

Une introduction aux sciences des données

Gabriel Peyré

CNRS & DMA, École Normale Supérieure, PSL

CHAPITRE 4

Transport optimal

9 septembre 2026

Comparer deux distributions demande de préciser le coût du déplacement de la masse entre les positions. Le transport optimal traduit cette idée en un problème d'optimisation. À partir d'un petit problème d'affectation, nous introduisons les plans de transport, expliquons quand ils correspondent à des applications, puis utilisons les distances de transport pour comparer des couleurs, des formes et des textes.

4.1 Le problème de Monge

En 1781, Gaspard Monge formule le problème du déplacement de terre d'un lieu à un autre à moindre coût [7]. Son coût est la distance parcourue par chaque unité de masse. D'autres applications suggèrent d'autres coûts, comme le temps de trajet ou le carré de la distance euclidienne.

4.1.1 Un problème d'affectation

Considérons six boulangeries et six cafés à Paris. Chaque boulangerie fournit un lot de croissants, et chaque café demande un lot. Dans le modèle d'affectation, chaque lot va à un seul café, et chaque café reçoit un seul lot. Les deux exigences sont nécessaires : donner simplement une destination à chaque boulangerie permettrait à plusieurs d'entre elles de choisir le même café.

Notons C_{ij} le temps de trajet de la boulangerie i au café j . Dans l'exemple de la figure 4.1, on a $C_{34} = 10$. Une affectation admissible est une permutation σ de $\{1, \dots, 6\}$. Par exemple,

$$(\sigma(1), \dots, \sigma(6)) = (5, 2, 6, 1, 3, 4). \quad (4.1)$$

Son coût total est

$$\text{Coût}(\sigma) = \sum_{i=1}^6 C_{i,\sigma(i)} = 10 + 7 + 15 + 10 + 14 + 9 = 65. \quad (4.2)$$

On additionne ici les six temps de trajet. Ce modèle ne décrit ni la tournée d'un véhicule unique, ni la durée totale de livraisons effectuées en parallèle.

Pour n lots identiques de chaque côté, le problème discret de Monge s'écrit

$$\min_{\sigma \in \Sigma_n} \sum_{i=1}^n C_{i,\sigma(i)}, \quad (4.3)$$

où Σ_n est l'ensemble des permutations. Dans notre exemple, la valeur optimale est 64, atteinte par $(5, 6, 2, 1, 3, 4)$. L'examen exhaustif des $6! = 720$ possibilités permet de le vérifier. Il y a $n!$ affectations en général, et $70! \approx 1,198 \times 10^{100}$: l'énumération devient donc vite impraticable.

4.1.2 Trier en dimension un

Supposons que les positions sources $x_1 \leq \dots \leq x_n$ et cibles $y_1 \leq \dots \leq y_n$ soient sur une droite, avec le coût $C_{ij} = |x_i - y_j|$. Appairer les points dans l'ordre croissant est optimal. L'inégalité essentielle est

$$|x - y| + |x' - y'| \leq |x - y'| + |x' - y| \quad \text{si } x \leq x' \text{ et } y \leq y'.$$

C_{ij}	y_1	y_2	y_3	y_4	y_5	y_6
x_1	12	10	31	27	10	30
x_2	22	7	25	15	11	14
x_3	19	7	19	10	15	15
x_4	10	6	21	19	14	24
x_5	15	23	14	24	31	34
x_6	35	26	16	9	34	15

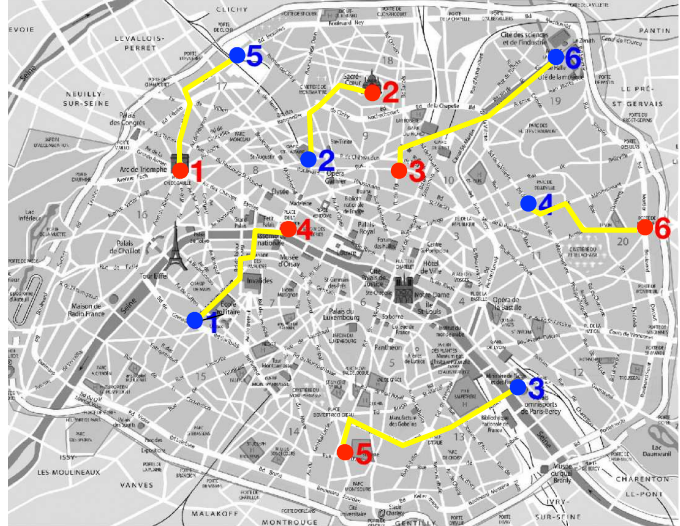


FIGURE 4.1. Matrice de coût et affectation des six boulangeries (rouge) aux six cafés (bleu). Les entrées surlignées donnent un temps de trajet cumulé de 65 minutes.

