

# Une introduction aux sciences de l'imagerie



Gabriel Peyré  
CNRS & DMA  
École Normale Supérieure  
[gabriel.peyre@ens.fr](mailto:gabriel.peyre@ens.fr)  
<https://mathematical-tours.github.io>

March 12, 2019



# Chapter 1

## Le Transport Optimal et ses Applications

### 1.1 Le Transport Optimal de Monge

Gaspard Monge, en plus d'être un grand mathématicien, a participé activement à la révolution Française, et a créé l'École Polytechnique ainsi que l'École Normale Supérieure. Motivé par des applications militaires, il a formulé en 1781 le problème du transport optimal [4]. Il s'est posé la question du calcul de la façon la plus économique de transporter de la terre entre deux endroits pour faire des remblais. Dans son texte original, il a fait l'hypothèse que le coût du déplacement d'une unité de masse est égal à la distance parcourue, mais l'on peut utiliser n'importe quel coût adapté au problème à résoudre.

**Le problème de Monge.** Pour illustrer le problème et sa formulation mathématique, intéressons-nous à la façon optimale de distribuer les croissants depuis les boulangeries vers le café, le matin dans Paris. Pour simplifier, nous allons supposer qu'il y a uniquement 6 boulangeries et cafés, que l'on peut voir à la figure 1.1 (les boulangeries sont en rouge et les cafés en bleu). Le coût à minimiser est le temps total des trajets, et l'on note  $C_{i,j}$  le temps entre la boulangerie  $i \in \{1, \dots, 6\}$  et le café  $j \in \{1, \dots, 6\}$ . Par exemple, on a  $C_{3,4} = 10$ , ce qui signifie qu'il y a 10 minutes de trajet entre la boulangerie numéro 3 et le café numéro 4.

Afin de satisfaire la contrainte d'approvisionnement (que l'on appelle aussi la conservation de la masse), il faut que chaque boulangerie soit connectée à un et un seul café. Comme il y a le même nombre de boulangeries que de cafés, ceci implique que chaque café est également connecté à une

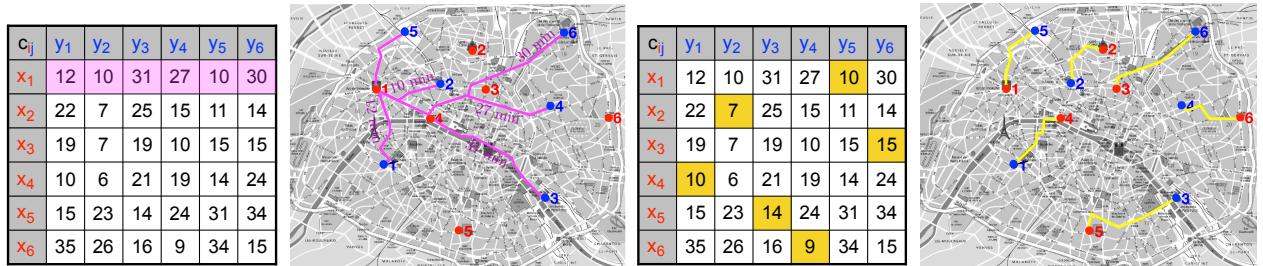


Figure 1.1: Matrice de coût et connexions associées. Gauche: une ligne de la matrice coût. Droite: un choix de permutation valide.

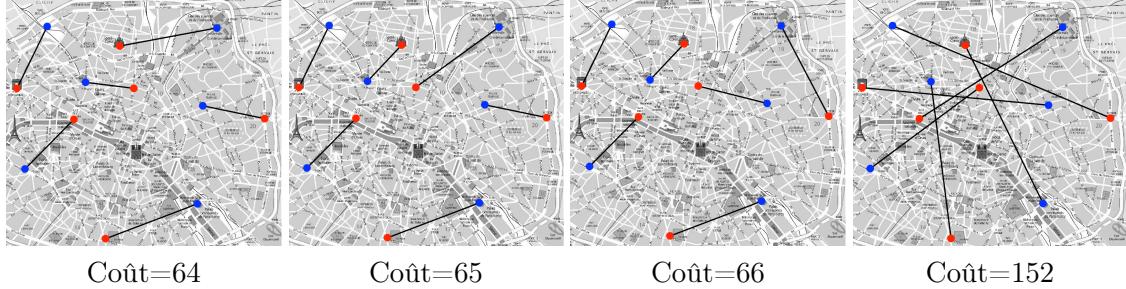


Figure 1.2: Choix de permutation avec différent coûts.

