Transport optimal de mesure : coup de neuf pour un très vieux problème

Bien connue depuis plusieurs siècles pour ses applications logistiques et économiques, la problématique du transport optimal de mesure connaît actuellement un renouveau spectaculaire à cause de ses liens insoupçonnés avec la mécanique des fluides, les équations aux dérivées partielles et d'autres domaines des mathématiques.

On dit souvent que l'une des plus grandes satisfactions des mathématiciens consiste à établir des liens entre domaines *a priori* éloignés. L'histoire récente du transport optimal de mesure est à cet égard très représentative. Initié à la fin du XVIIIe siècle par Monge, développé par Kantorovich au milieu du XXe siècle pour ses applications en économie, ce sujet a connu une renaissance spectaculaire dans les 15 dernières années, à partir des travaux de Brenier en mécanique des fluides. Les spécialistes actuels, peut-être frappés de délire monomaniaque voient maintenant du transport optimal partout : depuis les équations semi-géostrophiques en météorologie jusqu'aux problèmes isopérimétriques, en passant par les milieux granulaires, la physique statistique et les inégalités de Sobolev. Essayons de retracer quelques étapes de cette renaissance. Notons que, conformément à un phénomène assez courant, Brenier a « redécouvert » certains résultats déjà connus ; mais que cette redécouverte, loin d'être superflue, a apporté un nouvel éclairage au domaine qui, sans cela, n'aurait certainement pas acquis sa notoriété actuelle.

Pour comprendre le contexte dans lequel s'inscrivaient les travaux de Brenier, commençons par quelques rappels élémentaires de mécanique des fluides. L'**équation d'Euler incompressible** est l'une des équations les plus simples, les plus anciennes et les plus mystérieuses de toute la mécanique des fluides. Dans sa formulation dite Lagrangienne, elle peut se décrire comme suit. Soit Ω un ouvert de $\mathbb{R}^3$, modélisant le récipient contenant le fluide, de volume normalisé à 1 ; on notera λ la mesure de Lebesgue restreinte à Ω. L'état du fluide est modélisé par une application $g(t,x)$, où t est la variable de temps et x la variable d'espace ; $g(t,x)$ représente la position au temps t d'une « particule » qui au temps 0 se serait trouvée au point x. Nous supposerons que le flot est régulier, au sens où pour tout t, l'application $g(t,\cdot)$ est un difféomorphisme de Ω sur Ω. Pour traduire l'incompressibilité du fluide, on impose que g *préserve la mesure de Lebesgue* : $g\#\lambda = \lambda$ (encadré 1). En d'autres termes, le volume occupé par un ensemble de particules donné ne varie pas au cours du temps. La fonction $g(t,\cdot)$ appartient donc au groupe SDiff(Ω) des *difféomorphismes de Ω préservant la mesure de Lebesgue*.

Dans ce formalisme, l'équation d'Euler s'écrit

$$\frac{\partial^2 g}{\partial t^2}(t,x) = \nabla p(t,x)\,, \qquad (1)$$

où l'on note ∇p le vecteur des dérivées partielles de la fonction scalaire p (« pression ») par rapport à x_1, x_2, x_3. Il ne faut pas spécifier d'équation sur p : cette marge de manœuvre est indispensable pour compenser la contrainte $g \in \mathrm{SDiff}(\Omega)$. Une formulation plus connue de l'équation d'Euler, formellement équivalente à (1), porte sur le champ de vitesses u : $\partial u/\partial t + (u\cdot\nabla)u + \nabla p = 0$.

L'un des problèmes ouverts les plus célèbres de la théorie mathématique de la mécanique des fluides consiste à construire des solutions « raisonnables » de l'équation d'Euler. Sous certaines hypothèses de régularité, cela équivaut à construire des trajectoires qui *minimisent l'action, localement en temps* : pour tous temps t_0 et t_1 suffisamment proches, pour toute trajectoire $m(t,\cdot)$ à valeurs dans SDiff(Ω), telle que $m(t_0,\cdot) = g(t_0,\cdot)$ et $m(t_1,\cdot) = g(t_1,\cdot)$, on doit avoir

$$\int_{t_0}^{t_1}\left(\int_\Omega \left\|\frac{dg}{dt}\right\|^2 dx\right)dt \leqslant \int_{t_0}^{t_1}\left(\int_\Omega \left\|\frac{dm}{dt}\right\|^2 dx\right)dt. \qquad (2)$$

