x-1}.$$

Il reste à écrire le membre de droite comme une fonction développable en série entière, ce qui est une affaire de routine après l'avoir décomposé en éléments simples. On obtient ici, après avoir noté $x_1 = \dfrac{-1+\sqrt{5}}{2}$ et $x_2 = \dfrac{-1-\sqrt{5}}{2}$ les deux racines de X^2+X-1 :

$$\forall x \in \left]-\frac{1}{2}, \frac{1}{2}\right[, \quad S(x) = -\frac{x}{\sqrt{5}}\left(-\frac{1}{x_1}\cdot\frac{1}{1-\frac{x}{x_1}} + \frac{1}{x_2}\cdot\frac{1}{1-\frac{x}{x_2}}\right) = \sum_{n=0}^{+\infty} \frac{1}{\sqrt{5}}\left(\frac{1}{x_1^n} - \frac{1}{x_2^n}\right)x^n.$$

D'après l'équivalence (♣), on a donc :

$$\forall n \in \mathbb{N}, \quad u_n = \frac{1}{\sqrt{5}}\left(\frac{1}{x_1^n} - \frac{1}{x_2^n}\right) = \frac{x_2^n - x_1^n}{\sqrt{5}(x_1x_2)^n} = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right].$$

C'est aussi l'expression que vous auriez trouvée en explicitant la suite $(u_n)_{n\geqslant 0}$ *via* les méthodes classiques de 1re année.

Exercice 60. Compléter l'exemple ci-dessus :

1. Montrer : $\forall n \in \mathbb{N}$, $0 \leqslant u_n \leqslant 2^n$ (ne pas utiliser l'expression explicite de $(u_n)_{n\geqslant 0}$, ou le raisonnement est circulaire !). En déduire que $\sum\limits_{n\geqslant 0} u_n x^n$ converge pour tout $x \in \left]-\frac{1}{2}, \frac{1}{2}\right[$.
2. Montrer : $\dfrac{1}{X^2+X-1} = \dfrac{1}{x_1-x_2}\left(\dfrac{1}{X-x_1} - \dfrac{1}{X-x_2}\right)$. En déduire le développement en série entière de $x \mapsto -\dfrac{x}{x^2+x-1}$ proposé.

Le succès de notre raisonnement ci-dessus est dû au fait qu'un décalage d'indice pour des suites se traduise par une simple multiplication ou division par une puissance de x pour les séries génératrices associées : on se retrouve avec une équation très banale à résoudre. On a ainsi illustré en quoi le pont entre le monde des suites et le monde des fonctions permet d'expliciter les unes en fonction des autres.

Le produit de Cauchy est mis en valeur plus haut, parce qu'il est particulièrement efficace dans le cas où l'on étudie des suites définies par des sommes, ou qui vérifient des relations impliquant des sommes. On donne un exemple en exercice ci-dessous, que vous connaissez bien. Notez qu'un intérêt de l'approche de cet exercice est qu'elle ne nécessite aucunement d'avoir une idée du résultat à obtenir.

Exercice 61. On pose : $\forall n \in \mathbb{N}$, $u_n = \sum\limits_{k=0}^{n} k$.

1. Montrer que pour tout $x \in]-1,1[$, la série $\sum\limits_{n\geqslant 0} nx^n$ converge vers $\dfrac{x}{(1-x)^2}$. Trois pistes éventuelles pour le calcul de la somme :
   — faire un changement d'indice simple pour en déduire une relation vérifiée par la somme de cette série ;
   — utiliser le théorème de sommation par paquets ou reconnaître un produit de Cauchy de deux sommes qu'on sait calculer ;
   — faire une transformation d'Abel, si vous savez ce que c'est (notion hors programme).
2. Montrer : $\forall x \in]-1,1[$, $\sum\limits_{n=0}^{+\infty} u_n x^n = \dfrac{x}{(1-x)^3}$, en utilisant un produit de Cauchy.
3. ☆ Montrer : $\dfrac{1}{(1-x)^3} = \sum\limits_{n=0}^{+\infty} \dfrac{n(n+1)}{2} x^n$, en admettant que l'on peut dériver la somme géométrique $\sum\limits_{n=0}^{+\infty} x^n$ « comme on pense » (à moins que vous ne trouviez une démonstration valable avec le programme de MPSI, ce qui n'est pas à exclure).
4. Conclure en donnant la valeur de u_n pour tout $n \in \mathbb{N}$.

Dans le contexte où l'on utilise les séries entières, a_n est souvent, au choix : 1° une somme non usuelle, 2° le cardinal d'un ensemble fini, dont on connaît des relations de récurrence grâce à des arguments combinatoires.

Notons qu'on a mis l'accent sur l'implication directe de (♣), mais l'implication réciproque (qui, elle, est totalement triviale) a aussi sa pertinence, pour les mêmes que celles données ci-dessus : une relation compliquée entre fonctions a parfois une traduction plus simple avec des suites et il vaut mieux dans ce cas travailler avec elles. Une situation récurrente qui illustre mon propos est celui des équations différentielles, comme nous allons l'illustrer immédiatement avec le cas de l'équation différentielle $y' = y$ munie de la condition initiale $y(0) = 1$. Faisons comme si nous ignorions ses solutions, et retrouvons le fait que l'exponentielle soit l'unique solution (vous constaterez des similitudes avec la résolution de la section 8.2 ; la philosophie est cependant différente ici). Avant de poursuivre, rappelons qu'il est relativement facile de démontrer l'unicité d'une solution : ← page 33

Exercice 62. Soient f et g deux fonctions dérivables sur $\mathbb{R}$ et solutions de l'équation différentielle : $y' = y$, et vérifiant de plus : $f(0) = g(0) = 1$. Montrer : $f = g$.
Indication : montrer que g ne s'annule pas grâce au calcul de $g(x)g(-x)$ pour tout $x \in \mathbb{R}$, puis étudier la dérivée de $\frac{f}{g}$.

Par conséquent, il suffit de trouver une seule solution pour résoudre entièrement ce problème de Cauchy, et je *choisis* de la chercher sous la forme d'une fonction développable en série entière. Soit $f : x \mapsto \sum_{n=0}^{+\infty} a_n x^n$ une telle fonction (ne nous soucions pas de son domaine de définition dans cette exposition : en toute rigueur on le trouverait en faisant une analyse-synthèse, l'analyse permettant d'expliciter suffisamment la suite $(a_n)_{n\geqslant 0}$ pour déterminer les valeurs de x où il y a convergence). On veut : $f' = f$, et : $f(0) = 1$. Si f est dérivable terme à terme comme on le promet plus haut, alors ces deux égalités équivalent à :

$$\sum_{n=0}^{+\infty} a_{n+1}(n+1)x^n = \sum_{n=0}^{+\infty} a_n x^n, \quad \text{et :} \quad a_0 = 1,$$

pour tout x dans le domaine de définition de f. D'après l'équivalence (♣), cela équivaut à :

$$\forall n \in \mathbb{N},\ a_{n+1} = \frac{a_n}{n+1}, \quad a_0 = 1.$$

On en déduit aisément, par récurrence : $\forall n \in \mathbb{N}$, $a_n = \dfrac{1}{n!}$, et donc f est nécessairement l'application $x \mapsto \sum_{n=0}^{+\infty} \dfrac{x^n}{n!}$: c'est la fonction exponentielle.

Nous avions dit plus haut que certaines relations compliquées entre fonctions ont parfois une traduction plus simple avec les suites. Dans le cas de figure présent, cela tient au fait qu'une équation différentielle se ramène à une relation de récurrence ; or une telle relation permet d'expliciter le terme général en un nombre fini d'étapes. C'est souvent plus accessible.

☆ **Exercice 63.** En suivant la même approche que ci-dessus, donner les solutions développables en série entière des équations différentielles suivantes. On admettra les résultats énoncés dans cette section, et on ne se souciera pas du domaine de définition I des solutions :

(a) $\forall x \in I$, $x^2(1-3x)y''(x) + 2x(2-9x)y'(x) + 2(1-9x)y(x) = -4$;
(b) $\forall x \in I$, $4x^2(3x-1)y''(x) + 6x(6x-1)y'(x) + (9x+2)y(x) = -4$;
(c) $\forall x \in I$, $xy''(x) + 2(4x^2+1)y'(x) + 24xy(x) = 0$;
(d) $\forall x \in I$, $x^2(x^2-2)y''(x) + 8x(x^2-1)y'(x) + 4(3x^2-1)y(x) = 12x + 8$;
(e) $\forall x \in I$, $(x^2-3)y''(x) + 7xy'(x) + 8y(x) = 0$;
(f) $y^{(3)} = y$.

J'ai fait en sorte que les relations de récurrence obtenues dans chaque question induisent une petite différence, les unes par rapport aux autres, dans l'explicitation.

Plus généralement, être une fonction développable en série entière est une telle contrainte que cela implique beaucoup de régularité. Entre autres choses, que nous verrons en exercices : le maximum d'une telle fonction sur un disque (si l'on raisonne avec la variable complexe) est toujours sur son bord ; si on contrôle la taille de la fonction sur $\mathbb{C}$, on contrôle la taille de ses dérivées (réfléchissez à quel point c'est différent de ce que vous constatez selon la variable réelle, où des fonctions très « petites » peuvent avoir des variations erratiques et donc des dérivées « grandes ») ; etc.

9.2 Résumé : objectif du chapitre

Après avoir défini les séries entières, nous démontrerons que leur domaine de convergence est un disque du plan complexe, centré en zéro (avec une incertitude sur l'inclusion du bord au domaine de convergence). Le rayon de ce disque, appelé *rayon de convergence*, est une donnée fondamentale et nous donnerons des méthodes pour l'obtenir (théorème de comparaison, adaptation de la règle de D'Alembert à ce cadre). Le comportement sur le bord du disque est subtil, et nous donnerons un énoncé permettant de trancher ce qu'il se passe dans un cas particulier : le théorème d'Abel radial.

Nous donnerons ensuite les propriétés de régularité d'une fonction développable en série entière, montrant notamment qu'elle est de classe C^∞ sur l'intérieur de son disque ouvert de convergence, et que toutes les opérations naturelles de l'analyse sont valables sur ce disque ouvert (dérivation, intégration), avec un peu plus de prudence dans le cas de l'intégration (on se limite à des segments). Ces énoncés permettront d'en déduire les équivalences (♠) et (♣) données plus haut, qui jouent un rôle central dans toute la théorie des séries entières pour les raisons déjà évoquées.

Réciproquement, on se demandera : étant donné une fonction, est-elle développable en série entière ? Un lien avec les théorèmes de Taylor sera explicité.

En application, en plus des motivations déjà données pour toutes les séries de fonctions : les séries entières aideront à la résolution d'équations fonctionnelles et différentielles non triviales et à l'étude de certaines suites difficiles à expliciter, souvent d'origine combinatoire.

9.3 Une autre application : généraliser les fonctions usuelles

Cette sous-section peut être ignorée. Elle motive un prolongement de la théorie des séries entières qui n'est explicitement au programme que dans le chapitre *Fonctions vectorielles* (et j'en parle donc dans la section idoine, page 49), et éventuellement dans certains problèmes écrits. Mais elle ne constitue en aucun cas le cœur de nos préoccupations en MP.

Une autre motivation de la théorie des séries entières, que le programme n'aborde qu'à la marge dans le chapitre *Fonctions vectorielles*, est qu'en écrivant les fonctions usuelles comme combinaison linéaire de puissances entières (pour caricaturer), nous rendons possible leur généralisation à d'autres contextes que le cadre réel ou complexe. En effet, dans le membre de droite des identités :

$$\forall x \in \mathbb{R},\ \exp(x) = \sum_{n=0}^{+\infty} \frac{x^n}{n!}, \quad \forall x \in]-1,1[,\ \frac{1}{1-x} = \sum_{n=0}^{+\infty} x^n, \quad \forall x \in]-1,1[,\ \ln(1+x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{n+1}}{n+1},$$

il n'apparaît que des sommes, produits et puissances impliquant des nombres rationnels et une variable x. De telles quantités ont un sens dans bien d'autres ensembles que $\mathbb{R}$ ou $\mathbb{C}$: il suffit par exemple d'avoir un anneau contenant $\mathbb{Q}$ (on a besoin de la structure d'anneau pour que le produit et l'exponentiation aient un sens). Si A est un tel anneau, et qu'on peut définir une topologie sur A de sorte que la convergence d'une série y ait un sens (et soit facile à étudier en imitant le cas réel ou complexe), il devient alors possible d'y définir un analogue des fonctions usuelles réelles grâce aux identités ci-dessus. Si l'on ne veut pas se poser de question de convergence, et ne manipuler que des sommes à support fini, c'est possible à condition d'avoir des éléments nilpotents. Il suffit alors de poser, par exemple, pour tout $a \in A$ tel que la somme existe : $\exp(a) = \sum_{n=0}^{+\infty} \frac{1}{n!} a^n$. En résumé :

Les séries entières permettent de généraliser les fonctions usuelles à d'autres anneaux que $\mathbb{R}$: $\mathbb{C}$, $M_n(K)$, $\mathbb{Z}/n\mathbb{Z}$, etc.