et une seule boulangerie. On va noter

$$\sigma : \textcolor{red}{i} \in \{1, \dots, 6\} \longmapsto \textcolor{blue}{j} \in \{1, \dots, 6\}$$

un tel choix de connexion. La figure 1.1 illustre au centre et à droite l'exemple

$$\sigma(1) = \textcolor{red}{5}, \sigma(2) = \textcolor{blue}{2}, \sigma(3) = \textcolor{red}{6}, \sigma(4) = \textcolor{blue}{1}, \sigma(5) = \textcolor{red}{3}, \sigma(6) = \textcolor{blue}{4}. \quad (1.1)$$

La contrainte de conservation de masse signifie que  $\sigma$  est une bijection de l'ensemble  $\{1, \dots, 6\}$  dans lui-même. On dit ainsi que  $\sigma$  est une permutation.

Le coût de transport associé à une telle bijection est la somme des coût  $C_{i,\sigma(i)}$  sélectionnés par la permutation  $\sigma$ , c'est-à-dire

$$\text{Coût}(\sigma) \stackrel{\text{def.}}{=} C_{1,\sigma(1)} + C_{2,\sigma(2)} + C_{3,\sigma(3)} + C_{4,\sigma(4)} + C_{5,\sigma(5)} + C_{6,\sigma(6)}. \quad (1.2)$$

Par exemple, pour la bijection (1.1) montrée à la figure 1.1, on obtient comme coût

$$C_{1,5} + C_{2,2} + C_{3,6} + C_{4,1} + C_{5,3} + C_{6,4} = 10 + 7 + 15 + 10 + 14 + 9 = 65.$$

Le problème de Monge consiste à chercher la permutation  $\sigma$  qui a le coût minimum, c'est-à-dire résoudre le problème d'optimisation

$$\min_{\sigma \in \text{Perm}(\{1, \dots, 6\})} \text{Coût}(\sigma), \quad (1.3)$$

où l'on a noté  $\text{Perm}(\{1, \dots, 6\})$  l'ensemble des permutations de l'ensemble  $\{1, \dots, 6\}$ .

La figure 1.2 montre que la permutation (1.1) n'est pas la meilleure : il existe par exemple une autre permutation qui a un coût de 64. Mais est-ce la meilleure ? Il se trouve que oui, on peut en effet tester sur un ordinateur toutes les permutations de  $\{1, \dots, 6\}$  et calculer leur coût. Combien y a-t-il de permutations au total ? Pour effectuer ce dénombrement, on voit qu'il y a 6 choix d'affectation possible de 1 à  $\sigma(1) \in \{1, \dots, 6\}$ , puis 5 choix possibles pour affecter 2 à  $\sigma(2) \in \{1, \dots, 6\} - \{\sigma(1)\}$ , et ainsi de suite. Le nombre total de possibilité est donc factoriel(6), que l'on note  $6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$ . Si on considère un nombre  $n$  de boulangerie, alors le nombre de permutations à tester pour trouver la meilleure est  $n! = n \times (n-1) \times \dots \times 2 \times 1$ . Ce nombre croît extrêmement vite avec  $n$ , par exemple  $70! \approx 1,198 \times 10^{100}$ , à comparer avec les  $10^{11}$  neurones dans le cerveau et les  $10^{79}$  atomes dans l'univers. Cette stratégie de recherche exhaustive n'est donc possible que pour de toute petites valeurs de  $n$ .

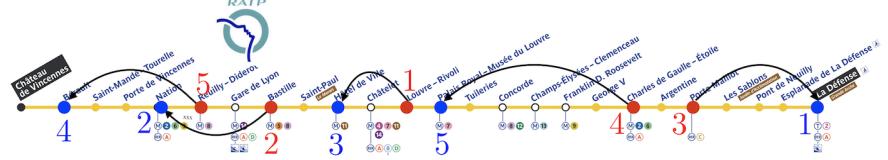


Figure 1.3: Le transport optimal en 1D le long d'une ligne de métro. La bijection optimale est  $\sigma : (1, 2, 3, 4, 5) \mapsto (3, 2, 1, 5, 4)$ .

