REPERES - IREM. N° 34 - janvier 1999

A PROPOS DU THÉORÈME DES ACCROISSEMENTS FINIS

Henri LOMBARDI
IREM de Franche-Comté

Introduction

Sans doute avez-vous un cousin qui ne fait pas de mathématiques mais qui conduit une voiture. La prochaine fois que vous le voyez demandez lui à brûle pourpoint :

"Suppose que tu roules pendant une heure entre 30 et 40 km /h, combien penses-tu que tu auras fait de kilomètres ?" (A)

La réponse sera sans doute

"Eh, tu me prends pour un imbécile ?"

ou bien

"Je ne vois pas ce qu'il y a de drôle."

Dans le premier cas, faites-vous tout petit (il serait absurde de se fâcher avec son cousin pour si peu).

Dans le deuxième cas, vous pouvez enchaîner

"Ce qu'il y a de drôle, c'est qu'on le démontre en mathématiques, qu'on a fait entre 30 et 40 kilomètres".

Et vous aurez peut-être fait faire un grand progrès au développement de l'esprit mathématique dans toutes les couches de la société.

Si vous indiquez en outre qu'il s'agit d'un théorème assez difficile à démontrer, vous aurez sans doute créé un nouvel adepte des démonstrations rigoureuses. Si vous avez enseigné les mathématiques en terminales dans les années 80 vous avez eu la lourde tâche d'avoir à démontrer l'évidente réponse à la question (A) ci-dessus au moyen de l'évidence (B) suivante :

"Suppose que pendant une heure tu ne recules à aucun moment, et bien au bout d'une heure tu n'auras pas reculé" (B).

Bref, vous l'aviez compris, il s'agissait de faire quitter à vos élèves la planète terre pour les emmener dans la planète "mathématiques formelles".

Et si les élèves n'y comprenaient rien, on peut se demander si par hasard, la nécessité de démonstrations convaincantes ne devrait pas être illustrée sur des exemples pour lesquels le résultat n'est pas *a priori* évident.

En fait la "démonstration" du théorème des accroissements finis, démontre, non pas la réponse à la question (A), mais l'adéquation du modèle "nombres réels, fonctions dérivables" pour la réalité "voiture qui roule avec un compteur de vitesse". Cette adéquation n'est d'ailleurs que partielle et approximative (il est douteux que R modélise parfaitement l'écoulement du temps, ou l'espace sur une ligne).

En résumé, on ne devrait jamais démontrer le théorème des accroissements finis (à l'université par exemple) en tant que preuve d'une chose évidente, mais en tant que vérification de l'adéquation d'un modèle mathématique à la réalité qu'il veut représenter.

Première discussion

Plaçons nous au niveau du DEUG, et supposons que nous ayons déjà fait sentir suffisamment en quoi la structure $\mathbb{R}$ est un assez bon modèle pour certaines choses, et en quoi une fonction réelle uniformément continue sur un intervalle [a,b] est un assez bon modèle pour la description du mouvement d'un mobile sur une ligne.

Nous en arrivons au chapitre "dérivée comme modélisation de la notion de vitesse".

Nous parlons de la vitesse moyenne, et nous arrivons au point : qu'il serait agréable d'avoir une fonction "vitesse instantanée".

Si nous reprenons ce qu'on nous a enseigné comme étant la Vérité, nous définirons la vitesse instantanée comme limite de la vitesse moyenne au sens de la définition usuelle. Il s'agit d'une définition "locale en un point", nous dirons dans la suite que c'est une définition "ponctuelle"

$$f'(t_0) = \lim_{h \to 0} \frac{f(t_0 + h) - f(t_0)}{h} \qquad (1)$$

Mais si nous prenons un peu de recul, et si nous nous débarrassons de ce réflexe de vouloir toujours rechercher des hypothèses minimales (même quand, dans la pratique mathématique, les problèmes ne se présentent jamais avec ces seules hypothèses minimales), nous serons amenés à nous demander si la définition minimale est, après tout, bien réaliste. En classe de physique (seconde ou après), la vitesse à l'instant t_0 est mesurée expérimentalement comme étant :

$$(f(t_0 + \Delta t) - f(t_0 - \Delta t)) / 2\,\Delta t,$$

avec Δt petit. Ceci suggèrerait une autre définition minimale, de la "dérivée de la fonction f au point t_0" qui serait :

$$f'(t_0) = \lim_{h \to 0} \frac{f(t_0 + h) - f(t_0 - h)}{2h} \qquad (2)$$

Avec cette nouvelle définition, certaines fonctions qui étaient non dérivables en t_0, deviennent dérivables, comme par exemple la fonction $f(t) = |\,t\,|$ (qui devient dérivable en 0 avec $f'(0) = 0$).

Évidemment, on sent bien que cette

définition est peut-être un peu trop minimale (1). Mais pourquoi alors la définition usuellement donnée en classe de mathématiques serait-elle "juste assez minimale mais pas trop" ? Si on ne veut pas utiliser des arguments purement esthétiques (c'est plus élégant, ça donne de plus jolis théorèmes, si vous remettez tout en cause, on ne sait plus où on va, etc.) on est bien obligé de se fatiguer un peu et d'ouvrir une discussion qui, ordinairement, est renvoyée dans les "notes historiques", mais qui, du point de vue de la formation mathématique, devrait être partie intégrante du cours.

Le fait de fixer, arbitrairement, un accroissement égal de chaque côté, est déplaisant pour une raison de fond (et pas seulement esthétique). Si la deuxième définition est trop minimale, la première l'est aussi, bien qu'à un degré moindre. En effet la première définition est équivalente à la suivante.

$$f'(t_0) = \lim_{\substack{h,k \to 0,\\ h\geq 0,\, k\geq 0,\, h+k>0}} \frac{f(t_0+h)-f(t_0-k)}{h+k} \qquad (3)$$

La propriété (3) résulte de la propriété (1) parce que le taux d'accroissement considéré en (3) est la moyenne pondérée (avec les poids h et k) des taux d'accroissements entre t_0 et $t_0 + h$ d'une part, entre $t_0 - k$ et t_0 d'autre part.