En termes géométriques, le problème est de construire des *géodésiques* dans l'espace SDiff(Ω). Remarquons bien qu'il y a deux problèmes possibles :

– Cédric Villani, École normale supérieure de Lyon, UMPA, 46 allée d'Italie, 69364 Lyon cedex 07.
cvillani@umpa.ens-lyon.fr

Encadré 1

MESURE IMAGE, MARGINALES

Rappelons quelques notions élémentaires qui serviront tout au long de l'article.
Soit T une application, μ et ν deux mesures de probabilité. On dit que ν est la **mesure image** *de μ par T, ou que T* **transporte** *la mesure μ sur la mesure ν, et on note (par exemple) $T\#\mu = \nu$, si pour toute application mesurable positive (ou bornée) b on a*

$$\int b(T(x))\,d\mu(x) = \int b(y)\,d\nu(y). \qquad (1)$$

Si $T\#\mu = \mu$, on dit que T **préserve la mesure** *μ. Par (1), on sait alors que $\int b \circ T\,d\mu = \int b\,d\mu$ pour tout $b \geqslant 0$.*
En particulier, pour tout p, $\|T\|^p_{L^p(d\mu)}$ est déterminé : c'est le moment d'ordre p de μ.
Si μ et ν ont des densités respectives f et g par rapport à la mesure de Lebesgue, et si l'application T est injective et définit un changement de variable admissible (par exemple si c'est un C^1 difféomorphisme), alors on établit aisément l'équation

$$f(x) = g(T(x))\,|\det(DT(x))|,$$

où DT désigne l'application Jacobienne de T. C'est la formule classique du changement de variables !
Quand μ et ν sont définies sur le même espace, le transport de mesure peut s'exprimer en termes purement physiques : imaginons μ et ν comme les densités de répartition d'un grand nombre de particules. Ecrire $T\#\mu = \nu$ revient à dire que si les particules sont au départ réparties selon la configuration μ et que l'on transporte chaque particule de l'emplacement x à l'emplacement $T(x)$, alors les particules seront réparties, après transport, selon la configuration ν.
Soit π une mesure de probabilité sur un espace produit $X \times Y$. On appelle **marginales** *de π les mesures de probabilité μ et ν définies comme mesures images de π par les projections $(x, y) \mapsto x$ et $(x, y) \mapsto y$. Prendre la marginale sur X revient à intégrer par rapport à $y \in Y$, et* vice versa. *De manière équivalente, pour toutes fonctions φ et ψ intégrables, on a*

$$\int [\varphi(x) + \psi(y)]\,d\pi(x,y) = \int \varphi\,d\mu + \int \psi\,d\nu.$$

– soit on se donne la valeur de $g(t, \cdot)$ en $t = t_0$ et $t = t_1$, soit g_0 et g_1, et on cherche à construire une solution (une trajectoire optimale) reliant g_0 à g_1. En d'autres termes, on connaît l'état du fluide à deux instants donnés et on cherche à reconstituer la trajectoire entre ces deux instants ;
– soit on se donne la valeur de $g(t, \cdot)$ en $t = t_0$, soit g_0, et sa dérivée (la vitesse initiale des particules), $dg/dt|_{t=t_0}$, et on cherche à prédire la trajectoire du fluide aux instants ultérieurs.

Les deux problèmes ne sont pas équivalents ; le premier est celui qui nous préoccupera. *A priori* plus simple que le second, il recèle cependant des surprises : par exemple, un résultat de Shnirelman implique qu'*il n'existe pas toujours* de trajectoire optimale.