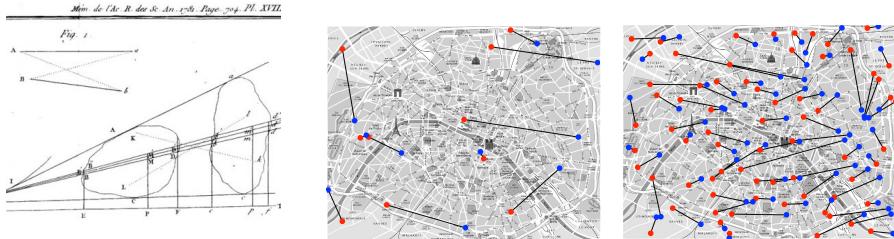


Figure 1.4: Gauche: extrait de l'article de Monge [4]. Droite: le transport optimal en 2D pour un cout Euclidien.

**En 1D et 2D.** La section suivante explique comment des avancées mathématiques ont permis de développer des techniques efficaces pour calculer un transport optimal  $\sigma$  même pour de grandes valeurs de  $n$ . Mais il aura fallu attendre près de 200 ans pour y arriver. Dans certains cas simples, on peut cependant calculer le transport optimal de façon simple. Le cas le plus élémentaire est lorsque les points à apparié sont le long d'un axe 1D, par exemple si les cafés et les boulangeries sont situés le long d'une ligne de métro. Il faut également que le coût  $C_{i,\sigma(i)}$  soit la distance le long de cet axe (par exemple le temps de trajet en métro entre les stations). Dans ce cas, il suffit de classer les index  $i$  et  $j$  par ordre croissant (donc de gauche à droite le long de la ligne de métro), et d'apparier le premier index  $i$  au premier index  $j$  ensemble, puis les deuxièmes index, etc. Ce procédé est illustré à la figure 1.3. Le temps de calcul nécessaire pour calculer le transport optimal en métro est donc le temps nécessaire pour classer les index. L'algorithme le plus simple pour effectuer un classement est celui utilisé habituellement pour trier un jeu de  $n$  cartes : il s'agit du tri par insertion (on insère itérativement chaque carte à sa place par rapport aux cartes déjà classées). Il effectue  $n(n-1)/2$  comparaisons. Pour  $n = 70$ , cela nécessite donc seulement 21415 opérations, ce qui rend la méthode utilisable, au contraire de la recherche exhaustive de toutes les  $n!$  permutations. On existe des algorithmes encore plus rapides (par exemple le tri par fusion), qui effectuent de l'ordre de  $n \log(n)$  opérations, et donc pour  $n = 70$ , de telles méthodes nécessitent moins de 1000 opérations.

Malheureusement, il n'est plus possible d'utiliser cette technique de classement dans des cas plus généraux. Pour des points dimension 2, si on prend comme coût  $C_{i,\sigma(i)}$  la distance Euclidienne (la distance en vol d'oiseau) entre les points, alors Gaspard Monge a montré dans son papier original (voir la figure 1.4, à gauche) qu'un transport optimal ne peut pas contenir de croisements. En effet, comme le montre la figure 1.4 (à droite), si l'on trace tous les segments entre les points  $i \mapsto j = \sigma(i)$  que l'on relie par la bijection définie par  $\sigma$ , ceux-ci ne se croisent jamais.

Cette observation géométrique n'est cependant pas suffisante pour calculer un transport optimal en 2-D : il existe en effet beaucoup de permutations  $\sigma$  tels que les segments associés ne se croisent pas. Il va falloir analyser de façon plus fine la structure des permutations optimales afin de pouvoir les

calculer de façon efficace. Nous allons maintenant voir comment Leonid Kantorovitch a reformulé le problème de Monge afin d'y parvenir.