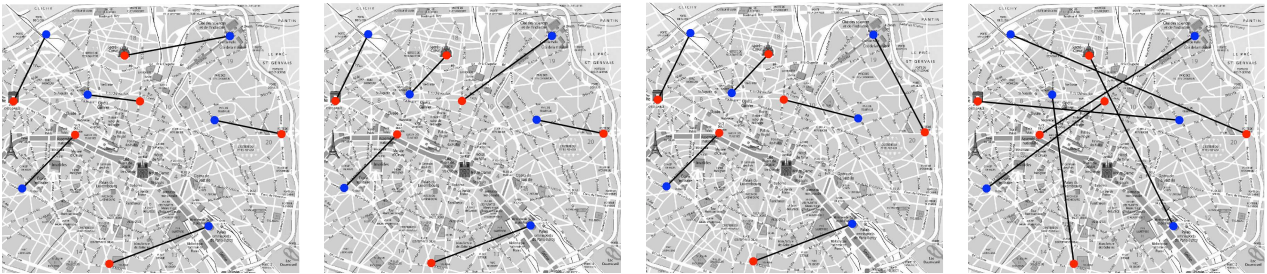


FIGURE 4.2. Quatre affectations de coûts 64, 65, 66 et 152, de gauche à droite.

On la vérifie en ordonnant les quatre points sur la droite. Lorsqu'une permutation contient une inversion, échanger les deux destinations n'augmente pas le coût. Éliminer successivement les inversions donne l'appariement trié. Une égalité est possible : cette affectation optimale n'est donc pas nécessairement unique.

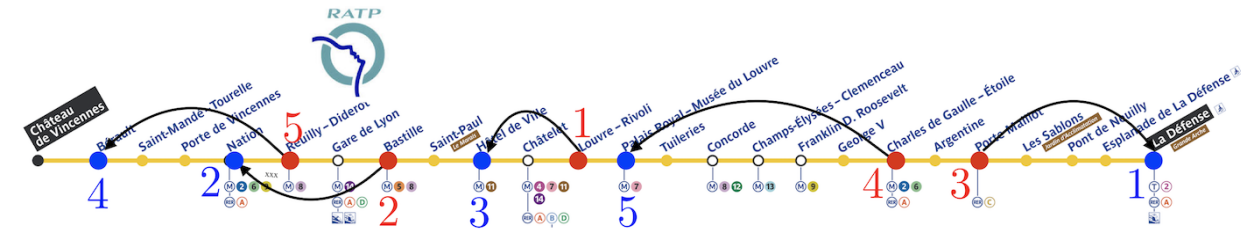


FIGURE 4.3. Appariement trié le long d'une ligne de métro, considérée comme un axe unidimensionnel. Pour les indices affichés, $\sigma : (1, 2, 3, 4, 5) \mapsto (3, 2, 1, 5, 4)$.

Le coût du calcul est ainsi celui du tri des deux listes. Le tri par insertion effectue au plus $n(n-1)/2$ comparaisons par liste, soit 2415 pour $n = 70$. Le tri fusion ramène ce coût à l'ordre de $n \log n$. Il s'agit du nombre de comparaisons pour le tri, et non d'une estimation de toutes les opérations d'une application de transport.

En dimension deux, aucun ordre unique ne résout le problème général. Pour le coût euclidien, deux segments d'un appariement optimal ne peuvent toutefois pas se croiser transversalement en leurs intérieurs. Si les segments $[x, y]$ et $[x', y']$ se rencontrent en z , l'inégalité triangulaire donne

$$|x - y'| + |x' - y| \leq |x - z| + |z - y'| + |x' - z| + |z - y| = |x - y| + |x' - y'|.$$

Pour un croisement transverse, l'inégalité est stricte : échanger les destinations diminue le coût. Des recouvrements colinéaires ou des extrémités communes sont des cas dégénérés qui ne donnent pas nécessairement

d'amélioration stricte. Cette observation ne suffit pas à calculer un optimum : de nombreux appariements sans croisement peuvent avoir des coûts différents.