GEODÉSIQUES APPROCHÉES

Au milieu des années 80, pour tenter d'y voir plus clair, Brenier cherchait à construire des géodésiques approchées, par une procédure de discrétisation du temps. Considérons le cas extrêmement simplifié où il n'y a que trois temps : t_0, t_1, et $t_{1/2} = (t_0 + t_1)/2$. On se donne $g_0 = g(t_0, \cdot)$, $g_1 = g(t_1, \cdot)$, et on note $\|m\|_{L^2} = \sqrt{\int m^2\,dx}$. Il est facile de s'apercevoir que la discrétisation du problème (2) consiste à rechercher $g_{1/2} \in \mathrm{SDiff}(\Omega)$ tel que pour tout $m \in \mathrm{SDiff}(\Omega)$,

$$\|g_1 - g_{1/2}\|^2_{L^2} + \|g_{1/2} - g_0\|^2_{L^2} \leqslant \|g_1 - m\|^2_{L^2} + \|m - g_0\|^2_{L^2}. \qquad (3)$$

Comme tous les éléments de $\mathrm{SDiff}(\Omega)$ ont la même norme L^2 (encadré 1), l'équation (3) peut se réécrire : pour tout $m \in \mathrm{SDiff}(\Omega)$,

$$\|g_{1/2} - h\|^2_{L^2} \leqslant \|m - h\|^2_{L^2}, \qquad (4)$$

où $h = (g_0 + g_1)/2$. Autrement dit, le $g_{1/2}$ que nous cherchons doit être la *projection orthogonale* du milieu h de g_0 et g_1, sur le groupe $\mathrm{SDiff}(\Omega)$, au sens de la norme L^2.

Comme le lecteur le vérifiera sans peine, le groupe $\mathrm{SDiff}(\Omega)$ n'est pas convexe ; par ailleurs, il n'est pas fermé au sens de la topologie L^2. La non-convexité empêche d'appliquer les théorèmes classiques de projection sur un convexe, mais ce n'est pas un problème fondamental très sérieux ; en revanche, on ne peut envisager de définir une projection sur un ensemble non fermé (que serait la projection d'un élément de $\overline{\mathrm{SDiff}(\Omega)} \setminus \mathrm{SDiff}(\Omega)$?)... Il faut en conclure que le problème de minimisation (4) est en général *mal posé* : il n'admet pas toujours de solution pour un h général.

Pour contourner ce problème, on va appliquer une procédure classique du calcul des variations : la *relaxation*, qui consiste à remplacer l'espace trop restreint $\mathrm{SDiff}(\Omega)$ par son adhérence $\overline{\mathrm{SDiff}(\Omega)}$. Ce dernier espace est constitué de *toutes* les applications s (pas nécessairement bijectives) préservant la mesure de Lebesgue ; on le notera $S(\Omega)$.

Il est possible de montrer par des théorèmes assez généraux que l'opération de projection sur $S(\Omega)$ est « presque toujours » bien définie. Mais on peut mieux faire et donner une construction plus explicite. Soit π la mesure image sur $\Omega \times \Omega$, définie par $\pi = (m \times h)\#\lambda$. Par définition de la mesure image,

$$\begin{aligned}\|m-h\|_{L^2}^2 &= \int_\Omega |m-h|^2 \\ &= \int_{\Omega\times\Omega} |x-y|^2\, d\pi(x,y).\end{aligned}$$

Par ailleurs, il est facile de vérifier que les *marginales* de π (encadré 1) sont la mesure de Lebesgue $\mu = \lambda$ d'une part, et la mesure $\nu = h\#\lambda$ d'autre part. Notre problème de minimisation peut maintenant être comparé au *problème plus général* qui consiste à minimiser la quantité

$$\int_{\Omega\times\Omega} |x-y|^2\, d\pi(x,y)$$

parmi tous les π, mesures de probabilité sur l'espace produit $\Omega \times \Omega$ dont les marginales sont λ et $h\#\lambda$. En développant le carré, on voit qu'il est équivalent de maximiser la quantité

$$\int_{\Omega\times\Omega} x \cdot y\, d\pi(x,y).$$

En termes probabilistes, nous cherchons à *maximiser les corrélations entre des variables aléatoires de lois respectives μ et ν, et dont la loi jointe serait l'inconnue π.*