Ainsi dans la définition usuelle, on peut considérer deux accroissements de part et d'autre de t_0 indépendants l'un de l'autre, mais on s'interdit de considérer deux accroissements distincts arbitraires, qui pourraient être éventuellement du même coté de t_0. En supprimant cette restriction, nous arrivons à la définition suivante (f étant supposée continue) :

la dérivée de f au point t_0 est la limite suivante, si elle existe

$$\lim_{h,k \to 0,\, h \neq k} \frac{f(t_0+h)-f(t_0+k)}{h-k} \qquad (4)$$

L'appellation officielle pour la propriété (4) est : "la fonction f est strictement dérivable en t_0". Dans un livre en tous points remarquable ([Dem]), Michel Demazure montre l'intérêt profond de cette définition, qui, généralisée en plusieurs variables, donne un cadre naturel pour le théorème d'inversion locale et par conséquent pour le théorème des fonctions implicites.

Une démarche plus systématique

Jusqu'à présent, nous avons raisonné, par jeu, en digression à partir de la définition "standard".

Mais nous avons en quelque sorte continué à jouer le même jeu, en ce sens que nous n'avons pas procédé systématiquement, ce qui nécessite de commencer par préciser les propriétés que nous voulons voir vérifiées par le modèle mathématique.

Personnellement, il me semble que **la** propriété que nous voulons voir vérifier à la fonction dérivée est la propriété (A). Celle justement que votre cousin, qui ne fait pas de mathématique, trouverait absolument aberrant qu'elle ne soit pas vérifiée.

En langage mathématique formalisé, nous obtenons la traduction (5) de la propriété (A)

(1) Avec cette définition le théorème de Rolle devient faux.

La fonction $g : [a, b] \to \mathbb{R}$, uniformément continue, sera dite *fonction dérivée de la fonction* $f : [a, b] \to \mathbb{R}$ si on a :

$$\left[\begin{array}{l}\forall x,y \in [a,b] \ \forall m, M \in \mathbb{R} \text{ avec } x > y \text{ et } m < M \\ g([x,y]) \subset [m,M] \Rightarrow \dfrac{f(y) - f(x)}{y - x} \in [m,M]\end{array}\right] \quad (5)$$

Insistons sur le fait que (5) est **exactement la traduction de** (A), avec l'*a priori* de modéliser au moyen de fonctions uniformément continues de $[a, b]$ vers $\mathbb{R}$.

Le réflexe minimaliste inciterait à ne pas demander la continuité uniforme, au moins pour la fonction g. Nous signalerons à ce sujet que, dans la mesure où on modélise au moyen de fonctions, il est à peu près inévitable de s'en tenir aux fonctions uniformément continues (pour un segment borné) ceci parce que :

– (a) la continuité est dans la nature de la modélisation (les mesures en physique sont toujours approchées, sauf lorsque la variable est dans un ensemble discret),
– (b) l'uniforme continuité est la forme manipulable de la continuité

Pour plus de détails voir l'encadré.

Notons que la définition (5) implique que en chaque point t_0 de $[a,b]$ la fonction f est strictement dérivable (définition (4)), avec $f'(t_0) = g(t_0)$: supposons en effet qu'on ait

$$|x - y| \le \delta \Rightarrow |g(y) - g(x)| \le \varepsilon$$

alors, pour $|h|$ et $|k| \le \delta$ et $h > k$ on aura $g([t_0+k, t_0+h]) \subset [g(t_0) - \varepsilon, g(t_0) + \varepsilon]$ et donc $\dfrac{f(t_0 + h) - f(t_0 + k)}{h - k} \in [g(t_0) - \varepsilon, g(t_0) + \varepsilon]$.

Équivalence de la définition globale avec une définition locale uniforme

La définition (5) peut être qualifiée de "globale". Elle implique la propriété (6) suivante, "locale au sens uniforme", qui est la définition de la dérivabilité donnée en mathématiques constructives.

$$\left[\begin{array}{l}\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \forall x,y \in [a,b] : \\ |y - x| \le \delta \\ \Rightarrow |f(y) - f(x) - g(x)(y - x)| \le \varepsilon |x - y|\end{array}\right] \quad (6)$$

En effet si on a

$$|x - y| \le \delta \Rightarrow |g(y) - g(x)| \le \varepsilon$$

alors, en fixant x, on obtient :

$$|x - y| \le \delta \Rightarrow g(y) \in [g(x) - \varepsilon, g(x) + \varepsilon]$$

et en appliquant (5) pour l'intervalle $[x - \delta, x + \delta] \cap [a, b]$ on obtient (6).

On notera que (6) implique que g(x) est la dérivée de f au point x au sens usuel mais aussi au sens de la stricte dérivabilité (définition (4)).

Avant de démontrer (6) ⇒ (5) notons que (6) admet la formulation équivalente suivante :

la fonction $z \mapsto g(z)$ est limite uniforme, lorsque y et x distincts tendent vers z, de la fonction "taux d'accroissement de f" (2)

$$h(x,y) := (f(y) - f(x)) / (y - x) \qquad (7)$$

Or cette définition peut être lue comme suit :

(2) On peut donc écrire (modulo une petite modification en t1) que la fonction g est limite uniforme de la suite de fonctions : $h_n(x) := n.(f(x+1/n) - f(x))$. Cela montre que l'hypothèse "g uniformément continue", nécessaire dans (5), est superflue dans (6).

la fonction dérivée est la limite de la fonction vitesse moyenne lorsque l'intervalle de temps tend vers 0.

Autrement dit, du point de vue strictement mathématique, la phrase "la vitesse instantanée est la limite de la vitesse moyenne" recèle un flou suffisamment important pour laisser ouvertes des interprétations très diverses, aussi bien celle correspondant à la définition ponctuelle que celle correspondant à la définition locale uniforme.