## 1.2 Le Transport Optimal de Kantorovitch

Leonid Kantorovitch est un mathématicien et un économiste soviétique qui a révolutionné la théorie du transport optimal pendant les années 40. Ces recherches sont issues de considération pratique qui l'on occupé avant et après la seconde guerre mondiale, durant laquelle a joué un rôle important pour l'allocation optimale de ressources, en particulier durant le siège de Léningrad. Il a par la même occasion, participé au développement de l'optimisation moderne, qui a eu un impact énorme dans de très nombreux domaines appliqués. Il a ainsi obtenu en 1975 le prix Nobel d'économie, car les premières applications (mais certainement pas les seules !) de sa théorie ont été dans ce domaine.

**Le problème de Kantorovitch.** L'idée centrale de Kantorovitch est de relaxer le problème de Monge en remplaçant l'ensemble des permutations par un ensemble plus grand mais plus simple. Tout d'abord on remarque que l'on peut représenter une permutation  $\sigma$  à l'aide d'une matrice de permutation  $P$  qui est une matrice binaire (remplie de 0 et de 1) de taille  $n \times n$  telle que  $P_{i,j} = 0$  sauf si  $j = \sigma(i)$  auquel cas  $P_{i,\sigma(i)} = 1$ . Par exemple, pour  $n = 3$  points, les permutations  $(1, 2, 3) \mapsto (1, 2, 3)$  (l'identité),  $(1, 2, 3) \mapsto (3, 2, 1)$  et  $(1, 2, 3) \mapsto (2, 1, 3)$  sont représentées par les matrices de taille  $3 \times 3$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Dans la suite, on note  $\mathcal{P}_n$  l'ensemble de  $n!$  matrices de permutation de taille  $n \times n$ .

Comme la matrice est binaire, avec seulement  $n$  éléments non-nuls égaux à 1, on peut remplacer la somme de  $n$  termes qui apparaît dans le coût  $\text{Coût}(\sigma)$  (1.2) du problème de Monge par une somme sur l'ensemble des  $n \times n$  index  $(i, j)$ , c'est-à-dire que si  $P$  est la matrice de permutation associée à  $\sigma$ , on a

$$\text{Coût}(\sigma) = \sum_{i=1}^n \sum_{j=1}^n P_{i,j} C_{i,j}.$$

On peut ainsi remplacer le problème de Monge (1.3) par le problème équivalent

$$\min_{P \in \mathcal{P}_n} \sum_{i=1}^n \sum_{j=1}^n P_{i,j} C_{i,j}. \quad (1.4)$$

Le génie de Kantorovitch a été de remarquer que l'on peut remplacer l'ensemble discret  $\mathcal{P}_n$  (c'est-à-dire composé d'un ensemble fini, mais très grand, de  $n!$  matrice) par un ensemble « continu » (donc en particulier infini) mais plus simple. On remarque en effet que les matrices de permutation de  $\mathcal{P}_n$  sont exactement les matrices qui ont un et un seul 1 le long de chaque ligne et de chaque colonne. Ceci peut aussi s'exprimer comme le fait qu'une matrice de permutation est une matrice binaire dont la somme de chaque ligne et de chaque colonne vaut 1, c'est-à-dire

$$\mathcal{P}_n = \left\{ P \in \{0, 1\}^{n \times n} ; \forall i, \sum_j P_{i,j} = 1, \forall j, \sum_i P_{i,j} = 1 \right\}.$$

Ce qui rend cet ensemble très compliqué, c'est la contrainte binaire, c'est-à-dire que ces matrices sont contraintes à être dans  $\{0, 1\}^{n \times n}$ . Kantorovitch propose alors de « relaxer » cette contrainte en supposant simplement que les entrées de  $P$  soient entre 0 et 1. Ceci définit un ensemble plus grand, l'ensemble des matrices bi-stochastiques

$$\mathcal{B}_n \stackrel{\text{def.}}{=} \left\{ P \in [0, 1]^{n \times n} ; \forall \textcolor{red}{i}, \sum_{\textcolor{blue}{j}} P_{\textcolor{red}{i}, \textcolor{blue}{j}} = 1, \forall \textcolor{blue}{j}, \sum_{\textcolor{red}{i}} P_{\textcolor{red}{i}, \textcolor{blue}{j}} = 1 \right\}. \quad (1.5)$$