Partis d'un problème de mécanique des fluides, nous avons abouti à un problème célèbre d'optimisation : le problème de **Monge-Kantorovich**. Sous sa version la plus générale, on peut l'énoncer ainsi : soient μ et ν deux mesures de probabilité sur des espaces respectifs X et Y, soit $c : X \times Y \to \mathbb{R}_+$ une fonction, dite « fonction de coût » ; le problème consiste à minimiser la *fonctionnelle de coût*

$$\int_{X\times Y} c(x,y)\, d\pi(x,y)$$

parmi toutes les mesures π sur $X \times Y$ admettant μ et ν pour marginales. Formulé pour la première fois sous cette forme par le célèbre économiste-mathématicien russe Leonid Kantorovich, ce problème avait été étudié dès 1780 par Gaspard Monge. L'existence d'un minimiseur est un exercice élémentaire d'analyse fonctionnelle, qui ne nous apporte cependant guère d'informations sur le problème initial. Pour en savoir plus, on peut appliquer un célèbre principe de dualité dû à Kantorovich, et que nous appellerons le *principe du convoyeur* (encadré 2). Dans le cas qui nous intéresse, il aboutit à l'identité de type « minimax »

$$\sup_\pi \int x \cdot y\, d\pi(x,y) = \inf_\varphi \int \varphi\, d\mu + \int \varphi^*\, d\nu, \quad (5)$$

où l'infimum est pris sur toutes les paires (φ, φ^*) de fonctions convexes conjuguées :

$$\begin{cases} \varphi^*(y) = \sup_x (x \cdot y - \varphi(x)) \\ \varphi(x) = \sup_y (x \cdot y - \varphi^*(y)). \end{cases}$$

Le problème variationnel à droite de (5) ne semble guère plus simple que celui qui se trouve à gauche... Cependant, en termes de calcul des variations, il est considérablement plus agréable : sans perte de généralité, on peut fixer la valeur de φ en un point (cela ne change pas la valeur de l'infimum) ; or, les paires de fonctions convexes conjuguées (φ, φ^*), définies sur un ouvert borné, ayant une valeur fixée en un point, forment un sous-ensemble *compact* de l'espace des fonctions continues sur cet ouvert. Le supremum est donc atteint dans le membre de droite de (5), par une paire de fonctions convexes.

Soit maintenant π une mesure optimale dans le problème de gauche de (5), et φ une fonction convexe optimale dans le problème de droite de (5) : en utilisant les propriétés de marginales, on peut écrire

$$\begin{aligned}&\int_{\Omega\times\Omega} x \cdot y\, d\pi(x,y) \\ &\qquad = \int_\Omega \varphi(x)\, d\mu(x) + \int_\Omega \varphi^*(y)\, d\nu(y) \\ &\qquad = \int_{\Omega\times\Omega} [\varphi(x) + \varphi^*(y)]\, d\pi(x,y),\end{aligned}$$

d'où

$$\int_{\Omega\times\Omega} [\varphi(x) + \varphi^*(y) - x \cdot y]\, d\pi(x,y) = 0.$$

Or, on a toujours $\varphi(x) + \varphi^*(y) \geqslant x \cdot y$, et donc nécessairement, pour π presque tous x et y,

$$x \cdot y = \varphi(x) + \varphi^*(y).$$

Dans le langage de l'analyse convexe, on dit que y appartient au sous-différentiel $\partial\varphi(x)$ de φ au point x. Pour peu que les mesures μ et ν soient absolument continues par rapport à la mesure de Lebesgue, on peut montrer que $\partial\varphi(x) = \{\nabla\varphi(x)\}$ et en déduire que $\nu = \nabla\varphi\#\mu$, où l'on considère $\nabla\varphi$ comme une application allant de Ω dans

$\mathbb{R}^n$. A partir de là, il est facile d'identifier notre projection orthogonale comme

$$s = \nabla\varphi^* \circ h = (\nabla\varphi)^{-1} \circ h. \tag{6}$$

Le résultat précédent peut sembler suspect : en composant les deux membres de (6) par $\nabla\varphi$, on obtient

$$h = \nabla\varphi \circ s : \tag{7}$$

l'application h est donc la composition de $\nabla\varphi$ par une application préservant la mesure de Lebesgue – or, nous n'avons fait aucune hypothèse sur h, si ce n'est que $h\#\lambda$ soit absolument continue... Il s'agit précisément du remarquable **théorème de factorisation polaire** de Brenier :

