Université Paris 9 - Dauphine

Calcul Différentiel et Optimisation

Un exo

Exercice. On considère

$$\forall A \in M_n(\mathbb{R}), f(A) = e^{tr(A)} A.$$

Montrer que f est Gateaux-différentiable, Fréchet-différentiable, et de classe $\mathcal{C}^1$.

Dérivabilité au sens de Gateaux :

Méthode 1 : on fixe A et H dans $M_n(\mathbb{R})$, et on pose $\varphi(t) = f(A + tH)$ pour $t \in \mathbb{R}$. Alors φ est dérivable et

$$\varphi'(t) = tr(H)e^{tr(A)+tr(H)t}(A + tH) + e^{tr(A)+tr(H)t}H.$$

$\varphi'(0) = e^{tr(A)}(tr(H)A + H)$ est le seul candidat possible pour être la différentielle de f en A dans la direction H.

Méthode 2 : on fixe A et H dans $M_n(\mathbb{R})$, et on calcule la limite

$$\frac{\varphi(t) - \varphi(0)}{t} = e^{tr(A)} \left(\frac{e^{tr(H)t} - 1}{t} A + He^{tr(H)t} \right) \longrightarrow_{t \to 0} e^{tr(A)}(tr(H)A + H)$$

On constate bien que, à A fixé, le candidat trouvé (par l'une des deux méthodes) est une application linéaire de H (et continue, ce qui est toujours le cas en dimension finie si on est linéaire). On a donc trouvé la différentielle $D_G f(A)(H) = e^{tr(A)}(tr(H)A + H)$ et montré la Gateaux-différentiabilité.

Dérivabilité au sens de Fréchet :

Pour montrer la Fréchet-différentiabilité, ce qui précède est insuffisant : on revient à la définition de la Fréchet-différentiabilité. Deux méthodes sont possibles : soit on a déjà montré la Gateaux-différentiabilité, et alors on connait le candidat pour la différentielle, soit on ne l'a pas fait. Bien sûr, si on veut montrer la Fréchet-différentiabilité, il semble plus rapide de ne pas traiter la Gateaux-différentiabilité, mais alors ce sera plus dur car on ne connait pas la différentielle. Voyez la différence dans les deux méthodes suivantes. La troisième méthode consiste à utiliser les règles de calcul du cours, et à décomposer la fonction f en produits et composées de fonction plus simples à différentier (comme vous le feriez si vous étiez en dimension 1).

Méthode 1 : on connaît la différentielle candidate (obtenue par la Gateaux-différentiabilité) : on fixe A dans $M_n(\mathbb{R})$; alors pour tout $H \in M_n(\mathbb{R})$,

$$\begin{aligned} f(A + H) - f(A) - e^{tr(A)}(tr(H)A + H) &= e^{tr(A)} \left[(e^{tr(H)} - 1 - tr(H))A + (e^{tr(H)} - 1)H \right] \\ &= e^{tr(A)} [o(tr(H))A + o(1)H] \\ &= o(\|H\|) \end{aligned}$$

Précisons la justification de la dernière ligne : l'avant dernière ligne signifie qu'il existe $\varepsilon_1, \varepsilon_2$ deux fonctions réelles telles que

$$f(A + H) - f(A) - e^{tr(A)}(tr(H)A + H) = e^{tr(A)} [\varepsilon_1(tr(H))A + \varepsilon_2(tr(H))H]$$

et

$$\forall \varepsilon > 0, \exists \delta > 0, \forall H \text{ telle que } |tr(H)| < \delta, |\varepsilon_1(tr(H))| \leq \varepsilon |tr(H)| \text{ et } |\varepsilon_2(tr(H))| \leq \varepsilon$$

Remarque 1 La bonne définition de $f(x) = o(g(x))$ au voisinage de 0 (par exemple) est bien $\forall \varepsilon > 0, \exists \delta > 0, \forall |x| < \delta, |f(x)| \leq \varepsilon |g(x)|$, qui évite de faire des quotients, ce qui peut poser problème lorsque g s'annule souvent, ce qui est le cas ici car il y a beaucoup de matrices de traces nulles.

Et en utilisant, via la continuité de la trace, que $|tr(H)| \leq C\|H\|$ pour une constante C indépendante de H (mais dépendante de la norme, que nous ne précisons pas), on montre que

$$\forall \varepsilon > 0, \exists \delta > 0, \forall H \text{ telle que } \|H\| < \delta / C, |\varepsilon_1(tr(H))| \leq C\varepsilon \|H\| \text{ et } |\varepsilon_2(tr(H))| \leq \varepsilon$$

Autrement dit, un "petit o de tr(H)" est un "petit o de $\|H\|$" (le contraire est faux).

Méthode 2 : on n'a pas regardé la Gateaux-différentiabilité, donc on ne connaît pas la différentielle. On doit linéariser $f(A + H)$ pour H petit :

$$\begin{aligned} f(A + H) &= e^{tr(A)} \left[e^{tr(H)}(A + H) \right] = e^{tr(A)} [(1 + tr(H) + o(tr(H))(A + H)] \\ &= f(A) + \underbrace{e^{tr(A)}(tr(H)A + H)}_{\text{linéaire en } H} + \underbrace{e^{tr(A)}[(A + H)o(tr(H)) + Htr(H)]}_{o(\|H\|)} \end{aligned}$$

où les justifications du o sont les mêmes que précédemment.