Le problème de Kantorovitch correspond à effectuer ce remplacement dans (??), et à résoudre le problème

$$\min_{P \in \mathcal{B}_n} \sum_{i=1}^n \sum_{j=1}^n P_{\textcolor{red}{i}, \textcolor{blue}{j}} C_{\textcolor{red}{i}, \textcolor{blue}{j}}. \quad (1.6)$$

L'immense avantage du problème de Kantorovitch (1.6) par rapport à celui de Monge (??) est que l'ensemble des matrices bistochastiques est convexe, c'est-à-dire que si l'on considère deux matrices bistochastiques  $P, Q \in \mathcal{B}_n$ , alors leur moyenne  $\frac{P+Q}{2} \in \mathcal{B}_n$  est encore bistochastique. Ceci n'est pas vrai pour les matrices de permutation, puisqu'en général la moyenne de deux matrices binaires n'est pas binaire (sauf si bien sûr si  $P = Q$ ). Cette nouvelle formulation a pu ainsi bénéficier d'une deuxième révolution initiée par George Dantzig [3], qui à la même époque, a proposé l'algorithme du simplex, qui permet de résoudre efficacement les problèmes de programmation linéaire, dont (1.6) est un cas particulier. Dans le cas du problème de Kantorovitch, il existe en effet un algorithme de simplex qui a une complexité de l'ordre de  $n^3$  opérations, ce qui permet de faire des calculs pour de grands  $n$ , de l'ordre de plusieurs milliers.

**L'équivalence Monge–Kantorovitch.** L'ensemble des matrices bistochastiques est plus grand que celui des matrices de permutations,  $\mathcal{P}_n \subset \mathcal{B}_n$ , de sorte que l'on a l'inégalité

$$\min_{P \in \mathcal{B}_n} \sum_{i=1}^n \sum_{j=1}^n P_{\textcolor{red}{i}, \textcolor{blue}{j}} C_{\textcolor{red}{i}, \textcolor{blue}{j}} \leq \min_{P \in \mathcal{P}_n} \sum_{i=1}^n \sum_{j=1}^n P_{\textcolor{red}{i}, \textcolor{blue}{j}} C_{\textcolor{red}{i}, \textcolor{blue}{j}} \quad (1.7)$$

entre les problèmes de Kantorovitch et de Monge. Mais de façon à première vue surprenante, un théorème fondamental du à George Birkhoff et à John von Neumann [1, 7] assure qu'en fait il y a égalité entre ces deux problèmes. En effet, ce théorème montre qu'il existe toujours une matrice solution du problème de Kantorovitch qui est binaire, de sorte qu'elle est aussi solution du problème de Monge (attention cependant, en général il n'y a pas unicité des solutions de ces problèmes). La conjonction de deux avancées spectaculaires, dues à Kantorovitch et à Dantzig, a permis de rendre le transport optimal applicables à des problèmes de grande taille, puisque l'algorithme du simplexe permet de résoudre en pratique ces problèmes.

**Le cas pondéré.** Outre son intérêt pratique, la formulation de Kantorovitch a aussi permis de généraliser le problème initial de Monge, en donnant le bon cadre pour le formaliser et l'étudier mathématiquement. En effet, le problème de Monge est très limité. Que se passe-t-il par exemple si il n'y a pas le même nombre  $n$  de café et  $m$  de boulangerie ? Le problème initial (1.3) n'a pas de solution, car on ne peut pas mettre en bijection deux ensembles de tailles différentes. Le bon concept n'est pas le nombre de boulangeries et de cafés, mais plutôt les distributions ( $a_1, \dots, a_n$ )