C'est d'ailleurs ce que vous fîtes pour définir de manière très satisfaisante l'exponentielle complexe, au moment d'aborder les séries numériques et les familles sommables, là où les définitions des années antérieures ont tout du bricolage (définition sur $\mathbb{R}$ soit comme solution d'une équation différentielle, soit comme réciproque du logarithme, puis définition sur $i\mathbb{R}$ à l'aide du cosinus et du sinus, et enfin extension à $\mathbb{C}$ par propriété de morphisme), et donnent des démonstrations inélégantes que les propriétés classiques sur $\mathbb{R}$ restent valables sur $\mathbb{C}$.

De même, le mathématicien qui voudrait définir le logarithme sur un domaine du plan complexe proposerait comme définition : $\ln(z) = \sum_{n=0}^{+\infty} (-1)^n \frac{(z-1)^{n+1}}{n+1}$. Malheureusement cette somme n'existe pas pour tout $z \in \mathbb{C}$ vérifiant $|z-1| > 1$, mais c'est mieux que rien. Pour $z \in \mathbb{C}$ tel que $|z-1| < 1$, on retrouve bien l'une des propriétés classiques du logarithme :

Exercice 64. *Sans se soucier de justifier la validité des opérations classiques de l'analyse*, vérifier que pour tout $z \in \mathbb{C}$ tel que : $|z-1| < 1$, on a : $\exp\left(\sum_{n=0}^{+\infty} (-1)^n \frac{(z-1)^{n+1}}{n+1}\right) = z$.

En revanche l'identité $\ln(ab) = \ln(a) + \ln(b)$ se généralise assez mal. N'en parlons pas.

L'intérêt de vouloir généraliser des fonctions usuelles à d'autres anneaux que $\mathbb{R}$, est de profiter de leurs propriétés remarquables en espérant qu'elles restent valables... Et c'est en général le cas ! En effet, toute identité vérifiée par des fonctions développables en série entière se ramène, en raisonnant coefficient par coefficient, à des identités polynomiales à coefficients entiers ou rationnels (par exemple l'identité $\exp(x+y) = \exp(x)\exp(y)$ équivaut à la formule du binôme de Newton, comme vous l'avez démontré cette année grâce aux produits de Cauchy). Et les identités polynomiales à coefficients entiers ont le don de pouvoir se généraliser à tout anneau A en remplaçant l'indéterminée X par n'importe quel élément $a \in A$, grâce au fait que $P \mapsto P(a)$ soit un morphisme de $\mathbb{Z}[X]$ dans A* !

Il faut simplement être plus prudent lorsqu'on veut généraliser une identité avec plus d'une seule variable x : elle ne se généralise qu'avec des éléments qui **commutent**, puisque dans le cas contraire $P(X,Y) \mapsto P(a,b)$ n'est plus un morphisme de $\mathbb{Z}[X,Y]$ dans A (ici $(a,b) \in A^2$).

Exercice 65. Soit $n \in \mathbb{N}$. Donner des exemples explicites, dans des anneaux non commutatifs A, où les identités suivantes sont *fausses* :

$$\forall (a,b) \in A^2, \quad (a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, \quad a^n - b^n = (a-b) \sum_{k=0}^{n-1} a^k b^{n-1-k},$$

et démontrer qu'elles sont valables si a et b commutent.

En dehors de ces légères complications, on retiendra le principe général de cette sous-section :

1. Les séries entières permettent de généraliser des fonctions réelles à d'autres anneaux que $\mathbb{R}$.
2. Les identités vérifiées par ces fonctions restent valables quand on les définit sur un autre anneau, à condition d'ajouter une hypothèse de commutation lorsque ces identités font intervenir au moins deux variables.

Illustrons ceci avec le développement en série entière de $x \mapsto \frac{1}{1-x}$. L'intérêt de généraliser cette fonction est évident : cela permet de fabriquer des inverses à peu de frais. Pour ne pas se soucier des questions de convergence, l'exercice ci-dessous ne nécessite d'utiliser son développement qu'avec des éléments nilpotents, de sorte que les sommes soient à support fini.

Exercice 66. Traiter chacune des questions ci-dessous en s'inspirant du développement en série entière de $x \mapsto \frac{1}{1-x} = (1-x)^{-1}$.

1. Trouver rapidement un inverse de 8 modulo 3^6.

*. Par souci de clarté dans l'exposition, je mets de côté les complications qui apparaissent lorsque A ne contient pas de sous-anneau isomorphe à $\mathbb{Q}$ et qu'on considère des identités avec des nombres rationnels.

2. Soient $a \in \mathbb{R}$, n un entier naturel non nul, et :

$$A = \begin{pmatrix} 1 & a & a^2 & a^3 & \cdots & a^{n-1} \\ 0 & 1 & a & a^2 & \cdots & a^{n-2} \\ 0 & 0 & 1 & a & \cdots & a^{n-3} \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \in \mathrm{M}_n(\mathbb{R}).$$

   Montrer que A est inversible et donner son inverse, *de tête*.

3. Soit A un anneau dont on note 0_A et 1_A les éléments neutres pour l'addition et la multiplication. Montrer que si $n \in A$ est un élément nilpotent (c'est-à-dire : $\exists k \in \mathbb{N} \setminus \{0\}$, $n^k = 0_A$), alors $1_A + n$ est inversible et donner son inverse.

   Traiter de même l'inversibilité de $u + n$ où u est un élément inversible de A commutant avec n.

4. Soit M une matrice d'ordre n nilpotente. Montrer que $(\mathrm{I}_n - M)^2$ est inversible et expliciter son inverse.

Pour les généralisations de l'exponentielle, nous en parlons plus amplement dans le chapitre 11 (*Fonctions vectorielles*). Voir plus spécifiquement la section 11.2. → page 50

10 Probabilités discrètes

Les probabilités de 2^e année consistent surtout en l'extension des résultats de 1^{re} année, dans un cadre où les univers peuvent éventuellement être de cardinal infini. C'est souvent nécessaire lorsqu'on veut modéliser une expérience faisant apparaître (au moins en esprit) une infinité d'étapes, ou bien un nombre fini d'étapes qui peut néanmoins être arbitrairement élevé. Un exemple standard est celui d'une expérience ayant probabilité p d'aboutir à un succès, et qu'on répète jusqu'à enfin observer un succès (ou, cela revient au même si l'on y réfléchit bien : on répète l'expérience une infinité de fois et on observe le rang d'apparition du premier succès).

Exercice 67. Soient $n \in \mathbb{N}\setminus\{0\}$ et $p \in]0,1[$. On lance n fois une pièce non équilibrée, ayant probabilité p de faire face. Mathématiquement, cela peut se modéliser par la donnée de n variables aléatoires indépendantes $X_1, \ldots, X_n$ suivant la loi de Bernoulli de paramètre p. Soit X la variable aléatoire valant $+\infty$ si l'on n'obtient jamais face, et valant $k \in [\![1, n]\!]$ si le premier face apparaît au k^{e} lancer. Autrement dit : $X = \inf\{i \in [\![1, n]\!] \mid X_i = 1\}$ (avec la convention $\inf \emptyset = +\infty$).

1. Donner la loi de X.
2. ☆ On reprend la même expérience , mais en lançant une infinité de fois la pièce. Que s'attend-on à avoir pour $\mathrm{P}(X = k)$, pour tout $k \in \mathbb{N} \setminus \{0\}$, dans ce cas ? (Nous manquons à ce stade de théorèmes pour justifier que le raisonnement très tentant à effectuer est effectivement correct.)

On définira la loi géométrique pour modéliser ce genre d'expérience.

L'extension aux univers infinis ne pose presque pas de difficulté, tant qu'on se limite aux univers *en bijection avec* $\mathbb{N}$ (c'est-à-dire aux ensembles dits *dénombrables*). Pour comprendre le type de problème posé hors de ce cadre, on pourra méditer l'exercice suivant :

Exercice 68. On lance une pièce une infinité de fois, ce qu'on modélise par l'univers : $\Omega = \{0,1\}^{\mathbb{N}}$. Il est tentant d'affirmer que la probabilité uniforme $\{0,1\}$ en induit une sur Ω, comme c'est le cas pour les produits cartésiens finis. Notons P une telle probabilité sur Ω, si elle existe.

1. Soit $\omega \in \Omega$. Pour tout $n \in \mathbb{N}$, notons A_n l'évènement : « les n premiers lancers coïncident avec les n premiers termes de ω. » Calculer $\mathrm{P}(A_n)$ pour tout $n \in \mathbb{N}$, et en déduire : $\mathrm{P}(\{\omega\}) = 0$.
2. Comparer $\mathrm{P}(\Omega)$ et $\sum_{\omega \in \Omega} \mathrm{P}(\{\omega\})$. Est-ce que le résultat vous paraît normal ?

Si le sens à donner à cette somme vous embête, remplacez Ω dans l'indexation par des parties FINIES *$\Omega_n \subseteq \Omega$ telles que : $\forall n \in \mathbb{N}$, $\Omega_n \subseteq \Omega_{n+1}$, et : $\Omega = \bigcup_{n\in\mathbb{N}} \Omega_n$, puis prenez la limite quand $n \to +\infty$.*

3. Vérifier que Ω n'est pas en bijection avec $\mathbb{N}$. Pour cela, raisonner par l'absurde en supposant l'existence d'une bijection $f : \mathbb{N} \to \Omega$, et montrer que l'élément de Ω suivant fournit une contradiction : $\omega = (1 - f(n)_n)_{n\in\mathbb{N}}$. *On a noté $f(n)_n$ le n^e terme de la suite $f(n)$.*
   Ce raisonnement par l'absurde est à comparer aux deux dernières questions de l'exercice 18. ← page 10

À partir du moment où le cadre théorique est correctement défini pour faire des probabilités sur des univers infinis (modulo une certaine contrainte formulée ci-dessus), de nouvelles questions peuvent se poser :

— sur la réalisation asymptotique d'un évènement lorsqu'on itère une expérience un nombre arbitrairement grand de fois ;

— sur la « convergence » d'une suite de variables aléatoires, ou du moins leur approximation (puisque le terme de convergence n'est pas défini en MP pour des variables aléatoires).

10.1 Étude asymptotique d'un évènement

À partir du moment où il est permis de formuler des expériences ayant une infinité d'étapes (ou du moins un nombre fini d'étapes pouvant être arbitrairement élevé, comme : « on répète le lancer d'une pièce tant qu'on n'a pas obtenu “face” »), il est possible de s'interroger sur la probabilité d'observer tel phénomène *au moins une fois* dans cette série infinie d'expériences (par exemple : avoir au moins une fois « face » dans une série infinie de lancers d'une pièce), ou d'observer tel phénomène *systématiquement*. Formuler de tels évènements mathématiquement nécessite d'écrire une réunion ou intersection infinie d'évènements :

Exercice 69. On lance une pièce une infinité de fois, et on le modélise avec l'univers $\Omega = \{0,1\}^{(\mathbb{N}\setminus\{0\})}$ (l'exercice 68 montre que cette modélisation peut être problématique, mais pour ce que je demande dans cet exercice ce ne sera pas le cas puisque je ne vais pas parler de probabilités : seulement d'évènements). Dans cette modélisation, on peut imaginer que le nombre 1 représente « face ».

Pour tout $i \in \mathbb{N} \setminus \{0\}$, soit A_i l'évènement : « le i^e lancer donne “face” ».

1. Écrire en termes ensemblistes l'évènement : « on obtient “face” à chaque lancer de cette série infinie de lancers ».
2. Écrire en termes ensemblistes l'évènement : « on obtient “face” au moins une fois dans cette série infinie de lancers ».
3. Écrire en termes ensemblistes l'évènement : « on obtient “face” une infinité de fois dans cette série infinie de lancers ».
4. Écrire en termes ensemblistes l'évènement : « on finit par obtenir exclusivement “face” au-delà d'un certain nombre de lancers ».

Avec les outils de 1re année, calculer la probabilité d'une réunion ou intersection infinie est inaccessible. En seconde année, nous donnerons un certain nombre de résultats le permettant, dépendant des hypothèses sur les évènements de la réunion ou intersection. L'un d'eux est le *théorème de continuité monotone* qui nous dit essentiellement que le raisonnement naïf suivant : « lorsqu'on veut calculer la probabilité d'un évènement se produisant en une infinité d'étapes, il suffit de faire le calcul pour n étapes, puis de prendre la limite quand $n \to +\infty$ », est un raisonnement correct.

Un autre cadre faisant intervenir une étude asymptotique de la probabilité d'un évènement, est donné par les chaînes de Markov : on répète une certaine expérience une infinité de fois, dont l'état à la $(n+1)^e$ itération dépend uniquement de l'issue de la n^e itération (absence de mémoire). On se demande quelles issues de l'expérience sont les plus susceptibles d'advenir après un très grand nombre d'étapes. Nous donnons de tels exemples dans les exercices ci-dessous :

Exercice 70. Une abeille va chaque jour sur l'une des deux fleurs A et B. Au jour 0, elle va à la fleur A. À chaque nouvelle journée, il y a une probabilité $p \in]0,1[$ qu'elle aille sur la même fleur que la veille. Pour tout entier n, on note A_n l'évènement : « l'abeille est sur la fleur A au jour n » et B_n l'évènement : « l'abeille est sur la fleur B au jour n ». On pose $a_n = \mathrm{P}(A_n)$ et $b_n = \mathrm{P}(B_n)$.