Théorème 1. *Soit Ω un ouvert borné de $\mathbb{R}^d$ ($d \geqslant 1$), λ la mesure de Lebesgue restreinte à Ω, et $h : \Omega \to \mathbb{R}^d$ un champ de vecteurs tel que la mesure image $h\#\lambda$ soit absolument continue. Alors il existe une unique décomposition de h sous la forme*

$$h = \nabla\varphi \circ s,$$

où $\nabla\varphi$ est un gradient de fonction convexe sur Ω, et $s : \Omega \to \Omega$ est une application préservant la mesure de Lebesgue. En outre, s est l'unique projection orthogonale de h sur l'espace $S(\Omega)$.

Ce théorème présente de nombreux points communs avec d'autres théorèmes classiques : en particulier la factorisation polaire bien connue des matrices, $M = SO$ (S symétrique, O orthogonale). Moins bien connue d'ailleurs, est la caractérisation du facteur O dans cette décomposition comme la projection orthogonale de M sur le groupe des matrices orthogonales ! Les géomètres pourront également reconnaître dans le théorème de Brenier une version non linéaire du théorème de décomposition de Hodge.

Laissons désormais de côté la mécanique des fluides et continuons à étudier le problème de minimisation de Monge-Kantorovich. Il existe en général de nombreuses manières de transporter des mesures l'une sur l'autre, et c'est un problème classique que de construire des transports « remarquables », en un certain sens. Or, nous venons de constater que deux mesures μ et ν sur Ω, absolument continues, pouvaient être transportées l'une sur l'autre par un gradient de fonction convexe. Voici un énoncé un peu plus général :

Théorème 2. *Soient $d\mu(x) = f(x)\,dx$ et $d\nu(y) = g(y)\,dy$ deux mesures de probabilité sur $\mathbb{R}^d$, absolument continues par rapport à la mesure de Lebesgue. Alors il existe un unique gradient de fonction convexe*

Encadré 2

LE PRINCIPE DU CONVOYEUR (« SHIPPER »)

C'est une façon imagée d'exprimer la **dualité de Kantorovich***. Soit un mathématicien industriel devant organiser le convoi de sa production de charbon depuis les mines jusqu'aux usines. La production et la consommation de charbon sont représentées par des mesures positives, les quantités totales étant en adéquation (les mesures ont même masse). Le souci de notre mathématicien est de minimiser le coût dépensé en transport, soit $\int c(x, y)\,d\pi(x, y)$, où $c(x, y)$ représente le coût du transport de x à y, et $d\pi(x, y)$ la quantité élémentaire de charbon transportée du point x au point y. Les marginales de π sont fixées : ce sont respectivement les densités de charbon produit, et consommé ; nous sommes donc en présence d'un problème de Monge-Kantorovich. Un autre mathématicien se manifeste alors et suggère de lui sous-traiter le problème de transport : « Je me contenterai de te faire payer un prix à l'embarquement et un prix au débarquement ; ces prix varieront en fonction de l'emplacement, et je suis prêt à octroyer des compensations financières (prix négatifs) pour certains endroits. Tu y seras forcément gagnant, car la somme du prix d'embarquement et du prix de débarquement sera* **toujours** *inférieure ou égale au prix que tu paierais pour faire transporter la marchandise ! » Bien sûr, l'affaire est conclue.*

Si l'on note $\varphi(x)$ le prix à payer pour embarquer au point x et $\psi(y)$ le prix à payer pour débarquer au point y, on voit que le convoyeur se fait payer

$$\int \varphi(x)\,d\mu(x) + \int \psi(y)\,d\nu(y). \tag{1}$$

Son problème est donc maintenant de fixer des prix φ et ψ de la manière la plus avantageuse, c'est-à-dire de façon à maximiser (1) tout en respectant la contrainte $\varphi(x) + \psi(y) \leqslant c(x,y)$, qui seule garantit que son offre est suffisamment attrayante pour qu'on lui laisse la charge de **tout** *le transport.*
Le **principe du convoyeur** *(dualité de Kantorovich) assure que la somme d'argent (1) peut être aussi proche que l'on souhaite du coût optimal de Monge-Kantorovich. En termes mathématiques,*

$$\inf_{\pi} \int c(x,y)\,d\pi(x,y) = \sup_{(\varphi,\psi)} \int \varphi\,d\mu + \int \psi\,d\nu,$$

où l'infimum à gauche est pris sur toutes les mesures π de marginales μ et ν et le supremum à droite est pris sur toutes les paires de fonctions de prix (φ,ψ) vérifiant l'inégalité $\varphi(x) + \psi(y) \leqslant c(x,y)$ pour tous x et y.