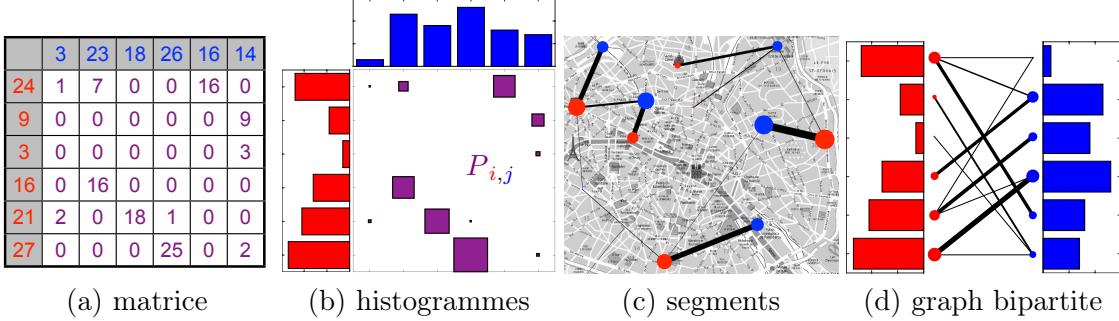


Figure 1.5: Différentes façons de représenter une matrice de couplage  $P \in \mathcal{B}(a, b)$ : (a) un tableau de nombre dont les lignes et colonnes ont des sommes prescrites ; (b) un histogramme bi-dimensionnel dont la taille de carré est proportionnelle à  $P_{i,j}$  ; (c) un ensemble de segments dont la largeur est proportionnelle est  $P_{i,j}$ .

de production (associées au boulangeries) et les distributions ( $b_1, \dots, b_m$ ) de consommation des cafés. Dans le cas initialement considéré, où  $n = m$ , toutes les quantités  $a_i$  et  $b_j$  sont égales à 1. Mais dans pour un problème « réalistes », ces quantités sont quelconques. Ces quantités doivent être positives, et vérifier

$$a_1 + \dots + a_n = b_1 + \dots + b_m,$$

de sorte qu'il ait autant de production que de consommation. La construction de Kantorovitch s'adapte naturellement à ce cas de distributions générales, en remplaçant les matrices bistochastique (1.5) par des matrices de « couplage » qui satisfont la contrainte de conservation de la masse

$$\mathcal{B}(a, b) \stackrel{\text{def.}}{=} \left\{ P \in [0, 1]^{n \times n} ; \forall i, \sum_j P_{i,j} = a_i, \forall j, \sum_i P_{i,j} = b_j \right\}.$$

Dans le cas initial où  $n = m$  et  $a_i = b_j = 1$ , alors  $\mathcal{B}(a, b) = \mathcal{B}_n$  et l'on retrouve des matrices bistochastiques. Dans le cas général, à chaque fois qu'une entrée  $P_{i,j}$  est non-nulle, ceci signifie que l'on transfert de la « masse » (ici une certaine quantité de croissants) entre  $i$  et  $j$ . Comme le montre la figure 1.5, on peut visualiser de différente façon une telle matrice  $P$  couplant deux distributions  $(a, b)$ .

Le problème de Kantorovitch qui généralise (1.6) s'écrit alors

$$\min_{P \in \mathcal{B}(a, b)} \sum_{i=1}^n \sum_{j=1}^n P_{i,j} C_{i,j}$$

ce qui signifie que l'on doit payer un coût  $C_{i,j}$  à chaque fois que l'on transfert une unité de masse entre  $i$  et  $j$ . Tout comme le problème original (1.6), on peut le résoudre de façon efficace avec l'algorithme du simplex.

### 1.3 Les applications

Bien que les motivations initiales de Monge et Kantorovitch étaient respectivement militaires et économiques, le transport optimal a trouvé d'innombrables applications, à la fois théoriques et plus