1. Pour tout entier n, exprimer a_{n+1} et b_{n+1} en fonction de a_n et b_n.
2. En remarquant que $a_n + b_n = 1$, déterminer les expressions de a_n et b_n en fonction de n pour tout $n \in \mathbb{N}$.
3. Vers quoi tendent les suites $(a_n)_{n\geqslant 0}$ et $(b_n)_{n\geqslant 0}$? Interpréter.

Exercice 71. On considère un carré ABCD et son centre, noté O. Lorsqu'un jeton est placé sur l'un de ces cinq points, il doit, à chaque tour, se déplacer vers l'un des points voisins de manière équiprobable. Au départ, le jeton est en A. Pour tout $n \in \mathbb{N} \setminus \{0\}$, on note p_n la probabilité que le jeton soit en O après n tours. Démontrer que : $\forall n \in \mathbb{N} \setminus \{0\}$, $p_{n+1} = \frac{1}{3}(1 - p_n)$, et en déduire l'expression de p_n en fonction de $n \in \mathbb{N} \setminus \{0\}$. Quelle est la limite de $(p_n)_{n\geqslant 1}$ quand n tend vers $+\infty$?

Dans ces deux exercices, on s'en sort parce qu'il n'y a qu'un nombre raisonnable d'issues à l'expérience à chaque étape. Si l'expérience est plus complexe, ce n'est plus possible... À moins de tirer profit des enseignements du chapitre de réduction des matrices (section 6). Plus précisément, ← page 17 l'exercice 39 illustre comment nous pouvons expliciter des suites réelles ou complexes, même lorsqu'elles vérifient des relations de récurrence sophistiquées, en passant par l'étude d'une suite de vecteurs colonnes $(X_n)_{n\geqslant 0}$ vérifiant une relation de récurrence d'ordre 1 : $\forall n \in \mathbb{N}$, $X_{n+1} = AX_n$, avec A une matrice carrée. Nous pourrons en faire autant avec les chaînes de Markov, grâce à un bon usage de la formule des probabilités totales :

Exercice 72. Pour étudier les capacités d'adaptation des souris, des biologistes ont mis au point le protocole suivant. Ils placent une souris dans une boîte comportant trois issues A, B et C. Les deux premières conduisent à un cul-de-sac alors que la troisième conduit à un morceau de gruyère. Après que la souris a choisi son issue, les scientifiques la remettent au centre de la boîte pour répéter l'expérience. Ils observent les résultats suivants :

— si la souris choisit la sortie A, elle sort la fois suivante en B ou C de façon équiprobable ;

— si elle choisit la sortie B, elle sort la fois suivante en A ou C avec la même probabilité ;

— si elle choisit la sortie C, elle la choisit de nouveau systématiquement la fois suivante.

Pour tout $n \in \mathbb{N} \setminus \{0\}$, on note A_n (respectivement B_n, C_n) l'évènement : « la souris a choisi la sortie A (respectivement B, C) à la n^{e} expérience », et on pose : $a_n = \mathrm{P}(A_n)$, $b_n = \mathrm{P}(B_n)$ et $c_n = \mathrm{P}(C_n)$, et enfin : $X_n = \begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix}$. Déterminer une matrice $M \in \mathrm{M}_3(\mathbb{R})$ telle que : $\forall n \in \mathbb{N} \setminus \{0\}$, $X_{n+1} = MX_n$.

On utilisera la formule des probabilités totales.

Exercice 73. Christian Renauld, Lionel Messire et Célian Baptiste font un « jeu » de survie. On suppose que chacun d'entre eux choisit au hasard (uniformément) une cible parmi les autres joueurs survivants, et lui tire dessus. Pour simplifier, les tirs des différents participants ont la même probabilité de succès $\frac{2}{3}$, et les succès des tirs sont indépendants les uns des autres. Tout le monde tire en même temps. On répète ce « jeu » jusqu'à ce qu'il n'y ait plus qu'un seul survivant, ou aucun.

Pour tout $n \in \mathbb{N} \setminus \{0\}$, on note les évènements Z_n : « il n'y a plus de survivant après n tirs (ou un nombre de tirs inférieur à n) », U_n : « il reste un survivant après n tirs (ou un nombre de tirs inférieur à n) », D_n : « il reste deux survivants après n tirs », et enfin T_n : « il reste trois survivants après n tirs ». Pour tout $n \in \mathbb{N} \setminus \{0\}$, soit X_n le vecteur colonne de $\mathrm{M}_{4,1}(\mathbb{R})$ dont les coordonnées sont respectivement $\mathrm{P}(\mathrm{Z}_n)$, $\mathrm{P}(\mathrm{U}_n)$, $\mathrm{P}(\mathrm{D}_n)$ et $\mathrm{P}(\mathrm{T}_n)$. Montrer :

$$\forall n \in \mathbb{N} \setminus \{0\}, \quad X_{n+1} = \begin{pmatrix} 1 & 0 & 4/9 & 2/27 \\ 0 & 1 & 4/9 & 5/9 \\ 0 & 0 & 1/9 & 1/3 \\ 0 & 0 & 0 & 1/27 \end{pmatrix} X_n.$$

La détermination de la dernière colonne nécessite de la patience, parce qu'il faut considérer plusieurs cas, pour passer de trois survivants à deux ou un, selon que chaque joueur ait une cible différente ou non. *On utilisera la formule des probabilités totales.*

Sachant étudier asymptotiquement des suites « géométriques » à valeurs matricielles, nous saurons en déduire des probabilités asymptotiques. Ce n'est pas le seul chapitre où l'algèbre linéaire viendra au secours d'autres domaines des mathématiques (voir le chapitre sur les équations différentielles linéaires, section 12). → page 52

10.2 Approximation de variables aléatoires

L'approximation de variables aléatoires est en fait une idée parmi les plus naturelles des probabilités, et nous allons l'approfondir (en restant dans le cadre du programme, qui n'introduit pas de variables aléatoires dont l'image est non dénombrable : il s'avère que c'est un facteur très limitant).

Un exemple où, je crois, l'idée est naturelle : si l'on veut modéliser (informatiquement ou en esprit) ce qu'est un tirage uniformément au hasard d'un nombre RÉEL de [0,1], on peut remplacer l'intervalle [0,1] par la donnée de N points équirépartis dans cet intervalle, et tirer au hasard un nombre parmi ces N points : si N est arbitrairement grand, l'intuition laisse entendre que « cela revient presque au même ». Formalisons :

Exercice 74. Pour tout $n \in \mathbb{N} \setminus \{0\}$, on note E_n l'ensemble : $E_n = \left\{ \dfrac{k}{n} \mid k \in [\![0, n]\!] \right\}$, et X_n est une variable aléatoire réelle discrète suivant la loi uniforme sur E_n.

1. Soit $x \in [0,1]$. Montrer : $\forall n \in \mathbb{N} \setminus \{0\}$, $\mathrm{P}(X_n \leqslant x) = \dfrac{\lfloor nx \rfloor + 1}{n+1}$, et en déduire :
$$\lim_{n \to +\infty} \mathrm{P}(X_n \leqslant x) = x.$$
2. Pour tout $(a, b) \in [0,1]^2$ tel que : $a \leqslant b$, montrer : $\lim\limits_{n \to +\infty} \mathrm{P}(X_n \in [a, b]) = b - a$.
3. Soit $(c, d) \in \mathbb{R}^2$ tel que : $c < d$. On pose à présent : $\forall n \in \mathbb{N} \setminus \{0\}$, $E_n = \left\{ a + k\dfrac{b-a}{n} \mid k \in [\![0, n]\!] \right\}$, et X_n est une variable aléatoire réelle discrète suivant la loi uniforme sur E_n. Pour tout $(a, b) \in [c, d]^2$ tel que : $a \leqslant b$, montrer : $\lim\limits_{n \to +\infty} \mathrm{P}(X_n \in [a, b]) = \dfrac{b-a}{d-c}$.

C'est le même principe qui est utilisé dans la méthode de Monte-Carlo.

Un autre exemple est donné dans l'exercice ci-dessous, qui formalise l'idée que quand on fait un tirage au hasard dans un ensemble extrêmement grand, tirer avec ou sans remise « revient presque au même », puisque même dans le cas d'une remise il est très peu probable de tomber deux fois sur le même élément. L'intérêt de l'approximation est qu'avec remise, le dénombrement est plus simple et on se ramène aisément à une loi usuelle.

Exercice 75. L'équipe de football de Lionel Messire, l'En Avant Montvendre (localisée près de Barcelonne dans la Drôme), possède N joueurs sous contrat fédéral, dont n joueurs qui ont régulièrement recours à des méthodes médicales illicites pour améliorer leurs performances.

Un beau jour, un contrôle anti-dopage fait irruption dans leur vestiaire, et choisit uniformément au hasard ℓ joueurs parmi les N de l'équipe, pour procéder à une analyse sanguine qui détecterait quasi-certainement les joueurs en fraude. On note $(\Omega, \mathcal{A}, \mathrm{P})$ un espace probabilisé modélisant cette expérience, et X_N la variable aléatoire comptant, parmi les ℓ joueurs choisis par le contrôle anti-dopage, combien sont *effectivement* dopés.

1. Donner $X_N(\Omega)$, et justifier : $\forall k \in X_N(\Omega)$, $\mathrm{P}(X_N = k) = \dfrac{\binom{n}{k}\binom{N-n}{\ell-k}}{\binom{N}{\ell}}$.
2. Montrer : $\forall k \in \mathbb{N}$, $\dbinom{M}{k} \underset{M\to+\infty}{\sim} \dfrac{M^k}{k!}$.
3. On admet dans cette question que $n = pN$ où $p \in]0,1[$ est une constante. Montrer : $\forall k \in X_N(\Omega)$, $\lim\limits_{N\to+\infty} \mathrm{P}(X_N = k) = \dbinom{\ell}{k} p^k(1-p)^{\ell-k}$. Que reconnaît-on ? À la lumière du paragraphe qui précède l'exercice, expliquer pourquoi ce résultat était prévisible.

Nous étudierons le cas plus particulier d'une suite de variables aléatoires suivant une loi binomiale dont le paramètre n (donnant le nombre d'épreuves) devient arbitrairement grand ; pour songer à des situations réelles où cela se produit, il suffit de penser à l'observation du nombre de personnes entrant dans un magasin dans une journée. Il est possible pour cela de regarder, à chaque minute de la journée, si quelqu'un est entré (cela correspond à $n = 24 \times 60 = 1440$). On y associe une probabilité p_n dont la détermination dans la vraie vie s'obtient par observation fréquentielle. Mais on peut affiner cette observation en observant plutôt à chaque *seconde* de la journée si quelqu'un est entré (on a alors $n = 86400$). Intuitivement, plus on segmente la journée en un nombre n d'intervalles de temps, et plus la modélisation est précise ; mais le prix à payer est que les calculs impliquant la loi binomiale de paramètre (n, p) deviennent de plus en plus complexes : songez à la présence de nombres tels que $n! = 86400!$... On va alors chercher une *approximation* de la loi binomiale, et obtenir ce qu'on appelle la *loi de Poisson*. L'exercice qui suit y initie et permet de conjecturer sa définition :

Exercice 76. Pour tout $n \in \mathbb{N} \setminus \{0\}$, soit X_n une variable aléatoire telle que : $X_n \sim \mathcal{B}(n, p_n)$, avec $p_n \in]0,1[$. On suppose : $\lim\limits_{n\to+\infty} np_n = \lambda \neq 0$. Montrer : $\forall k \in \mathbb{N}$, $\lim\limits_{n\to+\infty} \mathrm{P}(X_n = k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$.

Seulement ces approximations de lois de variables aléatoires par passage à la limite (le terme savant est celui de *convergence en loi* de variables aléatoires, mais il n'est pas au programme) ne fournissent pas suffisamment d'information. Par exemple, la sagesse populaire connaît la loi des grands nombres, que nous formulerons en ces termes : si $X_1, \ldots, X_n$ sont n variables aléatoires indépendantes suivant la même loi (imaginez par exemple qu'on lance un dé équilibré à six faces n fois, et que X_i vaut 1 si le i^{e} lancer, et 0 sinon), alors la moyenne empirique $\dfrac{X_1 + \cdots + X_n}{n}$ se rapproche de l'espérance de X_1, d'autant plus que n est très grand. C'est ainsi que vous avez pu observer que lorsqu'on lance un dé un très grand nombre de fois, chaque face apparaît presque avec la proportion $\dfrac{1}{6}$ (qui correspond bien à l'espérance de $X_1 \sim \mathcal{B}\left(\frac{1}{6}\right)$). Mais ce n'est pas très précis : combien de lancers doit-on faire si l'on espère avoir moins de 1% d'écart entre la moyenne empirique et l'espérance ? et si l'on veut 0,1% ? etc. Pour répondre à cette question, on doit plutôt étudier des probabilités du type :

$$\mathrm{P}(|X - Y| \geqslant \varepsilon),$$

où $\varepsilon > 0$ et X et Y sont deux variables aléatoires dont l'une est une approximation de l'autre. En illustration, nous montrerons la *loi faible des grands nombres* selon laquelle, en reprenant les notations ci-dessus :

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathrm{E}(X_1)\right| \geqslant \varepsilon\right) = \underset{n\to+\infty}{O}\left(\frac{1}{n}\right)$$

(l'énoncé sera même un peu plus précis que cela : c'est la variance qui permet de mesurer la taille du grand o). Vous pouvez vous convaincre que c'est bien une reformulation savante de la loi des grands nombres telle que nous la formulions ci-dessus.