(unique au sens de : déterminé $f(x)\,dx$-presque partout) $\nabla\varphi$, tel que

$$\nabla\varphi\#\mu = \nu.$$

De plus, si les seconds moments $\int |x|^2\,d\mu(x)$ et $\int |y|^2\,d\nu(y)$ sont finis, alors φ est l'unique minimiseur de la fonctionnelle de coût quadratique

$$\int |x - T(x)|^2\,d\mu(x)$$

parmi toutes les applications T qui transportent μ sur ν.

Si f, g et φ sont suffisamment régulières, alors on peut facilement en déduire (encadré 1) que φ est une solution de l'**équation de Monge-Ampère**

$$\det(D^2\varphi)(x) = \frac{f(x)}{g(\nabla\varphi(x))}.$$

L'étude de cette équation fort célèbre est considérée comme extrêmement ardue, du fait de son caractère « très non linéaire ». Le théorème 2 fournit donc une méthode, remarquablement simple et générale, pour construire des *solutions faibles* de cette équation. C'est un problème très délicat que de savoir si ces solutions faibles sont des solutions classiques... Dans une série d'articles difficiles, Caffarelli montre que si f et g sont strictement positives et de classe $C^{k,\alpha}$ ($k \in \mathbb{N}$, $\alpha \in]0,1[$), au sens où leurs dérivées d'ordre k sont Hölder-continues d'exposant α, alors φ est régulière, de classe $C^{k+2,\alpha}$. Ce résultat est bien sûr optimal dans sa catégorie.

INTERPOLATION DE McCANN

Changeons radicalement de paysage physique. Au début des années 1990, McCann travaillait à démontrer l'unicité des formes d'équilibre de certains systèmes physiques (étoile, gaz en interaction...) dont l'état est modélisé par une mesure de probabilité. Le problème mathématique se pose de la manière suivante : étant donnée telle ou telle fonctionnelle d'énergie F, peut-on démontrer qu'un minimiseur existe et qu'il est unique sous telle ou telle contrainte ? Le problème isopérimétrique et ses variantes (formes des cristaux, des gouttes d'eau, etc.) appartiennent à cette catégorie. Ces résultats d'unicité sont le plus souvent subordonnés à des propriétés de *stricte convexité*, selon un schéma de preuve très classique. En effet, si μ et ν sont deux minimiseurs, alors pour tout $t \in [0,1]$ on peut définir leur « interpolation linéaire »

$$\rho_t = (1-t)\mu + t\nu. \qquad (8)$$

Si F est strictement convexe, la fonction $t \mapsto F(\rho_t)$ l'est également (sauf si $\mu = \nu$), et présente un minimum strict pour un certain $t \in]0,1[$ – ce qui contredit l'hypothèse de minimalité pour μ et ν. Bien évidemment, cette procédure classique ne s'appliquait pas aux exemples considérés par McCann (cela aurait été trop facile !!). Il eut cependant l'idée de substituer à l'interpolation (8) une autre recette, basée sur le transport optimal. Définissons donc, pour tout t (« temps ») compris entre 0 et 1,

$$\rho_t = [(1-t)\mathrm{Id} + t\nabla\varphi]\#\mu. \qquad (9)$$

Bien sûr, $\rho_0 = \mu$, $\rho_1 = \nu$.

Les deux exemples de la figure ci-dessous montreront combien cette procédure est qualitativement différente de la procédure plus classique d'interpolation linéaire. Dans chacun des deux cas envisagés, la densité de départ est représentée en vert, la densité d'arrivée en rouge, la densité interpolée en jaune. Les deux figures du haut considèrent une interpolation par transport de mesure (interpolation par déplacement), les deux figures du bas une interpolation linéaire. Noter que, dans les deux cas, non seulement la forme générale, mais aussi les supports des mesures interpolées sont différents.