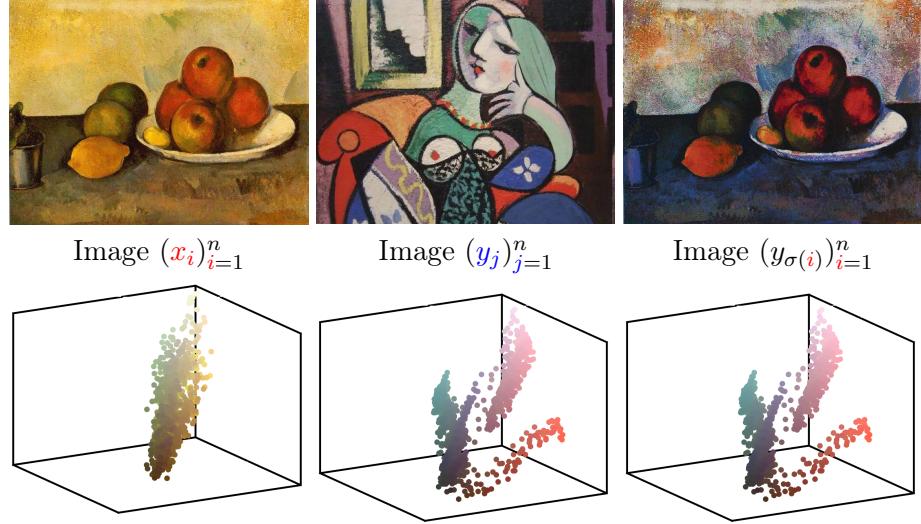


Figure 1.6: Exemple de transfert de palettes de couleurs à l'aide du transport optimal. Haut: les pixels sont sur la grille d'affichage pour former une image couleurs. Bas: les pixels sont placés à leurs positions dans  $\mathbb{R}^3$  pour former un nuage de point.

concrètes. Sur le plan mathématique, on peut considérer des distributions « continues » de masses, en quelque sorte la limite quand le nombre de point  $n$  tend vers l'infini. Ceci permet de définir le problème de transport entre des mesures de probabilité quelconque. Ce point de vu théorique est extrêmement fructueux, et c'est le mathématicien Yann Brenier qui a le premier montré l'équivalence dans le cadre continu des formulations de Monge et de Kantovotich [2]. Ces travaux pionniers ont montré la connexion entre le problème de transport et les équations aux dérivées partielles, et a débouché, entre autre, sur les médailles Fields de Cédric Villani (2010) et Alessio Figalli (2018).

Le transport optimal est depuis peu au coeur de problématiques plus appliquées en sciences des données, en particulier pour résoudre des problèmes en traitement d'image et en apprentissage statistique. La première idée, la plus immédiate, est d'utiliser la bijection  $\sigma$  afin de transformées des données, par exemple des images. Dans ce cas, on considère les pixels  $(x_i)_{i=1}^n$  et  $(y_j)_{j=1}^n$  de deux images couleurs. Chaque pixel  $x_i, y_j \in \mathbb{R}^3$  est un vecteur de dimension 3, qui représente les intensités de chacune des trois couleurs élémentaires, rouge, vert et bleu. Afin de changer les couleurs de la première image, et lui imposer la palette de la deuxième image, on calcule le transport  $\sigma$  pour la matrice de coût  $C_i = \|x_i - y_j\|^2$ , c'est-à-dire le carré de la distance Euclidienne entre les pixels. L'image avec les couleurs modifiées est  $(y_{\sigma(i)})_{i=1}^n$ , c'est à dire que l'on remplace dans la première image le pixel  $x_i$  par le pixel  $y_{\sigma(i)}$ . Cette image ressemble à la première, mais a la palette de couleurs de la deuxième image. La figure 1.6 illustre ce procédé pour imposer la palette de couleur de Picasso à un tableau de Cézanne.

On peut également utiliser le transport optimal pour des problèmes plus difficile, en n'utilisant de façon indirecte la bijection  $\sigma$  ou bien la matrice de couplage optimal  $P \in \mathcal{B}(a, b)$ . L'idée centrale est que la quantité associée à un couplage optimal  $P$

$$W(a, b) \stackrel{\text{def.}}{=} \sum_{i,j} P_{i,j} C_{i,j}$$

définit en quelque sorte l'effort nécessaire pour déplacer la masse de la distribution  $a$  vers la dis-

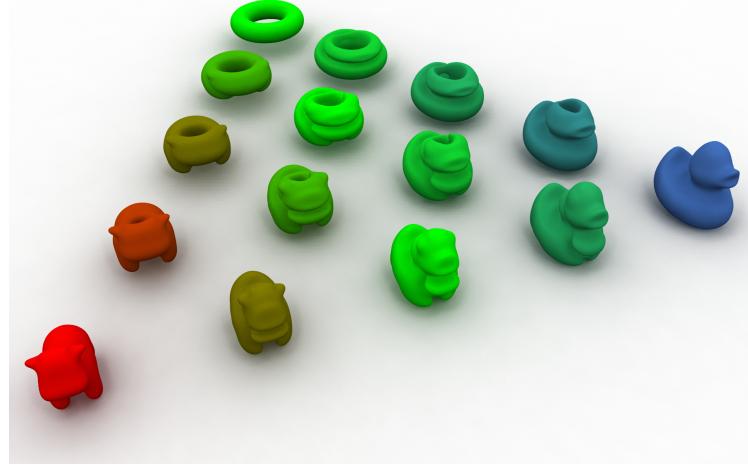


Figure 1.7: Exemple d’interpolation barycentrique entre des formes 3D, obtenu en minimisant (1.8).