Passons à la preuve que (6) $\Rightarrow$ (5).

Supposons (6), et, avec $x, y \in [a,b]$, supposons $g([x,y]) \subset [m,M]$. On veut montrer que le taux d'accroissement de la fonction f sur l'intervalle $[x,y]$ est compris entre m et M. Soit $\varepsilon > 0$, il suffira de montrer que le taux d'accroissement est inférieur à $M + \varepsilon$ (la même démonstration prouverait qu'il est supérieur à $m - \varepsilon$).

Considérons un δ convenant à ε dans (6), et découpons l'intervalle $[x,y]$ en un nombre fini d'intervalles $[x_i,x_{i+i}]$ de longueur $\leq \delta$.

On a alors

$$| f(x_{i+i}) - f(x_i) - g(x_i)(x_{i+i} - x_i) | \leq \varepsilon \; | x_{i+i} - x_i |$$

d'où

$$(f(x_{i+i}) - f(x_i)) / (x_{i+i} - x_i) \in [m - \varepsilon, M + \varepsilon]$$

en particulier

$$f(x_{i+i}) - f(x_i) \leq (M + \varepsilon)(x_{i+i} - x_i)$$

et enfin

$$f(y) - f(x) \leq (M + \varepsilon)(y - x)$$

On peut remarquer que cette preuve est essentiellement celle que Cauchy donne pour le théorème des accroissements finis. C'est en fait une preuve simple et naturelle, celle qui va de soi lorsqu'on a formalisé la notion de limite et qu'on cherche à en vérifier l'efficacité. Que cette preuve de Cauchy soit réputée fautive est à vrai dire un grand sujet d'étonnement, car la définition qu'il donne de la fonction dérivée (il ne donne pas celle de la dérivée en un point !) est trop imprécise pour qu'on puisse trancher entre l'interprétation maintenant usuelle et l'interprétation (6).

Signalons pour terminer sur l'équivalence entre (5) et (6) qu'on démontre en mathématiques classiques que cette définition équivaut aussi à "f est continument dérivable sur [a,b]" ou encore à "f est strictement dérivable en tout point de [a,b]".

Efficacité de la définition locale uniforme

La définition (6) "locale uniforme" est celle donnée en mathématiques constructives (*cf.* [BB]).

C'est une définition très efficace parce qu'elle permet de donner des démonstrations constructives, simples et naturelles des théorèmes essentiels sur les fonctions dérivées. En particulier nous avons vu que la preuve de l'inégalité des accroissements finis, sous la forme (6) $\Rightarrow$ (5) est tout à fait naturelle.

Nous laissons à la lectrice [3] le soin de vérifier que les preuves classiques concernant la dérivabilité d'une somme, d'un

(3) Avouez, lecteur mâle, que cela vous a fait un drole d'effet, cette interpellation au féminin. Et le principe de symétrie, alors ?

produit ou d'un quotient, ou celle qui donne l'existence d'une primitive pour une fonction uniformément continue, fonctionnent pratiquement sans aucun changement avec la définition (6), et sont des démonstrations entièrement constructives.

Ici il n'est sans doute pas inutile que nous expliquions un peu en quoi nous disons qu'une preuve peut être déclarée constructive. Par exemple avec le théorème des accroissements finis lui-même. Nous prenons dans tout ce texte l'expression "théorème des accroissements finis" au sens de ce qui est souvent appelé "inégalité des accroissements finis".

Preuves constructives de différentes variantes du théorème des accroissements finis

Le théorème des accroissements finis tel que nous l'avons démontré peut être formulé sous la forme équivalente suivante :

(TAF1) Si la fonction f est dérivable au sens (6) et si $f' \leq 0$ sur [a,b] alors $f(b) \leq f(a)$

Première variante

Lorsqu'on examine en détail la démonstration que nous avons donnée, elle donne en fait un résultat qui a une signification algorithmique plus forte.

(TAF2) Si f est dérivable au sens (6) et si $f(b) > f(a)$, alors on peut expliciter en un temps fini un nombre c dans [a,b] tel que $f'(c) > 0$

Plus précisément cette démonstration fournit le plan d'un algorithme qui traite les données f, a, b, (la fonction f étant dérivable au sens (6)) et fournit en sortie un rationnel $r \in [0,1]$ tel que $f'(a + r(b - a)) > 0$. Pour plus de précisions, consultez l'encadré *(page suivante)*.

Deuxième variante

À partir de la définition "usuelle" de dérivée, il est également possible de fournir une preuve algorithmique du théorème des accroissements finis

(TAF1bis) Si la fonction f est dérivable en tout point de l'intervalle [a,b] selon la définition usuelle et si $f' \geq 0$ sur [a,b] alors $f(b) \geq f(a)$

Nous allons voir qu'il s'agit d'une preuve nettement moins intuitive et que son caractère algorithmique est moins convaincant.

En effet, cette preuve ne fournit pas le théorème (TAF2) mais un théorème (TAF3) avec une hypothèse et une conclusion un peu plus faibles, qui implique (TAF1bis). Dans le théorème (TAF3) le réel x cherché n'est pas explicité en un temps fini : c'est seulement le n-ème terme d'une suite de rationnels convergeant vers x (de manière géométrique) qui est explicité en temps fini :

(TAF3) Si f est dérivable au sens usuel sur l'intervalle [a,b] et si $f(b) > f(a)$, alors on peut expliciter un réel $x \in [a, b]$ tel que $f'(x) > 0$, sous forme d'une limite d'une suite de rationnels convergeant géométriquement vers x

Voici cette preuve, qui fonctionne par dichotomie (quelques détails techniques supplémentaires devraient être rajoutés, mais l'essentiel est donné).

Encadré

Un problème qui se pose lorsqu'on examine le caractère algorithmique d'une preuve en analyse est le suivant : comment un algorithme traite-t-il une fonction uniformément continue en tant que donnée ? C'est-à-dire encore : quelle est la forme "manipulable" (algorithmique) prise par une fonction continue $[a,b] \to \mathbb{R}$? La seule réponse raisonnable semble être la suivante.