Le théorème suivant, dû à McCann, illustre l'intérêt de cette procédure :

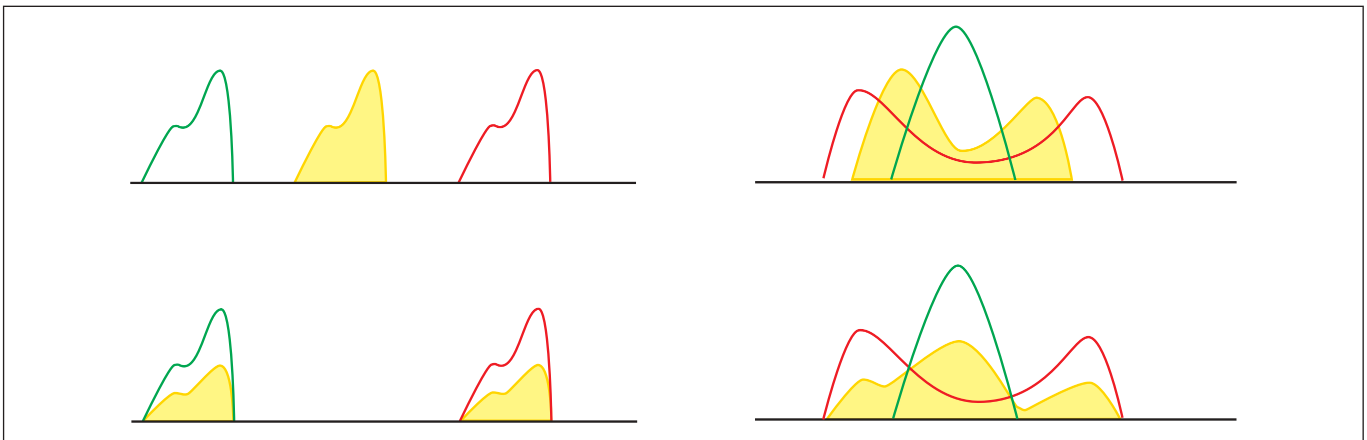

Figure - *Différence entre l'interpolation « linéaire » et l'interpolation par transport.*

Théorème 3. *Soit $W(z)$ une fonction convexe paire sur $\mathbb{R}^d$ et soit $\gamma \geqslant 1 - 1/d$, $\gamma \neq 1$. Alors la fonctionnelle*

$$\rho \longmapsto \frac{1}{\gamma - 1} \int \rho(x)^\gamma \, dx + \frac{1}{2} \int W(x-y)\, \rho(x)\, \rho(y)\, dx\, dy \qquad (10)$$

est convexe le long de l'interpolation (9).

La formule (10) contient deux termes ; on peut penser au premier comme à un terme d'« énergie interne » et au second comme à un terme d'« énergie potentielle ». L'hypothèse de convexité du potentiel d'interaction W peut sembler mystérieuse, mais elle a trouvé des exemples d'application, de manière quelque peu surprenante, dans l'étude des milieux granulaires.

Divers résultats d'unicité résultent de ce théorème de « convexité par déplacement ». Des études plus poussées ont mis en évidence l'intérêt de ces notions dans l'étude de certaines équations issues de la mécanique statistique. Otto et ses collaborateurs ont ainsi montré qu'il existait un lien naturel entre l'équation de la chaleur, le transport optimal et la fonctionnelle d'entropie de Boltzmann, $S(\rho) = -\int \rho \log \rho$. Voici une formulation imagée de leur découverte. Soit $\rho(t,x)$ la densité de particules diffusant librement, de sorte que ρ suit l'équation de la chaleur. Pour tout temps $t \geqslant 0$, $\rho(t) = \rho(t,\cdot)$ est une densité de probabilité. Entre deux instants infiniment proches t et $t + dt$, elle évolue de manière à **maximiser l'entropie**, tout en tenant compte d'une **pénalité** pour limiter les déplacements trop importants de particules : la quantité à maximiser est

$$S(\rho(t+dt)) - \frac{1}{2(dt)^2} T_2\big(\rho(t), \rho(t+dt)\big),$$

où $T_2(\mu,\nu)$ désigne le coût de transport optimal entre μ et ν, pour le coût $c(x,y) = |x-y|^2$. Cet énoncé peut se traduire de manière compacte grâce au concept, bien connu en physique et en mathématiques, de **flot gradient**.