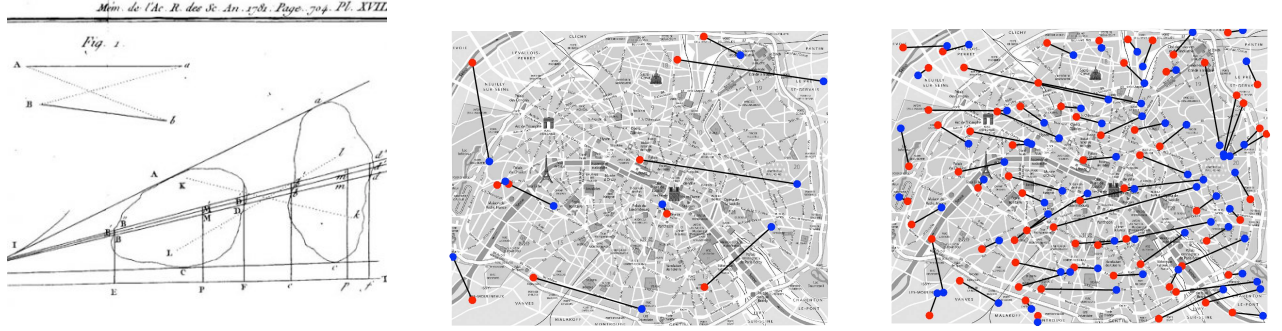


FIGURE 4.4. Une illustration de l'article original de Monge [7] et deux appariements euclidiens optimaux dans le plan.

4.2 La relaxation de Kantorovitch

La formulation de Kantorovitch autorise le partage de la masse entre plusieurs destinations [6]. Elle remplace les choix discrets par des quantités positives ou nulles soumises à des contraintes linéaires de conservation, faisant du transport optimal un problème de programmation linéaire.

4.2.1 Des permutations aux matrices

On représente une permutation par $P_{ij} = 1$ si $j = \sigma(i)$ et $P_{ij} = 0$ sinon. Par exemple, les permutations $(1, 2, 3)$, $(3, 2, 1)$ et $(2, 1, 3)$ correspondent à

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

Chaque ligne et chaque colonne a pour somme un. Si $\mathcal{P}_n$ désigne l'ensemble de ces matrices de permutation, le problème (4.3) équivaut à

$$\min_{P \in \mathcal{P}_n} \langle C, P \rangle, \quad \langle C, P \rangle = \sum_{i,j} C_{ij} P_{ij}. \quad (4.4)$$

Relâcher la contrainte binaire donne les matrices bistochastiques

$$\mathcal{B}_n = \left\{ P \in \mathbb{R}_+^{n \times n} : \sum_j P_{ij} = 1, \sum_i P_{ij} = 1 \right\}. \quad (4.5)$$

La positivité et les sommes des lignes impliquent déjà $P_{ij} \leq 1$. Le problème relaxé est

$$\min_{P \in \mathcal{B}_n} \langle C, P \rangle. \quad (4.6)$$

L'ensemble $\mathcal{B}_n$ est compact et convexe : pour $P, Q \in \mathcal{B}_n$ et $0 \leq t \leq 1$, la matrice $(1-t)P + tQ$ reste admissible. Un minimiseur existe car l'objectif est continu.

Le développement de la programmation linéaire, notamment de la méthode du simplexe de Dantzig [4], a rendu ces problèmes accessibles au calcul. Des algorithmes d'affectation spécialisés, comme certaines implémentations de la méthode hongroise, résolvent le problème dense $n \times n$ en $O(n^3)$ opérations. Cette borne ne doit pas être attribuée à la méthode générale du simplexe, dont le temps de calcul dans le pire cas n'est pas polynomial.

4.2.2 Pourquoi la relaxation est exacte pour des masses égales

Le théorème de Birkhoff–von Neumann affirme que toute matrice bistochastique est une combinaison convexe de matrices de permutation [2, 11] :

$$P = \sum_{k=1}^K \lambda_k P^{(k)}, \quad \lambda_k \geq 0, \quad \sum_k \lambda_k = 1.$$

Ainsi, $\langle C, P \rangle = \sum_k \lambda_k \langle C, P^{(k)} \rangle$ ne peut pas être inférieur au plus petit coût d'une permutation. Comme toute matrice de permutation est admissible pour la relaxation, on obtient

$$\min_{P \in \mathcal{B}_n} \langle C, P \rangle = \min_{P \in \mathcal{P}_n} \langle C, P \rangle. \quad (4.7)$$

Au moins un plan optimal est donc une permutation. Cela ne signifie pas que tout plan optimal soit une permutation : la moyenne de deux permutations optimales distinctes est également optimale. Si tous les coûts sont nuls, toute matrice bistochastique est optimale.

Pour certifier une valeur optimale, on peut chercher des nombres u_i, v_j tels que $u_i + v_j \leq C_{ij}$. Toute matrice $P \in \mathcal{B}_n$ vérifie alors

$$\langle C, P \rangle \geq \sum_{i,j} (u_i + v_j) P_{ij} = \sum_i u_i + \sum_j v_j.$$

Dans l'exemple parisien, $u = (2, 3, 3, 0, 5, 2)$ et $v = (10, 4, 9, 7, 8, 11)$ vérifient les 36 inégalités et la somme de leurs composantes vaut 64. Avec une affectation de coût 64, ils prouvent donc l'optimalité sans énumérer toutes les permutations.