La forme manipulable d'une fonction réelle continue sur un intervalle [a,b] est obtenue à travers deux ingrédients :

– d'une part, un module de continuité uniforme pour la fonction f, c.-à-d. une fonction mc : $\mathbb{N} \to \mathbb{N}$ vérifiant :

$$(\mid x - y \mid \leq 1/2^{mc(n)}) \Rightarrow (\mid f(x) - f(y) \mid \leq 1/2^n).$$

– d'autre part, pour chaque $n \in N$ et chaque nombre rationnel $t \in [a,b]$, nous devons être capables de calculer un rationnel z vérifiant

$$\mid f(t) - z \mid \leq 1/2^n.$$

Notez à ce sujet que toutes les fonctions qu'on calcule vraiment, et en particulier la plupart des fonctions utiles en mathématiques, peuvent être traitées de cette manière algorithmique.

Lorsqu'on cherche à expliciter le contenu algorithmique de la définition usuelle de la continuité en tout point pour une fonction f pour laquelle le deuxième ingrédient ci-dessus serait fourni, on tombe le problème suivant, un peu inattendu. En un point x irrationnel, la valeur de f(x) peut être a priori connue comme limite de n'importe quelle suite $f(x_n)$ où (x_n) est une suite de Cauchy de rationnels convergeant géométriquement vers x. Cependant pour expliciter f(x) nous devons savoir avec quelle vitesse la suite $f(x_n)$ converge vers f(x). *A priori,* selon la définition usuelle, cette vitesse dépend elle-même du point x (pour tout x et tout ε il existe δ tel que etc.). Il faudrait donc être capable de spécifier une vitesse de convergence particulière pour chaque réel x. Personne n'est capable d'un tel exploit. En tout cas personne n'en a été capable jusqu'à maintenant. Heureusement qu'il y a la continuité uniforme. Elle nous permet de supprimer le "pour tout x" qu'il est si difficile de rendre algorithmique.

Signalons à ce sujet dans le même ordre d'idées que la compréhension algorithmique des fonctions continues définies sur des espaces métriques non localement compacts est un sujet qu'on est bien loin d'avoir complètement éclairci.

On note $\Delta = (f(b) - f(a))/(b - a)$. On considère $\varepsilon > 0$, $\varepsilon_n = \varepsilon/2^n$. A l'étape numéro n on détermine un intervalle $[a_n,b_n]$ de longueur $(b - a)/2^n$ sur lequel le taux d'accroissement de la fonction f est $\geq \Delta - \varepsilon + \varepsilon_n$, les intervalles successifs étant emboîtés. L'initialisation avec n = 0 et l'intervalle [a,b] est claire. Ensuite on remarque que lorsqu'on divise un intervalle en deux intervalles égaux, le taux d'accroissement global T est la moyenne des deux taux d'accroissements T' et T" sur les intervalles de longueur moitié. Il suffit alors de calculer chacun des deux réels T' et T'' avec une précision suffisante, pour être sûr que l'un des deux au moins est supérieur à $T - \delta$ où δ est un rationnel strictement positif arbitraire. Donc si $T = T_n \geq \Delta - \varepsilon + \varepsilon_n$, en calculant T' et T'' avec une précision suffisante, on sera assuré que $T' \geq T - \varepsilon_n/2$ ou $T'' \geq T - \varepsilon_n/2$ c'est-à-dire $T' \geq \Delta - \varepsilon + \varepsilon_{n+1}$ ou $T' \geq \Delta - \varepsilon + \varepsilon_{n+1}$. La limite commune des deux suites (a_n) et (b_n) est un réel c de [a,b]. La suite (a_n) converge en croissant vers c et la suite (b_n) converge en décroissant vers c. Finalement le nombre dérivé f ' (c) est $\geq \Delta - \varepsilon$ car il est la limite du taux d'accroissement sur l'intervalle $[a_n,b_n]$ (utiliser la propriété (3)).

Comparaison des deux variantes

Si nous comparons le contenu algorithmique des deux preuves, nous voyons que la première fournit un rationnel r convenable en un nombre fini d'étapes, tandis que la seconde fournit un réel x comme limite d'une suite de Cauchy (explicite et explicitement convergente). Par ailleurs l'hypothèse dans la deuxième preuve est plus faible que dans la première.

En fait, bien que nous ayons réussi à faire fonctionner la définition ponctuelle à l'intérieur d'une preuve de nature algorithmique, cette définition usuelle de la dérivabilité en tout point *n'a pas de signification algorithmique connue.* Plus précisément, personne n'a la moindre idée de comment on pourrait rendre entièrement explicite la définition usuelle de la dérivabilité d'une fonction en tout point d'un intervalle. En effet, on ne sait pas expliciter le fait que "le nombre dérivé existe en tout point de l'intervalle" autrement qu'en exprimant ce nombre dérivé comme fonction uniformément continue de la variable x et en donnant un module de convergence, uniforme par rapport à x, qui explicite la manière dont le taux d'accroissement tend vers le nombre dérivé en x. Mais alors on tombe sur la définition (6) et non sur la définition usuelle.

Réexpliquons ce phénomène étrange sous une autre forme. Si nous faisons fonctionner la deuxième preuve de manière algorithmique sur un cas particulier où la fonction dérivée est seulement supposée exister au sens de la définition habituelle, nous construisons bien un réel x en lequel le nombre dérivé est $\geq \Delta - \varepsilon$, mais nous n'avons *a priori* aucune autre information plus précise sur ce nombre dérivé. Le nombre réel x est bien maîtrisé, mais pas le nombre dérivé en x. En effet, le taux d'accroissement entre a_n et b_n tend vers ce nombre dérivé, mais *avec une vitesse totalement inconnue*, car cette vitesse dépend *a priori* du nombre *précis* x et celui-ci n'est jamais connu *que de manière approchée.* Ainsi par exemple les hypothèses ne fournissent pas le moyen de connaître le nombre dérivé en x avec une précision qu'on aurait fixée *a priori*. Pour parler plus crûment, les soi-disant suites de Cauchy pour lesquelles la convergence n'est pas assurée de manière explicite sont de pauvres objets sans aucune utilité pratique, sauf celle de produire des théorèmes fort abstraits au moyen de preuves par l'absurde.