Méthode 3 : on va utiliser les règles de calcul des différentielles : on pose

$$\varphi_1(A) = tr(A), \quad \varphi_2(x) = e^x, \quad \varphi_3(A) = A$$

alors $f(A) = \varphi_2 \circ \varphi_1(A).\varphi_3(A)$, et les trois fonctions $\varphi_1, \varphi_2, \varphi_3$ sont très facilement (Fréchet-)dérivable, et donc f l'est d'après les théorèmes sur produit et composée. Si on a déjà calculé la dérivée directionnelle (première partie), on peut s'arrêter. Sinon on applique les règles de calcul de ces théorèmes) et

$$\varphi_1'(A)(H) = tr(H), \quad \varphi_2'(x)(h) = e^x h, \quad \varphi_3'(A)(H) = H.$$

(la fonction φ_2 est une fonction réelle, donc $\varphi_2'(x)(h)$ est la multiplication entre la dérivée comme vous la calculiez avant, et le réel h).
Les règles sur la dérivée d'une composée donnent :

$$\begin{aligned}[]
[\varphi_2 \circ \varphi_1]'(A)(H) &= [\varphi_2'(\varphi_1(A)) \circ \varphi_1'(A)](H) = \varphi_2'(\varphi_1(A))(\varphi_1'(A)(H)) = \varphi_2'(tr(A))(tr(H)) \\
&= e^{tr(A)} tr(H)
\end{aligned}$$

et celles sur la dérivée d'un produit donnent

$$\begin{aligned}
f'(A)(H) &= [\varphi_2 \circ \varphi_1]'(A)(H).\varphi_3(A) + [\varphi_2 \circ \varphi_1](A).\varphi_3'(A)(H) \\
&= e^{tr(A)} tr(H).A + e^{tr(A)}.H
\end{aligned}$$

En conclusion, on a montré que f est Fréchet-différentiable en tout $A \in M_n(\mathbb{R})$, et que $f'(A)(H) = e^{tr(A)}(tr(H)A + H)$

Remarque 2 Essayez de comprendre le cas de la dimension n=1, et faites le lien avec la dérivée habituelle de $f(x) = xe^x$.

Caractère $\mathcal{C}^1$:

Méthode 1 : on a déjà montré la Fréchet-différentiabilité : on doit montrer la continuité de l'application

$$\begin{array}{ccc}
M_n(\mathbb{R}) & \to & L(M_n(\mathbb{R}), M_n(\mathbb{R})) \\
A & \mapsto & \left(\begin{array}{ccc} M_n(\mathbb{R}) & \to & M_n(\mathbb{R}) \\ H & \mapsto & e^{tr(A)}(tr(H)A + H)) \end{array} \right)
\end{array}$$

On choisit donc sur $M_n(\mathbb{R})$ une norme, et cela fournit une norme subordonnée sur $\mathcal{L}(M_n(\mathbb{R}), M_n(\mathbb{R}))$ (norme d'applications linéaires). Et donc on peut écrire :

$$\begin{aligned}
\|f'(A)(H) - f'(B)(H)\|_{M_n(\mathbb{R})} &= \|(e^{tr(A)} - e^{tr(B)})H + (e^{tr(A)}A - e^{tr(B)}B)tr(H)\|_{M_n(\mathbb{R})} \\
&\leq |e^{tr(A)} - e^{tr(B)}| \|H\|_{M_n(\mathbb{R})} + C\|e^{tr(A)}A - e^{tr(B)}B\|_{M_n(\mathbb{R})} \|H\|_{M_n(\mathbb{R})}
\end{aligned}$$

où la constante C est telle que $|tr(H)| \leq C\|H\|_{M_n(\mathbb{R})}$. Donc par définition de la norme subordonnée, on a bien

$$\|f'(A) - f'(B)\|_{\mathcal{L}(M_n(\mathbb{R}), M_n(\mathbb{R}))} \leq |e^{tr(A)} - e^{tr(B)}| + C\|e^{tr(A)}A - e^{tr(B)}B\|_{M_n(\mathbb{R})}$$

et ce dernier terme tend bien vers 0 lorsque B tend vers A.

Méthode 2 : on a seulement montré la Gateaux-différentiabilité : on peut procéder exactement comme dans la méthode précédente, et utiliser le théorème du cours qui dit :

$$f : \Omega \subset E \to F \text{ Gateaux-différentiable en tout point de } \Omega \subset E, \text{ et } f'(= D_G f) \text{ continue sur } \Omega \Longrightarrow f \text{ de classe } \mathcal{C}^1$$

Conclusion : Les méthodes dépendent de la question posée, suivant qu'on vous demande de montrer la différentiablité Gateaux, la différentiablité Fréchet, ou le caractère $\mathcal{C}^1$. Il est à retenir qu'il est souvent utile d'étudier d'abord la dérivation directionnelle qui est plus simple, et pourra aider à prouver la Fréchet-différentiabilité. A retenir aussi, le dernier théorème cité ci-dessus, car il permet dans certain cas de courcircuiter la Fréchet-différentiabilité pour montrer le caractère $\mathcal{C}^1$.