APPLICATIONS A LA THÉORIE

C'est une banalité que de dire que les mathématiques « pures » trouvent sans cesse des applications ; mais de manière également universelle en mathématiques, des concepts dont le développement a été motivé par des problèmes concrets trouvent des applications dans des questions d'apparence purement théorique. Le transport optimal n'échappe pas à la règle : il a trouvé des applications spectaculaires dans le domaine des inégalités fonctionnelles à caractère géométrique. Considérons ainsi la célèbre (pour les spécialistes !) inégalité de Young optimale,

$$\|f * g\|_{L^r(\mathbb{R}^d)} \leqslant \left(\frac{C_p C_q}{C_r}\right)^d \|f\|_{L^p(\mathbb{R}^d)} \|g\|_{L^q(\mathbb{R}^d)},$$

$C_p^2 = p^{1/p}/(p')^{1/p'}$, $p' = p/(p-1)$, et $1/p + 1/q = 1 + 1/r$. Il en existe plusieurs démonstrations ; cependant, la plus spectaculaire est certainement celle qui a été mise au point par Barthe, reposant sur le transport optimal. Comme l'ont remarqué Cordero-Erausquin, Nazaret et l'auteur, il est également possible d'utiliser le transport de mesure pour démontrer avec peu d'efforts la non moins célèbre inégalité de Sobolev optimale,

$$\|f\|_{L^{p^\star}(\mathbb{R}^d)} \leqslant S_d(p) \|\nabla f\|_{L^p(\mathbb{R}^d)},$$

où $p^\star = dp/(d-p)$, $1 < p < d$, et $S_d(p)$ est la constante optimale. Bien sûr, le transport de mesure n'apparaît ici que comme un *outil*. Il en est de même pour le remarquable théorème suivant, dû à Caffarelli :

Théorème 4. *Soient F, G, H, J, K des fonctions continues positives sur $\mathbb{R}_+$, H et J étant croissantes, et soit $k \in \mathbb{R}$. Soit ρ une mesure de probabilité absolument continue, et soit $\lambda(\rho)$ la plus grande constante λ admissible dans l'énoncé suivant : si $\int K(h(x))\, d\rho(x) = 0$, alors*

$$F\left(\int_{\mathbb{R}^d} G(|h(x)|)\, d\rho(x)\right) \leqslant \frac{1}{\lambda} H\left(\int_{\mathbb{R}^d} J(|\nabla h(x)|)\, d\rho(x)\right). \qquad (11)$$

Soit maintenant γ la densité gaussienne standard : $\gamma(x) = e^{-|x|^2}/(2\pi)^{d/2}$; et soit v une fonction convexe, normalisée de sorte que $e^{-v}\gamma$ soit encore d'intégrale 1. Alors

$$\lambda(e^{-v}\gamma) \geqslant \lambda(\gamma).$$

En termes compacts : une inégalité fonctionnelle de la forme extrêmement générale (11) (qui inclut nombre d'inégalités de type isopérimétriques, trou spectral, etc.) ne peut être qu'améliorée par perturbation log concave d'une gaussienne. Encore une illustration des propriétés « miraculeuses » des fonctions gaussiennes...

POUR EN SAVOIR PLUS

Ambrosio (L.), Brenier (Y.), Buttazzo (G.), Caffarelli (L.), Evans (L.-C.), Pratelli (A.) et Villani (C.), « Optimal transportation and applications », *Actes de l'école d'été CIME tenue à Martina Franca, septembre 2001*. Lecture Notes in Mathematics 1813, *Springer-Verlag*, 2002.

Villani (C.), « Topics in optimal transportation », *Graduate Series in Mathematics*, *American Mathematical Society*, 2003.