Les « inégalités de concentration », qui sont d'autres inégalités du même type (en variant les hypothèses sur les variables aléatoires, etc.), sont une source féconde d'exercices de probabilités.

10.3 Résumé : objectif du chapitre

Après avoir étendu la théorie des probabilités sur un univers fini au cas infini (à condition que l'univers soit en bijection avec $\mathbb{N}$), ce qui consistera notamment en la démonstration de formules permettant de calculer la probabilité d'une réunion ou intersection *infinie* d'évènements, nous verrons comment les outils d'algèbre linéaire de 2^e seconde nous permettent d'étudier l'issue la plus probable d'une chaîne de Markov (qui correspond, en gros, à une expérience répétée plusieurs fois, où l'état du système à la $(n+1)^{\text{e}}$ expérience ne dépend que de l'issue de la n^{e}).

Puis une préoccupation majeure du chapitre concernera l'étude asymptotique d'une suite de variables aléatoires, qui se compartimente essentiellement en deux problématiques : 1° approcher des variables aléatoires par des variables aléatoires dont la loi est plus simple (lorsque les paramètres de leurs lois ont un comportement asymptotique appréciable), 2° mesurer la probabilité que l'écart entre une variable aléatoire X et son approximation Y soit supérieur à une quantité ε donnée, grâce à l'inégalité de Markov ou de Bienaymé-Tchebychev. C'est ainsi que nous en déduirons notamment la loi faible des grands nombres.

Enfin, nous introduirons (dans le cas de variables aléatoires à valeurs dans $\mathbb{N}$) la *fonction génératrice* d'une variable aléatoire : c'est une fonction développable en série entière. Nous en parlons dans la section 9.1, et disions entre autres qu'il était parfois plus intéressant d'étudier une suite $(a_n)_{n\geqslant 0}$ en ← page 37 passant par la fonction $x \mapsto \sum_{n=0}^{+\infty} a_n x^n$. Suivant cette même idée, nous verrons que pour toute variable aléatoire X à valeurs dans $\mathbb{N}$, la fonction $G_X : x \mapsto \sum_{n=0}^{+\infty} \mathrm{P}(X = n)x^n$ l'encode complètement : la connaissance explicite de la fonction G_X permet (au moins en théorie) d'obtenir toutes les informations utiles concernant X : sa loi, son espérance, sa variance, etc. Cela tombe bien, parce que pour certaines variables aléatoires (notamment celles définies comme une somme de variables aléatoires *indépendantes*), il est plus facile d'étudier sa fonction génératrice que la variable aléatoire elle-même, pour des raisons déjà exposées dans la section 9.1.

11 Dérivation et intégration des fonctions vectorielles

11.1 Fonctions à valeurs dans un espace vectoriel normé

Ce chapitre complète l'étude des fonctions définies sur un intervalle I de $\mathbb{R}$ et à valeurs dans un K-espace vectoriel normé E. Cette étude fut déjà faite en grande partie dans le chapitre de topologie, ← page 25 et elle se résumait en quelques mots ainsi : une fonction $f : I \to E$ vérifie une certaine propriété $\mathcal{P}$ si et seulement si ses composantes $f_i : I \to K$ dans une base de E vérifient $\mathcal{P}$. Ce qui est bien commode, puisque cela nous ramène à une situation connue et plus simple (l'espace ambiant de départ et d'arrivée est de dimension 1).

Cette équivalence fut établie pour : avoir une limite, être continue, être bornée, être lipschitzienne. Nous établissons dans ce chapitre pour la dérivabilité (que nous aurons bien sûr définie dans ce cadre) : si $(\vec{e}_1, \ldots, \vec{e}_n)$ est une base de E, et si : $\forall x \in I$, $f(x) = \sum_{i=1}^{n} f_i(x)\vec{e}_i$, alors f est dérivable si et seulement si f est dérivable pour tout i ; le cas échéant, on a : $\forall x \in I$, $f'(x) = \sum_{i=1}^{n} f_i'(x)\vec{e}_i$. En particulier, si $A : I \to \mathrm{M}_n(K)$ est une application dérivable, il suffit de dériver toutes ses composantes $a_{i,j} : I \to K$ pour obtenir A'.

Cette simplicité permet de généraliser plusieurs théorèmes connus pour les fonctions réelles ou complexes, en raisonnant coordonnée par coordonnée : caractérisation des fonctions dérivables constantes, théorème de la limite de la dérivée, etc. De même pour la classe C^k plus généralement.

Seulement il n'est pas toujours agréable de calculer la dérivée d'une fonction vectorielle en explicitant ses coordonnées. Par exemple, si A et B sont deux applications de I dans $\mathrm{M}_n(K)$ et dérivables, est-on vraiment obligé de se casser la tête à calculer chaque composante du produit matriciel $A \cdot B$, puis de les dériver, pour en déduire $(A \cdot B)'$? À cet effet nous donnerons des formules de dérivation qui auront l'avantage d'être intrinsèques (c'est-à-dire ne nécessitant pas un choix de base), notamment concernant la dérivation de :

— la composition d'une application $f : I \to E$ avec une application linéaire ;

— la composition d'une application $(f_i)_{1 \leqslant i \leqslant k} : I \to \prod_{i=1}^{k} E_i$ avec une application k-linéaire (le produit matriciel en étant un cas particulier).

Nous verrons en particulier que toutes les compositions avec des applications bilinéaires se dérivent selon le modèle du produit $(uv)' = u'v + uv'$:

$$\langle f, g\rangle' = \langle f', g\rangle + \langle f, g'\rangle, \quad (\det_{\mathcal{B}}((f, g)))' = \det_{\mathcal{B}}((f', g)) + \det_{\mathcal{B}}((f, g')), \quad \text{etc.}$$

Le cas de la dérivation étant réglé, son *alter-ego* (l'intégration) aura aussi droit à sa généralisation aux applications continues par morceaux de I dans E, sans subtilité majeure.

11.2 Une fonction vectorielle remarquable : exponentielle d'un endomorphisme, d'une matrice

Les chapitres *Topologie des espaces vectoriels normés*, *Suites et séries de fonctions* et *Séries entières*, conjointement à celui-ci, permettent de définir et d'étudier la fonction exponentielle d'un endomorphisme d'un espace vectoriel de dimension finie, ou d'une matrice carrée, en posant :

$$\forall f \in \mathrm{L}(E),\ \exp(f) = \sum_{n=0}^{+\infty} \frac{1}{n!} f^n, \quad \forall M \in \mathrm{M}_p(K),\ \exp(M) = \sum_{n=0}^{+\infty} \frac{1}{n!} M^n.$$

Exercice 77. Calculer $\exp(M)$ avec des exemples concrets de matrices M diagonales. Vérifier au moins dans ce cas que l'interprétation de la somme du membre de droite est facile.

Nous l'avons dit dans la section 9.3 : l'intérêt de définir la fonction exponentielle sur $\mathrm{L}(E)$ ou $\mathrm{M}_p(K)$ est de pouvoir résoudre les mêmes problèmes dans ces espaces que dans $\mathbb{R}$ ou $\mathbb{C}$, en espérant que les propriétés de l'exponentielle se préservent. Il serait impossible de résumer en peu de mots les nombreuses applications de l'exponentielle, mais en voici deux parmi d'autres pour illustrer mon propos : ← page 42

— la résolution des équations différentielles linéaires ;

— la détermination des racines complexes de l'unité (et plus généralement des racines d^{es} de n'importe quel nombre complexe, mais pour simplifier la discussion je vais me limiter aux racines de l'unité).

Concernant le premier point : on sait en effet démontrer que les solutions de l'équation différentielle : $y' = ay + b$ (avec a et b deux fonctions continues) sont de la forme $y : t \mapsto e^{A(t)} \left(\lambda + \int_{\bigstar}^{t} e^{-A(x)} b(x) \mathrm{d}x\right)$, où A est une primitive de a et λ une constante quelconque.

Exercice 78. Vérifier cette affirmation. Pour gagner du temps, on peut remarquer qu'elle se démontre très rapidement en multipliant l'égalité : $y' - ay = b$, par une fonction convenable, de sorte à reconnaître dans le membre de gauche la dérivée d'un produit de fonctions.

La clé, pour cela, est le fait que e^A ait pour dérivée ae^A.

Or nous verrons au chapitre *Équations différentielles linéaires* que TOUTE équation différentielle linéaire et TOUT système différentiel linéaire se ramènent à savoir résoudre : $Y' = AY + B$, où A et B sont des fonctions vectorielles à valeurs dans $\mathrm{M}_p(K)$ et $\mathrm{M}_{p,1}(K)$ respectivement (pour un certain entier naturel non nul p qui dépend de l'ordre des dérivées, ainsi que du nombre de fonctions inconnues, apparaissant dans l'équation ou le système), et Y une fonction vectorielle à déterminer. Suivant la discussion ci-dessus, il est raisonnable d'espérer que les solutions de l'équation $Y' = AY + B$ s'obtiennent encore une fois à l'aide de l'exponentielle, matricielle cette fois-ci.

Concernant le second point, je me permets une digression conceptuelle assez longue car, je pense, elle est instructive sur l'intérêt des morphismes en général : la raison pour laquelle l'exponentielle classique permet de produire si facilement des racines de l'unité provient du fait que ce soit un morphisme de groupes surjectif de $(\mathbb{R}, +)$ (ou $(\mathbb{C}, +)$) dans $(\mathbb{R}^*, \times)$ (ou $(\mathbb{C}^*, \times)$). Or un morphisme préserve relativement bien les « racines de l'élément neutre » (et même parfaitement bien si c'est un *isomorphisme*) dans un groupe *. Précisons ce qu'on entend par là :

Exercice 79. Soient G et H deux groupes, et f un morphisme de groupes de G dans H.

1. Soient $x \in G$ et $d \in \mathbb{N} \setminus \{0\}$. On pose : $y = f(x) \in H$. Montrer que si : $x^d = 1_G$, alors : $y^d = 1_H$, et montrer que l'implication réciproque est fausse en général.
2. On conserve les notations de la question précédente, et on suppose de plus que f est *injectif* (uniquement dans cette question). Montrer : $x^d = 1_G \iff y^d = 1_H$.
3. On suppose dans cette question que G et H sont des groupes commutatifs, et que f est *bijectif*. Montrer que les ensembles $\mathbb{U}_{G,d} = \{x \in G \mid x^d = 1_G\}$ et $\mathbb{U}_{H,d} = \{y \in H \mid y^d = 1_H\}$ sont des groupes et qu'ils sont isomorphes pour tout $d \in \mathbb{N} \setminus \{0\}$. Que reste-t-il de ce résultat si G et H ne sont pas supposés commutatifs ?

Autrement dit : si f est un isomorphisme alors, connaissant les racines d^{es} de l'unité de G, on en déduit celles de H (et inversement). Si f n'est pas bijectif, on a tout de même mieux que rien :

4. On suppose à présent que f est surjectif. Soient $y \in H$ et $d \in \mathbb{N} \setminus \{0\}$. Montrer :

$$\mathbb{U}_{H,d} = \left\{ y \in H \mid \exists x \in G,\ \exists e \in \ker(f),\ y = f(x) \text{ et } x^d = e \right\}.$$

La dernière question de cet exercice nous dit, en substance : pour obtenir les racines d^{es} de l'unité dans H, il suffit de calculer les racines d^{es} des éléments du noyau de f, où f est un morphisme de groupes surjectif d'un groupe G dans H (une fois que c'est fait, on prend l'image par f des racines ainsi obtenues) : l'affaire est rentable **s'il est plus facile d'extraire des racines d^{es} dans G que dans H, et c'est souvent le cas si la loi de G est additive** ! (Quand la loi est additive, l'équation $x^d = e$ d'inconnue x s'écrit plutôt $dx = e$, et il suffit de « savoir diviser par d » pour en déduire x.) C'est ce qui explique l'apport de l'exponentielle dans ce cadre : c'est un morphisme de groupes surjectif de $(\mathbb{C}, +)$ dans $(\mathbb{C}^*, \times)$, dont le noyau est $2i\pi\mathbb{Z}$. Par ce qui précède, pour obtenir les racines complexes d^{es} de l'unité, il suffit de savoir résoudre l'équation $dx = 2i\pi k$ d'inconnue $x \in \mathbb{C}$, ce qui est complètement trivial, puis de prendre l'image par l'exponentielle des solutions trouvées. On retrouve la description bien connue : $\mathbb{U}_d = \left\{ e^{\frac{2i\pi k}{d}} \mid k \in \mathbb{Z} \right\}$.