Nous entendons déjà le lecteur protester et nous dire que nous avons poussé le bouchon un peu trop loin, et que lui connaît bien des fonctions qui sont partout dérivables de manière explicite, mais dont la dérivée n'est pas continue. Il va nous donner l'exemple de la fonction

$$f(x) = x^2 \sin (1/x)$$

qui a une dérivée nulle au point 0 et égale à 2x sin (1/x) – cos (1/x) en tout point x distinct de 0. Eh bien, justement, cette fonction dérivée g(x) n'est pas une fonction qu'on puisse vraiment maîtriser d'un point de vue algorithmique. Si on vous demande de calculer g(x) de manière algorithmique, vous en êtes incapables lorsque la donnée x est un nombre réel défini de manière usuelle (à travers une suite de Cauchy de rationnels) pour lequel vous hésitez (peut être de manière inévitable) quant au test d'égalité à 0. En d'autres termes, la preuve classique qui montre que la fonction f ci-dessus est dérivable en tout point admet *une faille* lorsqu'on examine son contenu algorithmique. Au lieu de nous permettre de calculer g(x) en fonction de x, cette preuve nous informe seulement qu'il serait absurde de supposer qu'il existe un point x où la fonction f ne serait pas dérivable. Autrement dit, au lieu de nous donner l'information souhaitée dans l'énoncé, la preuve est une pure preuve par l'absurde. Mais on ne saurait valablement prouver de cette manière un fait de nature positive [(4)]. D'ailleurs, la fonction f(x) ne saurait absolument pas modéliser un mouvement concrètement existant dans la réalité physique, car le véhicule animé d'un tel mouvement subirait des accélérations en tous sens et prenant des valeurs arbitrairement grandes avant de s'arrêter (à l'instant 0) pour repartir immédiatement après dans une sarabande non moins effrénée, et tout ceci donnerait le mal de mer à des gens moins sensibles que Zénon d'Elée aux paradoxes de l'infini en acte.

(4) Voir mon article "Le raisonnement par l'absurde", *Repères IREM* n°29, octobre 1997.

Le théorème des accroissements finis est-il aussi un théorème d'algèbre ?

Puisque les théorèmes essentiels sur les fonctions dérivées sont démontrables constructivement avec la définition (6), nous obtenons une preuve complètement explicite du théorème (TAF2) lorsque la fonction f est une fonction polynome, ou une fraction rationnelle dont le dénominateur reste minoré par un réel $m > 0$ sur un intervalle $[a,b]$.

Ceci nous amène à poser la question suivante : le théorème des accroissements finis, sous la forme (TAF2), est-il un théorème d'algèbre ? C'est-à-dire encore :

> Le théorème (TAF2) est-il encore valable si f est un polynome ou une fraction rationnelle dont le dénominateur est supposé minoré par $m > 0$ sur l'intervalle considéré, et si le corps des réels est remplacé par un corps ordonné arbitraire ?

La réponse est sûrement oui lorsque l'on considère un corps ordonné $K \subset \mathbb{R}$, puisque la preuve de (TAF2) n'utilise aucun passage à la limite, contrairement à celle donnée pour (TAF3), mais seulement des minorations et des majorations. Le passage à la limite dans la preuve de (TAF3) est légitime parce que toute suite de Cauchy admet une limite dans $\mathbb{R}$. Dans les deux preuves de (TAF2) et (TAF3) on utilise le caractère archimédien du corps $\mathbb{R}$: tout réel est majoré par un entier, dans (TAF3) on utilise en plus le caractère complet (toute suite de Cauchy est convergente).

Il existe une preuve purement algébrique de (TAF2) dans le cas général d'un corps ordonné non nécessairement archimédien, mais seulement avec une fonction polynome, nous la donnons en annexe.

Nous revenons pour terminer à un autre aspect de la discussion générale sur le caractère plus ou moins constructif, algorithmique, des preuves en analyse. Et ceci, à partir de l'exemple du "théorème de Rolle".

Le théorème de Rolle : un théorème non algorithmique

Donnons tout d'abord une version constructive du théorème de Rolle :

(R1) Si la fonction f est dérivable au sens (6) et si $f(a) = f(b)$, alors $\inf\{|f'(t)| ; t \in [a,b]\} = 0$

Preuve algorithmique : soit un rationnel $\varepsilon > 0$ et découpons $[a,b]$ en intervalles $[x_i,x_{i+i}]$ de longueur $\leq \delta$ sur chacun desquels la fonction f ' varie de moins que $\varepsilon/4$. Pour chaque i, on calcule $f'(x_i)$ avec une précision meilleure que $\varepsilon/4$, ce qui permet d'affirmer au moins une des 2 éventualités :

$|f'(x_i)| \geq \varepsilon/2$

ou

$|f'(x_i)| \leq \varepsilon$

Si on avait $|f'(x_i)| \geq \varepsilon/2$ pour chaque i, alors $f'(x_i)$ resterait de signe constant (car $|f'(x_{i+i}) - f'(x_i)| \leq \varepsilon/4$), par exemple > 0. dans ce cas $f'(x)$ resterait $\geq \varepsilon/4$ sur $[a,b]$ et on aurait, par (5)

$f(b) \geq f(a) + \varepsilon/4\,(b - a)$

Cette preuve algorithmique nous donne le moyen de trouver (sous l'hypothèse donnée en (R1)), pour chaque entier n, en un nombre fini d'étapes, majoré à priori, un rationnel x_i tel que

$|f'(x_i)| \leq 1/2^n$

Elle ne nous permet cependant pas de réaliser explicitement le théorème de Rolle classique. Nous n'obtenons pas de méthode générale pour construire un nombre réel c (c'est-à-dire : construire une suite de Cauchy de rationnels) pour lequel $f'(c) = 0$.

