Présentation des chapitres de mathématiques de MP
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 $f : \mathbb{R} \to \mathbb{R}$:

— 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 1. 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 :

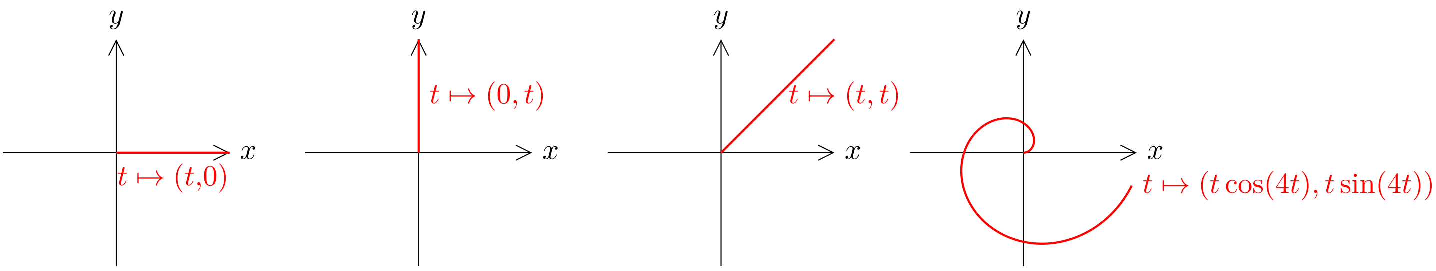
— $\lim\limits_{t\to 0^+} f(t,0)$ et $\lim\limits_{t\to 0^+} g(t,0)$ (on tend vers $(0,0)$ « horizontalement ») ;

— $\lim\limits_{t\to 0^+} g(0,t)$ (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 en topologie 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 en topologie, et pour notre discussion on se contentera de se souvenir que c'est lorsqu'on a une norme sur 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 2. 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 3. 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 $\iff f' = 0$;
- la caractérisation des fonctions dérivables croissantes : f croissante $\iff f' \geqslant 0$;
- 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 4.

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 en topologie, une généralisation des intervalles est possible, et elle est 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 5. 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 :

Exercice 6. 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 $\begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2}(\vec{a}) & \dfrac{\partial^2 f}{\partial x \partial y}(\vec{a}) \\ \dfrac{\partial^2 f}{\partial x \partial y}(\vec{a}) & \dfrac{\partial^2 f}{\partial y^2}(\vec{a}) \end{pmatrix}$.