J'insiste sur le fait que la loi du groupe de départ est additive : c'est la clé qui rend tout plus simple dans $\mathbb{C}$ que dans $\mathbb{C}^*$, et qui fait de l'exponentielle un morphisme remarquablement utile. Comparer ce discours avec celui tenu dans la section 4 (*Structures algébriques*), sur l'intérêt des morphismes (et plus spécifiquement des isomorphismes). ← page 11

Si je reformule ce problème classique en des termes si savants (parce qu'au fond, vous savez calculer des racines complexes d^{es} de l'unité sans parler de morphismes de groupes), c'est pour identifier ce dont on a réellement besoin pour, ensuite, faire la même chose dans $\mathrm{M}_p(K)$ voire dans tout anneau † : si l'on disposait d'un morphisme de groupes f surjectif de $(\mathrm{M}_p(K), +)$ dans $(\mathrm{GL}_n(K), \times)$ dont on connaîtrait le noyau, il suffirait de savoir résoudre l'équation $dM = E$, avec $E \in \ker(f)$ et $M \in \mathrm{M}_p(K)$ l'inconnue, pour en déduire les solutions de l'équation $N^d = \mathrm{I}_p$ d'inconnue $N \in \mathrm{GL}_p(K)$: il suffirait en effet de prendre l'image par f des solutions de $dM = E$. Comme un

*. Le terme adéquat, qui sera défini l'an prochain, est qu'un isomorphisme préserve les *ordres* (finis) des éléments d'un groupe.

†. Vous trouverez, dans mes feuilles d'exercices de l'an prochain, une démonstration que le groupe $(\mathbb{Z}/p^n\mathbb{Z})^\times$ est cyclique lorsque p est un nombre premier impair, qui fait usage d'une exponentielle « p-adique ». En effet, montrer que ce groupe est cyclique revient à démontrer l'existence d'une racine primitive d^{e} de l'unité où $d = \text{card}\left((\mathbb{Z}/p^n\mathbb{Z})^\times\right)$, et nous venons d'expliquer pourquoi l'exponentielle est une voie naturelle pour y parvenir.

morphisme qui convient est l'exponentielle dans le cas réel ou complexe, il est naturel d'espérer que l'exponentielle matricielle soit le morphisme recherché.

À la lumière de cette discussion, on comprend que si l'on souhaite définir l'exponentielle matricielle afin de résoudre des équations différentielles linéaires ou extraire des racines d^{es} dans $\mathrm{M}_n(K)$ (et encore ! je rappelle que je n'ai choisi que deux des applications possibles de l'exponentielle matricielle ; des motivations parfaitement complémentaires de la seconde auraient été possibles, notamment en géométrie), nous devons étudier :

— la dérivation de l'application $t \mapsto e^{A(t)}$ où A est une application dérivable à valeurs matricielles : obtient-on $t \mapsto A'(t)e^{A(t)}$? peut-on s'en servir pour résoudre des équations différentielles ?

— la propriété de morphisme de l'exponentielle matricielle : a-t-on toujours l'identité $\exp(A+B) = \exp(A)\exp(B)$? que peut-on en tirer comme conséquences ?

Nous verrons qu'il apparaît des complications du fait que $\mathrm{M}_p(K)$ ne soit pas un anneau commutatif. Les exercices suivants l'illustrent :

Exercice 80. On pose : $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

1. Montrer : $\forall t \in \mathbb{R}$, $\mathrm{ch}(t) = \sum_{n=0}^{+\infty} \frac{t^{2n}}{(2n)!}$, $\mathrm{sh}(t) = \sum_{n=0}^{+\infty} \frac{t^{2n+1}}{(2n+1)!}$.
2. ☆ En déduire : $\forall t \in \mathbb{R}$, $\exp(tA) = \begin{pmatrix} \mathrm{ch}(t) & \mathrm{sh}(t) \\ \mathrm{sh}(t) & \mathrm{ch}(t) \end{pmatrix}$. Comparer $t \mapsto \exp(tA)$ et $t \mapsto A\exp(tA)$. *On ne se souciera pas, pour le moment, des questions de convergence.*
3. Soit $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. A-t-on l'égalité : $\exp\left(B + B^\top\right) = \exp(B)\exp\left(B^\top\right)$?

Exercice 81. On pose : $\forall t \in \mathbb{R}$, $N(t) = \begin{pmatrix} 0 & t & t^2 \\ 0 & 0 & t^3 \\ 0 & 0 & 0 \end{pmatrix}$. Comparer l'application $t \mapsto N'(t)\exp(N(t))$ et la dérivée de l'application $t \mapsto \exp(N(t))$ (dériver les matrices composante par composante). Que conclure ?

Tout n'est donc pas aussi simple que dans $\mathbb{R}$ ou $\mathbb{C}$. Pas grave : nous résoudrons ces complications et apporterons des réponses satisfaisantes, bien que partielles, aux problèmes énoncés ci-dessus.

En plus de cela, le fil rouge de l'année de MP en algèbre linéaire s'appliquera également à l'exponentielle d'un endomorphisme ou d'une matrice, afin de simplifier son étude et son calcul : nous étudierons sa *réduction* (voir la section 6). ← page 17

12 Équations différentielles linéaires

12.1 La théorie

La théorie des équations différentielles linéaires est intégralement traitée en 1^{re} année dans les deux cas particuliers suivants :

— les équations différentielles linéaires du 1^{re} ordre ;

— les équations différentielles linéaires du 2^{e} ordre, à coefficients constants, avec un second membre « sympathique » (exponentielle, fonction trigonométrique, application polynomiale).

Nous irons plus loin en seconde année en donnant la structure algébrique générale de l'ensemble des solutions d'une équation différentielle linéaire, peu importe son ordre, ou d'un système différentiel linéaire. En particulier, nous verrons qu'un problème de Cauchy de la forme :

$$\left\{ \begin{array}{rcl} y^{(p)} + a_{p-1}y^{(p-1)} + \cdots + a_1 y' + a_0 y & = & b, \\ \forall i \in [\![0, p-1]\!],\ y^{(i)}(t_0) & = & y_i \end{array} \right.$$

où les $a_i : I \to K$ et $b : I \to K$ sont des applications continues, t_0 un réel fixé de l'intervalle d'étude I, et les y_i des scalaires fixés, admet toujours une *unique* solution.

Pour cette équation différentielle linéaire, comme pour toutes les autres (ainsi que pour les systèmes différentiels faisant intervenir plusieurs fonctions), l'idée sera de se ramener à une équation équivalente : $Y' = AY + B$, où A et B sont des fonctions vectorielles à valeurs dans $\mathrm{M}_p(K)$ et $\mathrm{M}_{p,1}(K)$ (pour un certain entier naturel non nul p qui dépend de l'ordre des dérivées, ainsi que du nombre de fonctions inconnues, apparaissant dans l'équation ou le système), et Y une fonction vectorielle à déterminer. Ce n'est pas très difficile, comme en atteste l'exercice suivant :

Exercice 82.

1. Soient $f, g, h : \mathbb{R} \to \mathbb{R}$ trois applications de classe C^1 sur $\mathbb{R}$. Définir des applications continues $A : \mathbb{R} \to \mathrm{M}_3(\mathbb{R})$ et $B : \mathbb{R} \to \mathrm{M}_{3,1}(\mathbb{R})$, et une application dérivable $Y : \mathbb{R} \to \mathrm{M}_{3,1}(\mathbb{R})$ (où A et B ne dépendent pas de f, g et h), telles que :
$$\forall t \in \mathbb{R}, \begin{cases} f'(t) &= -2f(t) + g(t) + h(t) + 1 \\ g'(t) &= f(t) - 2g(t) + h(t) + t \\ h'(t) &= f(t) + g(t) - 2h(t) + t^2 \end{cases} \iff Y' = AY + B.$$
2. Soit $y : \mathbb{R} \to \mathbb{R}$ une application de classe C^3 sur $\mathbb{R}$. Définir une application continue $A : \mathbb{R} \to \mathrm{M}_3(\mathbb{R})$ et une application dérivable $Y : \mathbb{R} \to \mathrm{M}_{3,1}(\mathbb{R})$ (où A ne dépend pas de y), telles que : $y''' + y'' - 10y' + 8y = 0 \iff Y' = AY$.
3. Soient $f : \mathbb{R} \to \mathbb{R}$ et $g : \mathbb{R} \to \mathbb{R}$ deux applications de classe C^2 sur $\mathbb{R}$. Définir des applications continues $A : \mathbb{R} \to \mathrm{M}_3(\mathbb{R})$ et $B : \mathbb{R} \to \mathrm{M}_{3,1}(\mathbb{R})$, et une application dérivable $Y : \mathbb{R} \to \mathrm{M}_{3,1}(\mathbb{R})$ (où A et B ne dépendent pas de f ni de g), telles que :
$$\forall t \in \mathbb{R}, \begin{cases} f''(t) &= f(t) + g(t) + \exp(t) \\ g'(t) &= f(t) + f'(t) + g(t) + \exp(-t) \end{cases} \iff Y' = AY + B.$$

Cette réduction illustre un phénomène général : quitte à augmenter la dimension, on peut ramener toute équation linéaire à une équation d'ordre 1. C'est également ce qui est fait dans l'exercice 39 ← page 24 avec des suites, dans la dernière question. On peut imiter l'approche pour résoudre entièrement certains systèmes différentiels :

Exercice 83. On reprend les notations de l'exercice 39. Soient f, g et h trois applications dérivables sur $\mathbb{R}$. On pose : $Y = \begin{pmatrix} f \\ g \\ h \end{pmatrix}$, et : $Z = P^{-1}Y$.

1. Montrer : $Y' = AY \iff Z' = \begin{pmatrix} 5 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} Z$, et en déduire une expression explicite des composantes de Z.
2. En déduire les solutions du système différentiel :
$$\begin{cases} f' &= f + 2g + 2h, \\ g' &= 2f + g + 2h, \\ h' &= 2f + 2g + h. \end{cases}$$
3. Adapter la méthode des deux questions précédentes au système différentiel :
$$\forall t \in \mathbb{R}, \quad \begin{cases} f'(t) &= f(t) + 2g(t) + 2h(t) + e^t, \\ g'(t) &= 2f(t) + g(t) + 2h(t) - e^{-t}, \\ h'(t) &= 2f(t) + 2g(t) + h(t) + 1. \end{cases}$$

4. En s'inspirant de ce qui précède, donner un énoncé général sur la forme des solutions d'une équation de la forme : $Y' = AY + B$ d'inconnue $Y : I \to \mathrm{M}_{p,1}(K)$ dérivable, lorsque $A \in \mathrm{M}_p(K)$ est semblable à une matrice diagonale D (notez bien que A n'est pas une application $I \to \mathrm{M}_p(K)$, ou du moins elle est constante) et $B : I \to \mathrm{M}_{p,1}(K)$ est une application continue.

La dernière question de cet exercice donne l'ensemble des solutions dans un cas très particulier (A à coefficients constants et semblable à une matrice diagonale). La discussion de la section 6.1 et ← page 18 ses exercices permettraient d'en faire de même pour la plupart des équations différentielles linéaires à coefficients constants. Mais cela ne donne pas encore le cas général. Nous l'obtiendrons grâce à un bel argument topologique (qui est, pour votre serviteur, une des plus belles applications de la topologie à un théorème hors de ce domaine), en remarquant que l'équation $Y' = AY + B$ munie d'une condition initiale $Y(t_0) = Y_0$ équivaut à l'équation :

$$\forall t \in I, \quad Y(t) = Y_0 + \int_{t_0}^{t} (A(x)Y(x) + B(x))\mathrm{d}x.$$

L'intérêt d'intégrer est de rajouter de la régularité et de faciliter les raisonnements impliquant des inégalités, puisque la dérivation est rarement compatible avec icelles.

Autrement dit, montrer l'existence et l'unicité d'une solution à l'équation différentielle équivaut à l'existence et l'unicité d'un point fixe de l'application $F : Y \mapsto \left(t \mapsto Y_0 + \int_{t_0}^{t} (A(x)Y(x) + B(x))\mathrm{d}x\right)$. Nous l'obtiendrons en montrant un analogue du théorème du point fixe de Banach-Picard (déjà montré en exemple dans le chapitre introductif), où il faut néanmoins s'adapter parce que F est une application de $\mathrm{C}^0(I, K)$ dans lui-même, et elle n'est pas nécessairement contractante. Ces complications techniques étant résolues, nous aurons démontré *le théorème de Cauchy linéaire* (existence et unicité d'une solution de $Y' = AY + B$ vérifiant une condition initiale prescrite), qui est le théorème majeur de ce chapitre.