Quant à la preuve usuelle du théorème de Rolle classique, elle est basée sur le fait que la fonction f atteint son maximum sur l'intervalle [a,b]. Et ce théorème d'analyse classique est fortement non algorithmique : en effet, on peut construire une fonction f uniformément continue sur un intervalle [a,b], de manière complètement explicite, et telle que cependant f n'atteigne son maximum en aucun point "mécaniquement calculable" de l'intervalle [a,b]. En conséquence, aucun algorithme ne peut fournir, à partir d'une fonction uniformément continue arbitraire explicite f, un point x où cette fonction atteindrait son maximum. Même dans l'hypothèse où f est supposée dérivable au sens (6), il semble impossible de rendre le théorème de Rolle entièrement explicite par un algorithme général, parce que cet algorithme général serait capable de réaliser le théorème de la valeur intermédiaire pour la fonction dérivée qui peut être *a priori* n'importe quelle fonction continue. Et on peut montrer qu'il n'existe pas d'algorithme général pour le théorème de la valeur intermédiaire.

Conclusion

Une préoccupation de réalisme nous a tout d'abord incité à demander un changement de statut pour l'inégalité des accroissements finis : elle démontre, non pas une vérité, mais l'adéquation d'un modèle à une réalité qu'il voulait représenter.

Ceci nous a conduit à proposer une

définition alternative pour la notion de fonction dérivable, basée sur l'inégalité des accroissements finis. Nous avons abouti à une formulation équivalente à celle donnée par les mathématiques constructives. Cette définition alternative de fonction dérivable s'avère en fin de compte très efficace du point de vue algorithmique.

Pour passer de la formulation usuelle "vitesse instantanée = limite de la vitesse moyenne" à la formulation constructive alternative, il suffit de lire le mot limite au sens uniforme.

Ceci n'est pas sans rappeler une mésaventure qui arriva à Cauchy. Il proposa une définition de la continuité (5) et il s'attaqua à un théorème qui, à l'époque, avait quasiment le statut d'une évidence : "si une suite de fonctions continues admet une limite, sa limite est une fonction continue".

Ce théorème n'est ni vrai, ni faux. Tout dépend du contexte, du sens précis qu'on attribue au mot limite. Cauchy donna en 1821 une démonstration "fausse". On connut rapidement des contre-exemples. Mais l'erreur dans la démonstration ne fut découverte que 26 ans plus tard. C'est Seidel qui mit clairement à jour le point faible : la notion de limite d'une suite de fonctions devait être prise au sens uniforme et non au sens point par point, le premier qui venait à l'esprit. Pour plus de détails sur ce sujet, on lira *Preuves et réfutations* de Imre Lakatos, pp. 165-182.

(5) Voir mon article "L'uniformité, un concept implicite efficace chez Cauchy", *Repères IREM* n°5, 1991.

BIBLIOGRAPHIE

La bibliographie qui suit donne des ouvrages cités dans l'article mais aussi quelques références pour la lectrice intéressée par les mathématiques constructives.

[Bee] BEESON M. J. : *Foundations of Constructive Mathematics* (1985) Springer. (c'est un livre de logique). On pourra aussi consulter dans la brochure Epiphymat de Besançon (1992), la traduction légèrement abrégée, par H. Lombardi, d'un article de M. Beeson : "principes problématiques en mathématiques constructives".

[Bis] BISHOP E. : *Foundations of Constructive Analysis*, McGraw Hill (1967) (épuisé).

[BB] BISHOP E., BRIDGES D. : *Constructive Analysis*, Springer-Verlag (1985).

[BR] BRIDGES D., RICHMAN F. : *Varieties of Constructive Mathematics*, London Math. Soc. LNS 97. Cambridge University Press (1987). (comparaison dans la pratique des trois principales tendances du constructivisme)

[Dem] DEMAZURE M. : *Catastrophes et bifurcations*, Ellipses (1989).

[Lak] LAKATOS I. : *Preuves et réfutations*, versions française, Hermann (1984).

[Lom] LOMBARDI H. : *Mathématiques constructives*, Brochure IREM de Besançon. (1994).

COLLECTIF : *"PENSER LES MATHÉMATIQUES"*, Collection Seuil Points.

On comparera avec intérêt les articles de Dieudonné et de Apéry. Une bibliographie intéressante se trouve à la fin de l'article d'Apéry.

Commentaires

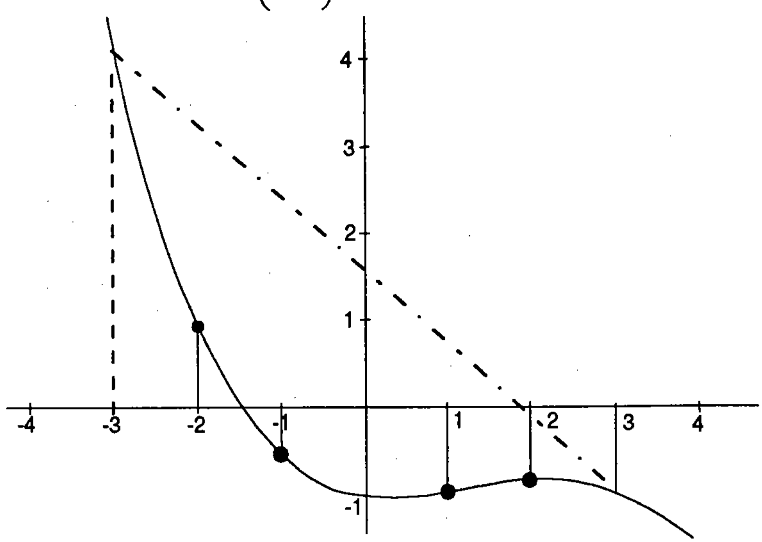
Il n'existe pas actuellement de livre de référence en français donnant un exposé de base

sur les mathématiques constructives. Ceci existe en langue anglaise : voir [BB] et [MRR]. Les livres [Bee] et [BR] font une étude comparée des différents points de vue constructifs existants.
Les divergences de point de vue existant entre différents mathématiciens constructifs (ou intuitionnistes) ne doivent pas cacher l'existence d'un socle fondamental commun, basé sur la critique de la notion de vérité absolue en mathématiques. Le livre de Bishop, repris dans [BB], qui reste le plus prudent, contient ce socle fondamental commun.
Enfin, il faut signaler que, même si la position philosophique constructive n'est pas largement représentée chez les mathématiciens, le courant dans la recherche mathématique qui s'intéresse particulièrement aux résultats de nature effective et algorithmique a toujours été vivant et connaît actuellement un fort développement.