tribution  $b$ . Elle permet donc de quantifier combien ces deux distributions sont « proches ». Par exemple, si  $C_{ij} = \|\mathbf{x}_i - \mathbf{y}_j\|^p$  est associé à la distance Euclidienne entre des points, pour  $p \geq 1$ , alors la quantité  $W(a, b)^{1/p}$  est une distance entre les distributions, en particulier elle vérifie  $W(a, b) = 0$  si et seulement si  $a = b$ , et elle vérifie l’inégalité triangulaires. Ces propriétés sont très importantes pour permettre d’appliquer le transport à des problèmes pratiques.

Un exemple typique d’application consiste à calculer des barycentres entre des distributions. La figure 1.7 montre un exemple où l’on considère trois distributions  $a^1, a^2, a^3$  (montrées aux trois sommets du triangle) qui sont des distributions uniformes de masse à l’intérieur de formes 3D (c’est-à-dire que la masse  $a_i^j$  associé à chaque point est 0 à l’extérieur de la première forme, et constant à l’intérieur). On calcule un barycentre pondéré  $b$  de ces trois distributions en minimisant une somme pondérée de distance de transport optimal

$$\min_b \lambda_1 W(a^1, b) + \lambda_2 W(a^2, b) + \lambda_3 W(a^3, b). \quad (1.8)$$

En modifiant les poids  $(\lambda^1, \lambda^2, \lambda^3)$ , qui sont positifs et tels que  $\lambda^1 + \lambda^2 + \lambda^3 = 1$ , on modifie la forme obtenue en se déplaçant à l’intérieur d’un triangle de transport optimal. On peut utiliser cette distance  $W$  pour bien d’autres applications où l’on doit comparer des distributions de probabilité. C’est le cas par exemple en apprentissage automatique, par exemple pour comparer des textes à l’aide des distributions des mots qui les compose. La figure 1.8 illustre les histogrammes d’apparition des mots pour deux textes, où la taille des lettres du mot  $i$  est proportionnel à la masse  $a_i$ . Une question difficile dans ce cas est de savoir quel matrice de coût  $C_{ij}$  utiliser entre deux mots  $(i, j)$ , il s’agit d’un travail de linguistique (caractériser la proximité sémantique entre des mots du langage), que l’on peut chercher à résoudre en même temps que le transport optimal.

## Conclusions

Le transport optimal a connu de nombreuses révolutions, et sous l’impulsion de mathématiciens tels que Monge, Kantorovitch, Dantzig et Brenier, il est progressivement devenu un outil



Figure 1.8: Exemple d’histogrammes de mots.

théorique et numérique fondamental. Il est maintenant au cœur de problème fondamentaux en science des données pour modéliser, résoudre numériquement et analyser théoriquement les problèmes de l’apprentissage automatiques. Les opportunités pour développer de nouvelles théories et des algorithmes performants sont immenses. Pour plus d’information sur les aspects théoriques du transport optimal, on pourra consulter les livres [? 6]. Les aspects numériques et applicatifs sont couverts dans le livre [5].



# Bibliography

- [1] Garrett Birkhoff. Tres observaciones sobre el álgebra lineal. *Universidad Nacional de Tucumán Revista Series A*, 5:147–151, 1946.
- [2] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.
- [3] George B Dantzig. Application of the simplex method to a transportation problem. *Activity Analysis of Production and Allocation*, 13:359–373, 1951.
- [4] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences*, pages 666–704, 1781.
- [5] Gabriel Peyré and Marco Cuturi. Computational optimal transport, 2018.
- [6] Filippo Santambrogio. *Optimal transport for applied mathematicians*. Birkhauser, 2015.
- [7] John Von Neumann. A certain zero-sum two-person game equivalent to the optimal assignment problem. *Contributions to the Theory of Games*, 2:5–12, 1953.