Ce résultat est équivalent à la donnée d'une certaine bijection, qui est un isomorphisme lorsqu'il n'y a pas de second membre (c'est-à-dire lorsque B est identiquement nulle). Notons S l'espace vectoriel des solutions $Y : I \to \mathrm{M}_{p,1}(K)$ de $Y' = AY$, et fixons $t_0 \in I$. On a alors l'isomorphisme suivant, équivalent au théorème de Cauchy linéaire :

$$\Phi_{t_0} : \begin{cases} S & \to & K^p \\ Y & \mapsto & Y(t_0) \end{cases}.$$

C'est bien plus riche d'implications, parce qu'on a *rajouté de la structure*. Surtout : cela donne un isomorphisme entre un espace mal connu (que l'on cherche à expliciter, justement) et l'espace vectoriel par excellence K^p dans lequel on sait tout faire (notamment trouver des bases). Or un isomorphisme conserve tout ce qui est relatif à la structure : ainsi, au lieu de démontrer certains résultats concernant des fonctions dans S, nous aurons simplement à les démontrer dans K^n. On l'illustre dans l'exercice 85.

Exercice 84. Vérifier l'affirmation selon laquelle l'existence et l'unicité d'une solution $Y : I \to \mathrm{M}_{p,1}(K)$ dérivable de l'équation $Y' = AY + B$, avec la condition initiale $Y(t_0) = Y_0$ (ici $t_0 \in I$ et $Y_0 \in \mathrm{M}_{p,1}(K)$ sont fixés), équivaut à l'isomorphisme ci-dessus.

Exercice 85. Utiliser l'isomorphisme ci-dessus pour répondre aux questions suivantes.

1. Montrer : $\dim(S) = p$.
2. Soient $Y_1, \ldots, Y_p$ des solutions de : $Y' = AY$. Montrer que $(Y_1, \ldots, Y_p)$ est une base de S si et seulement s'il existe $t_0 \in I$ tel que la famille $(Y_1(t_0), \ldots, Y_p(t_0))$ soit libre, si et seulement si la famille $(Y_1(t), \ldots, Y_p(t))$ est libre pour tout $t \in I$.
3. Soit $\mathcal{B}$ une base de $\mathrm{M}_{p,1}(K)$. Montrer que l'application $W : t \mapsto \det_{\mathcal{B}}\left((Y_1(t), \ldots, Y_p(t))\right)$, appelée *wronskien* de $Y_1, \ldots, Y_p$, ne s'annule pas sur I ou est identiquement nulle.

Chaque question de cet exercice sera traitée en cours et aura des conséquences heureuses : la donnée de la dimension nous assure qu'il suffit de trouver p solutions indépendantes de $Y' = AY$, peu importe par quel moyen, pour en déduire une base de l'espace vectoriel des solutions, et la seconde question nous dit qu'il suffit de vérifier la liberté des évaluations en un seul réel pour avoir une base de l'espace de solutions ! C'est utile aussi bien en théorie qu'en pratique.

La troisième question semble dire la même chose que la seconde. Pas tout à fait : l'application W permet d'utiliser des techniques d'analyse pour avoir plus d'informations sur le comportement d'une famille $(Y_1, \dots, Y_p)$ *quand on fait varier* t, au contraire de l'isomorphisme Φ_{t_0} : c'est en effet une application continue (et même mieux) et à valeurs dans K, dont l'étude en tant que fonction de la variable réelle, donc une étude *analytique*, permet d'avoir des informations *algébriques* (annulation ou non d'un déterminant) !

12.2 La pratique

Les exercices 32 et 83 donnent déjà des exemples de résolutions pratiques, que nous rendrons plus méthodiques. Pour une description exhaustive et élégante des solutions, plus facile à manipuler notamment dans les exercices théoriques, nous utiliserons l'exponentielle matricielle mentionnée dans la section 11.2, et montrerons que les solutions de $Y' = AY + B$ s'écrivent exactement comme les solutions dans le cas réel $y' = ay + b$, *dans le cas où A est à coefficients constants.* Comme le calcul d'une exponentielle de matrice est élémentaire dans le cas diagonal, nous étudierons aussi plus spécifiquement le cas d'une matrice A semblable à une matrice diagonale.

← pages 21 et 53
← page 50

Reste le cas des coefficients non constants, bien plus délicat, et que nous étudierons en général dans le cas particulier des équations différentielles scalaires d'ordre 2 : $y'' + ay' + by = c$, où a, b et c sont des applications continues. Nous le verrons, *le plus dur est de trouver* UNE *solution non triviale.* Pour y parvenir, la méthode privilégiée est la recherche de solutions sous la forme d'une fonction développable en série entière, comme on l'a illustré dans la section 9.1 et plus spécifiquement dans l'exercice 63. Une fois qu'on a trouvé une solution non triviale f, on sait en produire une seconde linéairement indépendante g à l'aide d'une méthode qui n'est pas sans rappeler la méthode de variation de la constante, et (f, g) est alors une famille libre de l'espace vectoriel des solutions de l'équation différentielle, qui est de dimension 2 comme on l'a dit plus haut ; c'est donc une base. L'exercice ci-dessous donne une idée de l'approche :

← page 37
← page 41

Exercice 86. Considérons a et b deux applications continues sur un intervalle I et à valeurs complexes. Soit f une application deux fois dérivable sur I, solution de l'équation différentielle $(E) : y''+ay'+by = 0$. Soit z une application deux fois dérivable sur I. On pose : $g = f \cdot z$.

1. Montrer que g est solution de (E) si et seulement si z' est solution d'une équation différentielle du premier ordre dont les coefficients dépendent de a et b
2. En déduire l'expression d'une solution g de (E) qui ne soit pas proportionnelle à f.

Cependant dans certains cas nous devrons nous contenter d'informations qualitatives sur les solutions, à défaut de pouvoir les expliciter.

12.3 Résumé : objectif du chapitre

Dans un premier temps, nous expliquerons comment toute équation et tout système différentiels se ramènent à une équation de la forme : $Y' = AY + B$, comme exposé ci-dessus ; nous serons plus précis sur le sens de « s'y ramener », afin de s'assurer que les solutions de $Y' = AY + B$ donnent *exactement* les solutions de l'équation différentielle de départ. Nous démontrerons alors le théorème de Cauchy linéaire, assurant l'existence et l'unicité d'une solution de cette équation vérifiant une condition initiale prescrite. Cela permet d'en déduire un isomorphisme riche en implications : il permettra notamment d'avoir la structure de l'ensemble des solutions d'une équation différentielle (à commencer par sa dimension) et des conditions nécessaires et suffisantes simples pour qu'une famille de solutions soit une famille génératrice (dans le cas où c'est un espace vectoriel).

Ces résultats théoriques ont pour implication pratique que pour trouver toutes les solutions d'une équation différentielle, il suffit de trouver autant de solutions linéairement indépendantes que l'ordre de l'équation. La théorie de la réduction (section 6) et l'exponentielle matricielle (section 11.2) permettront d'expliciter les solutions dans le cas simple de coefficients constants. Lorsque les coefficients

← page 17
← page 50

ne sont pas constants, on s'adapte : on trouve une ou plusieurs solutions à la main, le plus méthodique étant de les chercher sous forme de fonction développable en série entière ; une fois qu'on a une solution, trouver les suivantes est moins difficile. Pour trouver des solutions particulières dans le cas d'un second membre, on introduira la méthode de variation *des* constantes.

Nous donnerons enfin, en exercice, des situations où les solutions ne sont pas explicites, et nous verrons comment les outils introduits (les isomorphismes et le wronskien) permettent malgré tout d'avoir le comportement asymptotique des solutions (limite en $+\infty$, caractère borné, etc.).

13 Espaces préhilbertiens réels et euclidiens

Le programme de 1re année définit le produit scalaire dans des espaces vectoriels abstraits, afin de les « géométriser », et que l'intuition héritée de la géométrie en dimension 2 ou 3 permette de résoudre des problèmes dans des espaces qui n'ont *a priori* rien à voir avec la géométrie au sens classique. Mais comme souvent en mathématiques, la théorie reste assez pauvre tant qu'on n'introduit pas des *applications* interagissant avec la structure géométrique qu'on a définie. Nous introduirons à cet effet :

— la notion d'endomorphisme adjoint ;

— les endomorphismes symétriques ou autoadjoints (c'est synonyme), et d'autres de la même veine en exercices ;

— les isométries.

13.1 Les isométries

Vous avez déjà entendu parler des isométries dans vos jeunes années de géométrie : ce sont les applications linéaires (quoique vous ayez vu quelques exemples qui ne le sont pas : les translations) qui conservent les distances et les angles non orientés. Avec le vocabulaire de géométrie de MP, une isométrie vectorielle f est un endomorphisme d'un espace euclidien $(E, \langle\cdot,\cdot\rangle)$ vérifiant :

$$\forall(\vec{x},\vec{y}) \in E^2, \quad \langle f(\vec{x}), f(\vec{y})\rangle = \langle \vec{x}, \vec{y}\rangle \,, \tag{$*$}$$

ou de manière équivalente : $\forall \vec{x} \in E$, $\|f(\vec{x})\| = \|\vec{x}\|$, avec $\|\cdot\|$ la norme euclidienne associée au produit scalaire $\langle\cdot,\cdot\rangle$. L'ensemble des isométries vectorielles de E est noté $\mathrm{O}(E)$.

Exercice 87. Vérifier l'équivalence des deux définitions. Pour cela, se souvenir des formules permettant d'exprimer une norme en fonction du produit scalaire associé, et inversement.

Exercice 88. Comprendre pourquoi l'égalité $(*)$ coïncide avec la définition « antique » d'une isométrie (application qui conserve les distances et les angles non orientés). Pour cela, vous devrez notamment vous demander comment $\langle \vec{x}, \vec{y}\rangle$ s'exprime en fonction de l'angle entre $\vec{x}$ et $\vec{y}$.

Exercice 89. Lorsque vous définissiez les isométries dans vos jeunes années, la précision « linéaire » n'apparaissait pas. Pour comprendre pourquoi :

1. Montrer que si f est une application de E dans E vérifiant $(*)$ et $f(\vec{0}) = \vec{0}$, alors f est linéaire (*indication : calculer* $\|f(\vec{x}+\lambda\vec{y}) - (f(\vec{x}) + \lambda f(\vec{y}))\|^2$ *avec un usage adéquat de* $(*)$).
2. Montrer que si f est une application de E dans E vérifiant $(*)$, alors f est affine (c'est-à-dire la composition d'une translation et d'une application linéaire).

Ainsi on ne « perd » pas d'isométries en omettant la précision « linéaire ».

Des exemples d'isométries vectorielles dans le plan ont déjà été donnés depuis le début de votre pratique de la géométrie : les symétries axiales et les rotations.

Exercice 90. Soient E un espace euclidien et F un sous-espace vectoriel de E.

1. Montrer que la symétrie par rapport à F et parallèlement à $F^\perp$ est une isométrie vectorielle (*un dessin permet de se convaincre que la conservation des normes est évidente ; l'utiliser comme support à une démonstration rigoureuse*).

2. Montrer que la projection par rapport à F et parallèlement à $F^\perp$ n'est pas une isométrie vectorielle, sauf si $F = E$ (*là encore, un support visuel devrait aider votre intuition géométrique*).
3. Montrer que si G est un sous-espace vectoriel supplémentaire de F, mais différent de $F^\perp$ (pour que ce soit possible, cela implique en particulier que F n'est pas $\{\vec{0}\}$ ni E), alors la symétrie par rapport à F et parallèlement à G n'est jamais une isométrie vectorielle. Cette question est plus subtile que les précédentes.

L'objectif de l'étude des isométries en MP est de classifier les isométries en dimension 2 et 3 (sachant que la méthode employée permet en principe de faire de même en toute dimension), c'est-à-dire : toutes les décrire, et donner des critères simples pour les distinguer et les reconnaître (étant donnée une isométrie vectorielle f dont on donne une description analytique ou matricielle, comment savoir de quelle isométrie il s'agit ?).

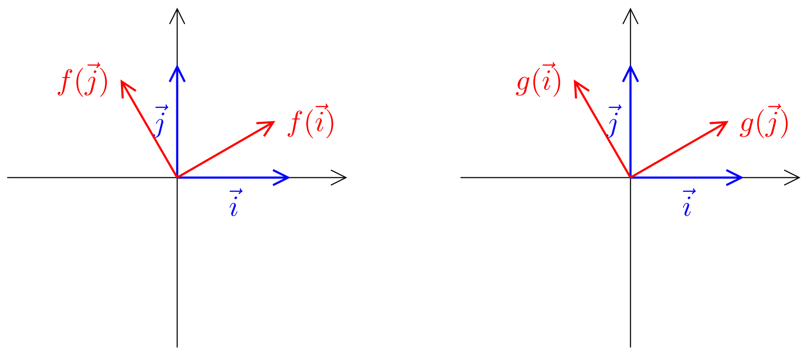
En dimension 2, nous verrons que les isométries connues depuis toujours (les rotations et les symétries axiales) sont les seules. En dimension 3, il y en a d'autres, mais elles restent obtenues à l'aide de compositions de rotations et de symétries d'un certain type (dites *orthogonales* : ce sont les symétries dont les caractéristiques géométriques sont orthogonales, dont il est question dans l'exercice 90). Pour distinguer les isométries entre elles, et en particulier trancher s'il s'agit d'une rotation ou d'une symétrie orthogonale (ou d'autre chose), on verra qu'il y a deux axes possibles :

— regarder la dimension de l'espace des points fixes : $\{\vec{x} \in E \mid f(\vec{x}) = \vec{x}\}$;

— regarder si l'*orientation* est préservée ou renversée.