Annexe 1 : le théorème algébrique des accroissements finis

Le théorème algébrique des accroissements finis affirme que le taux d'accroissement d'un polynome de degré fixé sur un intervalle est une moyenne, pondérée par des coefficients positifs, de valeurs de la dérivée en des points connus de l'intervalle, indépendants du polynome.

$$y = \frac{1}{200}x^4 + \left(-\frac{1}{12}\right)x^3 + \frac{1}{4}x^2 + \frac{1}{8}(-x) - 1$$



Exemple : Pour tout polynome de degré ≤ 4 on a l'identité :

$$\frac{P(a)-P(b)}{a-b} = \left(\frac{1}{3}P'\left(\frac{a}{6}+\frac{5b}{6}\right)+\frac{1}{6}P'\left(\frac{a}{3}+\frac{2b}{3}\right)+\frac{1}{6}P'\left(\frac{2a}{3}+\frac{b}{3}\right)+\frac{1}{3}P'\left(\frac{5a}{6}+\frac{b}{6}\right)\right)$$

Plus généralement on a les résultats suivants :

Lemme : Il existe deux suites $(\lambda_{i,j})_{1\le i\le j; j=1,2,...,n,...}$ et $(r_{i,j})_{1\le i\le j; j=1,2,...,n,...}$ de rationnels $\in]\,0, 1\,[$ telles que, pour tout polynome $P \in \mathbb{Q}[X]$ de degré $\le n$, on ait :

$$P(a) - P(b) = (a - b) \times \sum_{i=1}^{n} r_{i,n} \, . \, P'(a + \lambda_{i,n}(b - a))$$

Théorème : (théorème algébrique des accroissements finis pour les polynomes)
Il existe deux suites $(\lambda_{i,j})_{1\le i\le j; j=1,2,...,n,...}$ et $(r_{i,j})_{1\le i\le j; j=1,2,...,n,...}$ de rationnels $\in]\,0, 1\,[$ telles que, pour tout corps ordonné **K** et tout polynome $P \in \mathbf{K}[X]$ de degré $\le n$, on ait :

$$P(a) - P(b) = (a - b) \times \sum_{i=1}^{n} r_{i,n} \, . \, P'(a + \lambda_{i,n}(b - a))$$

En particulier, si P' est de signe positif sur un intervalle, la fonction polynome est croissante sur tout l'intervalle, et si le taux d'accroissement de P est 0 entre a et b la dérivée de P est > 0 en l'un au moins des points $a + \lambda_{i,n}(b - a)$.

preuve> Le théorème est une conséquence immédiate du lemme : ce dernier fournit en effet des identités algébriques concernant les variables "a, b, et les coefficients du polynome" qui s'appliquent alors dans tout anneau commutatif qui est une $\mathbb{Q}$-algèbre, et en particulier dans les corps ordonnés.

Démontrons le lemme. Par changement de variable affine, on peut supposer que $a = -1$ et $b = 1$.

Considérons le degré n fixé. Et proposons nous de résoudre la question d'abord dans le cas du corps des réels R. Soit E l'espace des polynomes de degré $\le$ n. C'est un espace vectoriel de dimension n + 1.

L'application $P \le P(1) - P(-1)$ est une forme linéaire sur E ne faisant pas intervenir le coefficient constant. Les formes linéaires sur E ne faisant pas intervenir le coefficient constant forment un espace vectoriel réel F de dimension n. Pour tout choix de n réels $\mu_{i,n}$, les n formes linéaires $P \le P'(\mu_{i,n})$ forment une base de F. En effet, ce sont des éléments de F car elles ne font pas intervenir le coefficient constant. Et elles sont indépendantes parce qu'on retrouve les coefficients de P' à partir des valeurs $P'(\mu_{i,n})$ par interpolation.

Il correspond donc à tout choix de n réels $\mu_{i,n}$ des réels $r_{i,n}$ qui rendent la formule vraie pour tout polynome de degré $\le$ n :

$$P(1) - P(-1) = \sum_{i=1}^{n} r_{i,n} \, . \, P'(\mu_{i,n})$$

En outre, les $r_{i,n}$ peuvent être calculés à partir des $\mu_{i,n}$ en résolvant un système de Cramer (les n équations de ce système de Cramer sont obtenues en remplaçant dans l'identité ci-dessus le polynome P successivement par les monomes $X, X^2, \dots, X^n$). On voit donc que les $r_{i,n}$ dépendent continument et rationnellement des $\mu_{i,n}$.

La difficulté consiste à déterminer des $\mu_{i,n}$ rationnels $\in]-1, 1[$ tels que les $r_{i,n}$ correspondants soient dans $]\,0, 1\,[$. Les formules de quadrature de Gauss correspondent à un tel choix, mais avec des réels alors que nous voulons des rationnels. Il suffit alors de choisir des $\mu_{i,n}$ rationnels suffisamment voisins des $\mu_{i,n}$ de Gauss (ce sont les zéros des polynomes de Legendre) pour que les $r_{i,n}$ correspondants restent positifs.