L'orientation est une notion subtile à définir : nous ne le ferons pas ici, en nous basant seulement sur la compréhension intuitive que l'on vous a donnée jusqu'à présent. On dit qu'une isométrie préserve l'orientation si et seulement si elle ne change pas l'orientation d'une base (l'image d'une base orientée dans le sens direct, c'est-à-dire dans le sens trigonométrique, reste une base orientée dans le même sens). La figure de l'exercice suivant vous permet de regarder qu'effectivement, la simple question de l'orientation permet de trancher entre une rotation et une symétrie axiale, du moins dans le plan :

Exercice 91. Soient f et g deux isométries de $\mathbb{R}^2$. La figure suivante donne l'image par f et par g des vecteurs de la base canonique :



Donner la nature de f et g : rotation ou symétrie axiale ? Si c'est une rotation : dire comment obtenir son angle. Si c'est une symétrie axiale : représenter son axe, ou dire comment l'obtenir.

Mais en algèbre linéaire, il est souvent commode de caractériser matriciellement les propriétés. Cela fera partie de l'objet de la classification : reconnaître une isométrie à sa matrice représentative dans une base adéquate. Or les analogies matricielles de résultats géométriques sont possibles dès lors qu'on sait les traduire en termes de base, et il s'avère qu'il est possible de caractériser les isométries vectorielles à l'aide de leur action sur une base convenable : du fait qu'elles préservent les distances et les angles, elles préservent aussi les vecteurs unitaires et les relations d'orthogonalité.

Exercice 92. Soient E un espace euclidien et f un endomorphisme de E.

1. Montrer que si f est une isométrie et $\mathcal{B}$ une base *orthonormée* de E, alors $f(\mathcal{B})$ est une base orthonormée de E.

2. Étudier la réciproque : s'il existe une base orthonormée $\mathcal{B}$ de E telle que $f(\mathcal{B})$ soit une base orthonormée de E, peut-on conclure que f est une isométrie ?

Ainsi, suivant le principe selon lequel une propriété formulée en termes de bases a son pendant matriciel, nous obtiendrons notamment le résultat suivant :

Exercice 93. Soit f une isométrie de $\mathbb{R}^2$. Soit $\mathcal{B}$ une base orthonormée de $\mathbb{R}^2$ (prenez la base canonique si vous le souhaitez : c'est presque sans importance).

1. Montrer qu'il existe $(x, y) \in \mathbb{R}^2$ tel que : $x^2 + y^2 = 1$, et tel que la matrice représentative de f dans la base $\mathcal{B}$ soit :
$$\begin{pmatrix} x & y \\ y & -x \end{pmatrix}, \quad \text{ou :} \quad \begin{pmatrix} x & y \\ -y & x \end{pmatrix}.$$
Vous verrez l'an prochain que pour savoir laquelle des deux matrices donne la matrice de f, cela dépend de l'orientation *de $f(\mathcal{B})$.*
2. En déduire qu'il existe $\theta \in \mathbb{R}$ tel que la matrice représentative de f dans la base $\mathcal{B}$ soit :
$$\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}, \quad \text{ou :} \quad \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}.$$
Reconnaître, dans chaque cas, de quelle isométrie bien connue depuis votre tendre enfance il s'agit. Ne pas hésiter à *dessiner* l'image par f des vecteurs de la base $\mathcal{B}$ dans chaque cas, pour y voir plus clair.

Plus généralement, l'étude des matrices d'isométries nous conduira à l'étude des matrices dites *orthogonales*, dont nous donnerons les propriétés.

Exercice 94. Revoir, dans votre cours de 1$^{\text{re}}$ année, les propriétés utilisant le principe nommé ci-dessus, selon lequel les propriétés formulées en termes de bases se lisent matriciellement ; et ce, afin de s'en imprégner.

Enfin, en exercice, nous verrons que l'ensemble $\mathrm{O}(E)$ des isométries vectorielles a pour avantage d'être un « petit » sous-ensemble de $\mathrm{GL}(E)$, aussi bien algébriquement (il peut être engendré par un nombre restreint d'endomorphismes) que topologiquement. De plus, on peut s'y ramener de toutes sortes de manières, ce qui le rend remarquable ; nous l'illustrerons et verrons en quoi c'est utile.

13.2 Les endomorphismes adjoints et autoadjoints

Vous avez peut-être été intrigués que, durant tout votre cours d'algèbre linéaire de 1$^{\text{re}}$ année, il n'ait jamais été question d'un lien entre l'application linéaire associée à une matrice (canoniquement ou non) et celle associée à sa transposée. Et pour cause : un tel lien existe en toute généralité (grâce à la notion d'espace dual, de base duale), mais il est assez peu pratique parce qu'il fait manipuler des applications linéaires dans des espaces vectoriels très différents ; c'est d'autant plus embêtant pour traduire en termes d'applications des égalités où apparaîtraient les deux matrices en même temps.

Néanmoins, lorsque E est un espace euclidien, la difficulté suggérée ci-dessus est résolue : à tout endomorphisme f de E, on peut associer un endomorphisme f^* de E (appelé endomorphisme adjoint de f) tel que, pour toute base *orthonormée* $\mathcal{B}$ de E, la matrice de f^* dans $\mathcal{B}$ soit la matrice transposée de f dans $\mathcal{B}$. Il peut être surprenant que la solution s'exprime joliment dans un cadre géométrique, alors qu'*a priori* la notion de matrice transposée ne nécessite pas de produit scalaire pour être définie.

Ainsi il devient possible de montrer avec souplesse des liens entre une matrice et sa transposée (par exemple, vous savez qu'une matrice et sa transposée ont même rang, mais y a-t-il un lien entre leurs images ? leurs noyaux ?), en passant par les endomorphismes ; l'intérêt des endomorphismes étant comme souvent qu'ils ne sont pas assujettis à une base donnée.

Pour avoir une idée du genre de lien que nous allons étudier : voir l'exercice ci-dessous.

Exercice 95. On munit $\mathrm{M}_{n,1}(\mathbb{R})$ du produit scalaire usuel $(X, Y) \mapsto X^\top Y$. Soit $A \in \mathrm{M}_n(\mathbb{R})$.

1. Montrer : $\ker(A) = \operatorname{im}\left(A^\top\right)^\perp$, et : $\operatorname{im}(A) = \ker\left(A^\top\right)^\perp$ (*commencer par une inclusion, et montrer l'égalité des dimensions*).
2. Retrouver ainsi le fait que A et $A^\top$ ont même rang.

Les résultats obtenus avec les endomorphismes adjoints sont d'autant plus intéressants lorsqu'ils sont associés à des matrices ayant une relation remarquable avec leurs transposées : c'est le cas en particulier des matrices symétriques et antisymétriques. Nous étudierons donc ces matrices (plus particulièrement les symétriques) grâce à ce cadre géométrique qui les rend très éclairantes. Un endomorphisme dont la matrice dans toute base orthonormée est symétrique, est lui-même appelé endomorphisme symétrique, ou autoadjoint (puisque cela correspondra à la situation où : $f = f^*$). Le résultat central de ce chapitre sera le *théorème spectral*, selon lequel il existe toujours une base orthonormée dans laquelle la matrice d'un endomorphisme *autoadjoint* est diagonale. Ce sont les meilleures bases qui soient : d'une part parce qu'en géométrie, tout est plus simple dans une base orthonormée (le produit scalaire s'exprime comme le produit scalaire usuel, de même la norme, etc.), et d'autre part parce qu'en algèbre matricielle tout est plus simple avec des matrices diagonales. Ce théorème sera appliqué en d'innombrables circonstances.

13.3 Résumé : objectif du chapitre

Après quelques rappels et compléments du programme de 1re année, nous définirons la notion d'endomorphisme adjoint, et verrons ce qu'elle entraîne. Parmi cela, il y a le lien avec la matrice transposée. Ainsi nous pourrons relier géométriquement toute matrice et sa transposée, ce qui permettra plus facilement de transférer les propriétés de l'une à l'autre.

Dans le cas particulier d'une matrice égale à sa transposée, les relations entre un endomorphisme et son adjoint seront d'autant plus instructives que nous pourrons en déduire un joli théorème de réduction, le théorème spectral, qui assure l'existence d'une base à la fois commode pour la géométrie (elle est orthonormée) et pour l'étude d'un endomorphisme associé à une matrice symétrique (dans cette base, la matrice de cet endomorphisme est diagonale).

Puis nous introduirons les isométries, donnerons quelques exemples remarquables (parmi lesquels les symétries orthogonales) et les représenterons matriciellement : la matrice d'une isométrie « se reconnaît à l'œil nu ». Cela donnera la notion de matrice *orthogonale*. En comprenant les isométries, nous comprendrons ces matrices, et réciproquement.

Nous étudierons plus spécifiquement les isométries du plan et de l'espace tridimensionnel, afin de montrer que les isométries étudiées dans vos jeunes années sont essentiellement les seules (rotations et symétries). En traduisant matriciellement cette classification, nous en déduirons des résultats de *réduction* des isométries (et donc des matrices orthogonales).

14 Calcul différentiel

Les chapitres *Topologie des espaces vectoriels normés* et *Dérivation et intégration des fonctions vectorielles* permettent d'étendre les résultats de 1re année sur les fonctions $f : \mathbb{R} \to \mathbb{R}$: ← page 25 ← page 49

— aux fonctions $f : E \to F$ *continues*, avec E et F deux espaces vectoriels normés ;

— aux fonctions $f : \mathbb{R} \to F$ *dérivables* avec F un espace vectoriel normé ;

et il ne reste plus qu'à étendre les résultats de dérivabilité aux fonctions $f : E \to F$ dérivables.

Il n'est pas difficile de gérer le cas où l'espace d'arrivée est de dimension élevée : quitte à raisonner composante par composante, on se ramène au cas de fonctions à valeurs réelles. En revanche c'est plus délicat lorsque c'est l'espace de départ qui est dimension strictement supérieure à celle de $\mathbb{R}$. La raison à cela est que, pour calculer la limite de $f : E \to F$ en $\vec{a} \in E$, il y a de nombreuses façons de tendre vers $\vec{a}$ si E est de dimension supérieure ou égale à 2 (par la gauche, la droite, en diagonale, par « au-dessus ou en-dessous », en décrivant une spirale autour de $\vec{a}$...), alors qu'en dimension 1 on ne peut tendre vers un point $a \in \mathbb{R}$ que « par la gauche ou par la droite ». Il y a donc bien plus de chemins à considérer et le comportement de f sur chacun d'entre eux peut différer.

Exercice 96. Soient f et g les applications définies sur $\mathbb{R}^2 \setminus (\{0\} \times \mathbb{R})$ et $\mathbb{R}^2 \setminus \{(0,0)\}$ respectivement, par $(x, y) \mapsto \dfrac{y}{x}$ et $(x, y) \mapsto \dfrac{xy}{x^2+y^2}$. Calculer les limites suivantes :

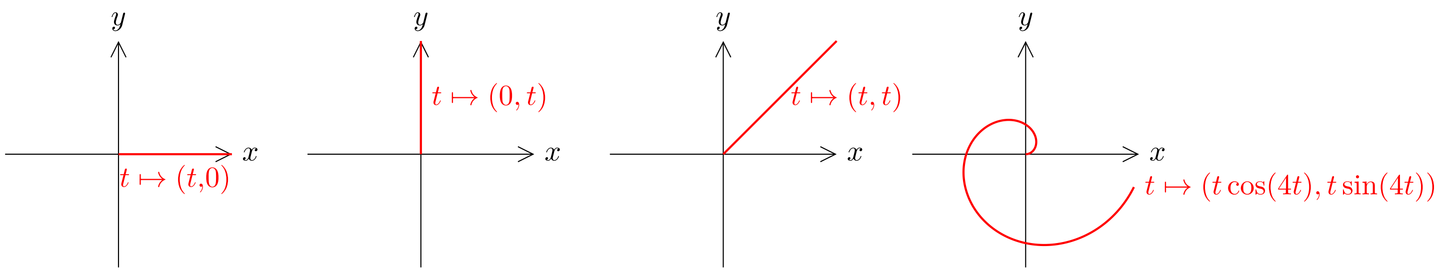
— $\lim\limits_{t\to 0^+} f(t,0)$ et $\lim\limits_{t\to 0^+} g(t,0)$ (on tend vers $(0,0)$ « horizontalement ») ;

— $\lim\limits_{t\to 0^+} g(0,t)$ (on tend vers $(0,0)$ « verticalement ») ;

— $\lim\limits_{t\to 0^+} f(t,t)$ et $\lim\limits_{t\to 0^+} g(t,t)$ (on tend vers $(0,0)$ « en suivant la droite d'équation $y = x$ ») ;

— $\lim\limits_{t\to 0^+} f(t\cos(4t), t\sin(4t))$ et $\lim\limits_{t\to 0^+} g(t\cos(4t), t\sin(4t))$ (on tend vers $(0,0)$ « en suivant une spirale ») ;

et les comparer : que remarque-t-on ? Si l'on devait donner un sens à $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ ou $\lim\limits_{(x,y)\to(0,0)} g(x,y)$, que proposerait-on ? Vous pouvez encore varier les choix des chemins (suivre une branche de parabole, d'hyperbole, d'autres droites...).



Néanmoins, malgré cette difficulté, les techniques du chapitre de topologie permettent sans trop de peine de définir et montrer l'existence de limite ou la continuité d'applications définies sur un espace vectoriel différent de $\mathbb{R}$: on explique dans les grandes lignes à la section 7 comment faire. C'est plus délicat pour la dérivabilité : bien que la notion de limite soit généralisée aux applications $f : E \to F$, c'est ici la définition du taux d'accroissement qui pose souci : on ne peut pas écrire $\dfrac{f(\vec{x}+\vec{h}) - f(\vec{x})}{\vec{h}}$ si $\vec{x}$ et $\vec{h}$ sont des vecteurs : c'est la division par $\vec{h}$ qui pose bien sûr problème.

Pour généraliser la notion de dérivée sans avoir ce problème, il suffit de se souvenir que la dérivée d'une fonction $f : \mathbb{R} \to \mathbb{R}$ est ce qui permet d'approcher son accroissement par une fonction *linéaire*. Plus précisément, f est dérivable en a si et seulement s'il existe $\ell \in \mathbb{R}$ tel que :

$$\underbrace{f(a+h) - f(a)}_{\text{accroissement}} = \underbrace{h\ell}_{\substack{\text{terme}\\\text{linéaire}}} + \underbrace{\underset{h\to 0}{o}(h)}_{\substack{\text{terme}\\\text{correctif}}},$$

et lorsqu'un tel ℓ existe on a alors : $f'(a) = \ell$. (Profitons-en pour signaler que c'est la façon la plus mûre de comprendre ce qu'est la *grande idée* derrière la dérivée et le calcul différentiel : approcher l'accroissement de f par une fonction linéaire, de sorte que les propriétés de f puissent découler de celles de la fonction linéaire qui est toujours plus simple à étudier ; à condition de maîtriser la taille du terme correctif. C'est ainsi qu'on parvient à déduire la monotonie de f autour de a à partir de celle de $h \mapsto hf'(a)$, qui dépend du signe de sa pente.)

Et là, il n'est pas difficile de proposer un analogue de la dérivée, que l'on appellera cependant autrement : si E et F sont deux espaces vectoriels munis d'une norme $\|\cdot\|$ (nous en parlons dans la section 7, et pour notre discussion on se contentera de se souvenir que c'est lorsqu'on a une norme sur ← page 25 un espace vectoriel qu'on peut faire de l'analyse), on dit que $f : E \to F$ est *différentiable* en $\vec{a} \in E$ s'il existe une application linéaire $L : E \to F$ telle que :

$$\underbrace{f(\vec{a}+\vec{h}) - f(\vec{a})}_{\text{accroissement}} = \underbrace{L(\vec{h})}_{\substack{\text{terme}\\\text{linéaire}}} + \underbrace{\underset{\vec{h}\to\vec{0}}{o}\left(\|\vec{h}\|\right)}_{\substack{\text{terme}\\\text{correctif}}}.$$

Le cas échéant, on définit la différentielle de f au point $\vec{a}$, dans la direction $\vec{h}$, comme étant : $\mathrm{d}f(\vec{a})\vec{h} = L(\vec{h})$. Avec ces notations, $\mathrm{d}f(\vec{a})$ est une application linéaire qui « approche » l'accroissement de f entre $\vec{a}$ et $\vec{a}+\vec{h}$. Cela semble très différent de l'approche de 1re année, où vous aviez plutôt étudié des fonctions de deux variables à travers leurs dérivées partielles. En vérité il y a un lien, que nous expliciterons dans le cours (et pas seulement pour deux variables) :

Exercice 97. Soient $\|\cdot\|$ la norme euclidienne usuelle de $\mathbb{R}^2$ et f l'application $(x, y) \mapsto xy$.

1. Pour tout $(x, y) \in \mathbb{R}^2$ et tout $(h, k) \in \mathbb{R}^2$, simplifier $f((x, y)+(h, k))-f(x, y)$, et en déduire que f est différentiable en tout point de $\mathbb{R}^2$, avec : $\forall (x, y) \in \mathbb{R}^2$, $\forall (h, k) \in \mathbb{R}^2$, $\mathrm{d}f(x, y)(h, k) = xk+yh$.
2. Calculer $\dfrac{\partial f}{\partial x}$ et $\dfrac{\partial f}{\partial y}$ en tout point, et écrire $\mathrm{d}f(x, y)$ en fonction de ces deux dérivées partielles.
3. À l'aide d'un résultat de votre cours de 1re année, montrer que l'égalité trouvée à la question précédente est vraie pour toute fonction de classe C^1.

Exercice 98. Soit $(n, p) \in (\mathbb{N} \setminus \{0\})^2$. Soit f une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$. Pour tout $\vec{a} \in \mathbb{R}^n$, expliciter l'application $\mathrm{d}f(\vec{a})$.

La différentielle a également un lien avec le gradient, que vous avez peut-être trouvé en traitant l'exercice ci-dessus, et que je ne développe dans cette présentation.

Après avoir donné plusieurs exemples d'applications différentiables (et l'expression de leurs différentielles) et les propriétés élémentaires de ces applications, notamment pour permettre leur calcul en pratique, l'objectif est de trouver un analogue en toute dimension des résultats connus pour les fonctions dérivables sur un intervalle de $\mathbb{R}$:

— la caractérisation des fonctions dérivables constantes : f constante $\Longleftrightarrow f' = 0$;
— la caractérisation des fonctions dérivables croissantes : f croissante $\Longleftrightarrow f' \geqslant 0$;
— l'inégalité des accroissements finis ($|f(b) - f(a)| \leqslant \|f'\|_\infty |b - a|$) ;
— le lien entre extremums et points critiques (si f admet un extremum local en un point a de l'intérieur de son intervalle de dérivabilité, alors $f'(a) = 0$).

Pour avoir de tels énoncés, il faut d'abord se demander ce qui doit jouer l'analogue des intervalles ouverts dans ces énoncés. On sait en effet que plusieurs de ces résultats deviennent faux si l'on n'a plus un intervalle, ou s'il n'est pas ouvert :

Exercice 99.

1. Donner un exemple de fonction dérivable en tout point de son domaine de définition, dont la dérivée est nulle en tout point, mais qui n'est pas constante.
2. Donner un exemple de fonction dérivable en tout point de son domaine de définition, dont la dérivée est de même signe en tout point, mais qui n'est pas monotone.
3. Donner un exemple de fonction dérivable, qui admet un extremum local en un point où la dérivée n'est pas nulle.

Comme on l'a dit dans la section 7.2, une généralisation des intervalles est possible, et elle est ← page 27
donnée par ce qu'on appelle des parties *connexes par arcs* ou des *convexes* (qui sont des cas particuliers plus simples de parties connexes par arcs). En remplaçant les intervalles ouverts par les parties ouvertes connexes par arcs (ou convexes pour faciliter certaines démonstrations), et les applications dérivables (resp. de classe C^1) par les applications différentiables (resp. de classe C^1), nous saurons généraliser les résultats ci-dessus aux fonctions de plusieurs variables. La stratégie de démonstration sera à chaque fois de se ramener au cas bien connu des fonctions d'une variable réelle, par composition avec un chemin bien choisi.

Une partie significative des applications tournera autour de l'optimisation : minimiser la valeur d'un produit scalaire, d'un déterminant, des coordonnées d'une solution d'une équation linéaire, etc., en cherchant les points où une différentielle est identiquement nulle.

L'absence de notion claire de monotonie, pour une fonction qui n'est pas définie sur un intervalle de $\mathbb{R}$ (du fait que pour aller d'un point à un autre, il peut y avoir beaucoup de chemins différents si l'on est en dimension strictement supérieure à 1), empêche de conclure aussi facilement sur la nature des points critiques. Pour une fonction $f : \mathbb{R} \to \mathbb{R}$ dérivable, un tableau de variation permet immédiatement de repérer les maximums et minimums. Pour une fonction $f : E \to \mathbb{R}$ avec $\dim(E) \geqslant 2$, une fois que nous aurons déterminé des points critiques de f grâce aux lieux où la différentielle est identiquement nulle, nous aurons parfois besoin d'une information supplémentaire pour déterminer s'il s'agit d'un extremum ou non ; et si oui, d'un maximum ou minimum. Pour comprendre comment faire, remarquons que pour les fonctions de la variable réelle, la dérivée *seconde* peut parfois nous y aider : considérons en effet une fonction $f : \mathbb{R} \to \mathbb{R}$ de la variable réelle, de classe C^2, ayant un extremum local en $a \in \mathbb{R}$. Comme $f'(a) = 0$, le développement limité de f à l'ordre 2 en a donne, d'après la formule de Taylor-Young :

$$f(x) - f(a) = \frac{f''(a)}{2}(x-a)^2 + \underset{x \to a}{o}\left((x-a)^2\right) \approx \frac{f''(a)}{2}(x-a)^2.$$

Comme $(x-a)^2 \geqslant 0$, le signe de $f(x) - f(a)$ (pour x proche de a) est donné par le signe de $f''(a)$: si $f''(a) > 0$, alors $f(x) - f(a) > 0$, donc $f(x) > f(a)$ pour tout x proche de a : on en déduit que a est un *minimum* local. Si $f''(a) < 0$, alors on conclut à l'inverse que $f(x) < f(a)$, si bien que a est un maximum local.

Exercice 100. Démontrer *rigoureusement* ce que l'on vient d'affirmer. Pour cela, il s'agit de remplacer l'égalité « $\approx$ » par une gestion adéquate du terme $\underset{x \to a}{o}\left((x-a)^2\right)$. Traduire en termes epsilonesques le fait d'être un petit o de $(x-a)^2$ quand $x \to a$, et faire un bon choix de ε pour conclure.

Nous avons un résultat analogue pour les fonctions en dimension supérieure, à condition de définir la matrice *hessienne* d'une fonction de classe C^2 qui est :

- intrinsèquement, à une petite imprécision près, la différentielle de la différentielle, c'est-à-dire la différentielle de l'application $\vec{a} \mapsto df(\vec{a})$ (par analogie avec la dérivée seconde qui est la dérivée de la dérivée) ;
- extrinsèquement, pour une fonction définie sur $\mathbb{R}^n$, une matrice dont le coefficient (i, j) est la dérivée partielle seconde par rapport aux i^{e} et j^{e} variables.

Cette matrice permet de généraliser la formule de Taylor-Young à l'ordre 2 et donc l'heuristique ci-dessus. Par exemple, pour une fonction f de classe C^2 sur $\mathbb{R}^2$, on aurait :

$$f(\vec{a}+\vec{h}) = f(\vec{a}) + h_1\frac{\partial f}{\partial x}(\vec{a}) + h_2\frac{\partial f}{\partial y}(\vec{a}) + h_1^2\frac{\partial^2 f}{\partial x^2}(\vec{a}) + 2h_1h_2\frac{\partial^2 f}{\partial x \partial y}(\vec{a}) + h_2^2\frac{\partial^2 f}{\partial y^2}(\vec{a}) + \underset{\vec{h} \to \vec{0}}{o}\left(\|\vec{h}\|^2\right).$$

Grâce à des conditions sur les dérivées partielles secondes (assez complexes : elles se reformulent plus agréablement en termes de trace et déterminant de la matrice hessienne), nous aurons un moyen de détecter les extremums locaux parmi les points critiques. Pour avoir une idée du type de conditions, voir l'exercice suivant (qui a été rajouté à la dernière minute pour avoir un nombre premier d'exercices) :

Exercice 101. On suppose : $\dfrac{\partial^2 f}{\partial x \partial y} > 0$.

1. Écrire la quantité $h_1^2\dfrac{\partial^2 f}{\partial x^2}(\vec{a}) + 2h_1h_2\dfrac{\partial^2 f}{\partial x \partial y}(\vec{a}) + h_2^2\dfrac{\partial^2 f}{\partial y^2}(\vec{a})$ sous la forme : $(\alpha h_1 + \beta h_2)^2 + \gamma h_2^2$, où α, β et γ s'expriment exclusivement à l'aide des dérivées partielles secondes de f.
2. En déduire une *conjecture* liant la nature d'un point critique $\vec{a}$ (maximum ou minimum) au déterminant de la matrice hessienne $\begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2}(\vec{a}) & \dfrac{\partial^2 f}{\partial x \partial y}(\vec{a}) \\ \dfrac{\partial^2 f}{\partial x \partial y}(\vec{a}) & \dfrac{\partial^2 f}{\partial y^2}(\vec{a}) \end{pmatrix}$.

Table des matières