NB : On trouvera les formules de quadrature de Gauss dans les livres d'analyse numériques. Un très bon exposé se trouve dans : *Analyse Numérique, F. Scheid, Série Schaum (1986). Il serait par ailleurs intéressant de trouver une preuve plus directe du lemme, sans passer par la formules de quadrature de Gauss. Notez aussi que notre preuve ne donne pas une suite de formules bien carrées, mais seulement un moyen explicite pour obtenir, pour chaque entier n, une formule convenable.*

Annexe 2 : la stricte dérivabilité et le théorème d'inversion locale

Dans cette annexe nous rappelons la preuve naturelle du théorème d'inversion locale dans le cadre des fonctions strictement dérivables.
Une fonction réelle continue f de plusieurs variables définie dans un ouvert U contenant un point $m \in \mathbb{R}^n$ est dite strictement dérivable au point m, avec une dérivée égale à φ en m (où φ est une forme linéaire sur $\mathbb{R}^n$) si la fonction affine

$p \mapsto f(m) + \varphi(p - m)$

est une bonne approximation de la fonction f au voisinage de m au sens suivant :

$\forall \varepsilon > 0 \ \exists \delta > 0 \ \forall p, q \in U$
si $\| p - m \| < \delta$, $\| q - m \| < \delta$ et $p \neq q$ alors $| f(p) - f(q) - \varphi(p - q) | < \varepsilon \| p - q \|$

La forme linéaire φ a pour coefficients (sur la base duale canonique) les dérivées partielles $\delta f/\delta x_j$.
De même une application continue F d'un ouvert U de $\mathbb{R}^n$ et à valeurs dans $\mathbb{R}^p$ est dite *strictement dérivable au point* m, avec une dérivée égale à ϕ en m (où ϕ est une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$) si la fonction affine

$p \mapsto f(m) + \varphi(p - m)$

est une bonne approximation de la fonction f au voisinage de m au sens suivant :

$\forall \varepsilon > 0 \ \exists \delta > 0 \ \forall p, q \in U$
si $\| p - m \| < \delta$, $\| q - m \| < \delta$ et $p \neq q$ alors $> f(p) - f(q) - \varphi(p - q) - \phi \| p - q \| < \varepsilon \| p - q \|$

En fait l'application F est donnée par p fonctions réelles ($f_1, \dots, f_p$) et l'application linéaire ϕ a pour matrice la matrice jacobienne des dérivées partielles $\delta f_i / \delta x_j$
Le théorème d'inversion locale concerne une application continue F d'un ouvert U de R^n et à valeurs dans $\mathbb{R}^n$. Sous l'hypothèse que F est strictement dérivable au point m et que la dérivée ϕ en m est une application linéaire inversible (c.-à-d. le jacobien de F est non nul) alors il existe un ouvert V contenant m et contenu dans U tel que F réalise une bijection entre V et l'ouvert F(V). En outre la bijection réciproque $F^{-1} : F(V) \to V$ est strictement dérivable en F(m) et sa dérivée en ce point est égale à ϕ^{-1}.
La preuve d'une simplicité étonnante est une simple application du théorème du point fixe. C'est la suivante. Quitte à remplacer F par $\phi^{-1} \circ F$ on peut supposer que $\phi = \mathrm{Id}$, ce que nous supposons désormais.
On suppose aussi que $m = F(m) = 0$ l'origine de R^n, ce qui n'est évidemment pas une restriction.

Considérons l'ouvert $W = B_0(r)$ qui est une boule de centre $m = 0$ contenue dans U et dont le rayon r est inférieur ou égal à la valeur δ correspondant à $\varepsilon = 1/2$ dans la définition de la stricte dérivabilité (n'importe quelle valeur plus petite que 1 pourrait faire l'affaire).
On définit une application v comme étant l'écart de F avec l'identité :

$$F(p) = p - v(p)$$

On a donc $\| v(p) - v(q) \| \leq \| p - q \| / 2$ pour tous p et q dans W. En particulier

$$\| F(p) - F(q) \| \geq \| p - q \| - \| p - q \| / 2 = \| p - q \| / 2$$

et la restriction de F à W est injective.
Pour un y suffisamment proche de 0 on cherche x dans W vérifiant

$$F(x) = x - v(x) = y$$

on cherche x sous la forme $y + \Delta y$ où Δy est un accroissement inconnu qui doit vérifier

$$\Delta y = v(x) = v(y + \Delta y)$$

autrement dit on cherche un point fixe Δy pour l'application

$$v_y : z \mapsto v(y + z)$$

or on a

$$\| v_y(z) - v_y(t) \| = \| v(y + z) - v(y + t) \| \leq \| (y + z) - (y + t) \| / 2 = \| z - t \| / 2$$

donc l'application v_y est (1/2)-lipschitzienne et à ce titre elle admet bien un point fixe unique, qui est la limite de la suite

$$y_n = (v_y \circ v_y \circ \ldots \circ v_y)(0) \text{ (n fois)}$$

on a la majoration

$$\| y_{n+1} - y_n \| = \| v_y(y_n) - v_y(y_{n-1}) \| \leq \| y_n - y_{n-1} \| /2 \leq \| y_1 - y_0 \| /2^n = \| y \| /2^n$$

et donc le point fixe Δy limite de la suite y_n vérifie

$$\| \Delta y \| \leq 2\, y$$

Finalement la rigueur du calcul précédent peut être établie si on précise ce que veut dire "y suffisamment proche de 0" au début du calcul : il faut se limiter à des y qui vérifient

$$\| 3y \| \leq r$$

Autrement dit pour tout y dans la boule $B_0(r/3)$ le calcul précédent est valable et fournit un point $x = y + \Delta y$ vérifiant $x \in B_0(r)$ et $F(x) = y$.
Finalement on prendra pour V l'ouvert $F^{-1}(B_0(r/3))$ qui est contenu dans W.
Nous laissons à la lectrice les vérifications concernant la continuité dans $B_0(r/3)$ et la stricte dérivabilité en 0 de la bijection réciproque de F.