4.2.3 Des nombres de points et des masses différents

Supposons maintenant que n boulangeries fournissent des quantités $a_i \geq 0$ et que m cafés demandent des quantités $b_j \geq 0$, avec la même masse totale

$$\sum_{i=1}^n a_i = \sum_{j=1}^m b_j = s.$$

Un plan de transport appartient à

$$\mathcal{B}(a, b) = \left\{ P \in \mathbb{R}_+^{n \times m} : \sum_j P_{ij} = a_i, \sum_i P_{ij} = b_j \right\}.$$

Pour $s > 0$, le plan $P_{ij} = a_i b_j / s$ prouve que cet ensemble est non vide. Pour $s = 0$, le seul plan est nul. Le problème de transport s'écrit

$$\min_{P \in \mathcal{B}(a, b)} \sum_{i=1}^n \sum_{j=1}^m C_{ij} P_{ij}. \quad (4.8)$$

Pour des coûts finis, la compacité assure encore l'existence d'un minimiseur. L'entrée P_{ij} est la quantité expédiée de i vers j , et C_{ij} son coût par unité de masse.

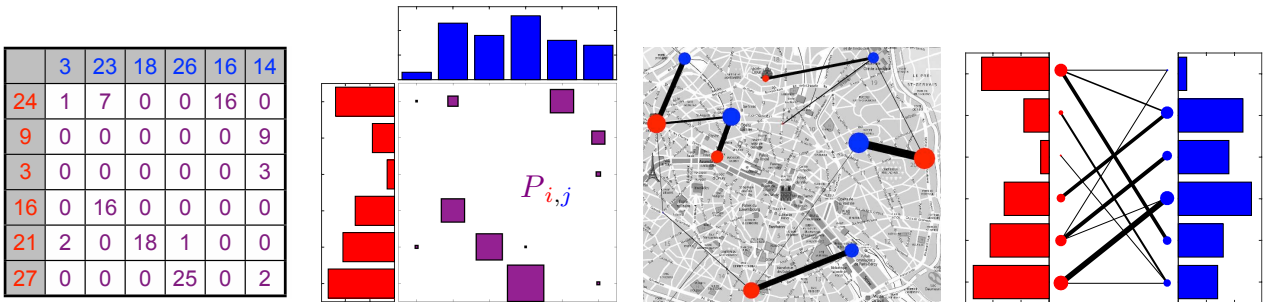


FIGURE 4.5. Quatre représentations du même couplage : ses entrées matricielles, leurs amplitudes représentées par des carrés, des segments pondérés dans l'espace et un graphe biparti. Les sommes des lignes et des colonnes donnent les masses prescrites a et b .

Le partage de la masse peut devenir nécessaire. Une boulangerie de masse 2 et deux cafés demandant chacun une masse 1 admettent pour seul plan $P = (1 \ 1)$. Une application qui envoie toute la masse source vers une seule cible ne peut pas réaliser ces marginales. Un plan de transport est donc plus général qu'une application de transport. Pour des masses égales $a_i = b_j = 1$, on retrouve $\mathcal{B}_n$; pour des masses de probabilité $a_i = b_j = 1/n$, les plans correspondants sont les P/n .

4.3 Mesures et distances de transport

Les mêmes idées s'appliquent aux mesures de probabilité, notamment aux distributions discrètes $\mu = \sum_i a_i \delta_{x_i}$ et $\nu = \sum_j b_j \delta_{y_j}$. Un couplage $\pi \in \Pi(\mu, \nu)$ est une mesure de probabilité sur les couples (x, y) dont les marginales sont μ et ν . Une application mesurable T induit un couplage $(\text{Id}, T)_\# \mu$ lorsque $T_\# \mu = \nu$; cette notation signifie qu'appliquer T à une variable aléatoire de loi μ donne une variable de loi ν .

Pour des mesures de probabilité sur $\mathbb{R}^d$ à moments d'ordre deux finis, la distance de transport quadratique est

$$W_2(\mu, \nu)^2 = \min_{\pi \in \Pi(\mu, \nu)} \int |x - y|^2 d\pi(x, y). \quad (4.9)$$

Dans le cas discret, on retrouve (4.8) avec $C_{ij} = |x_i - y_j|^2$. La distance dépend des masses et des positions : des vecteurs de poids égaux placés à des endroits différents ne décrivent pas nécessairement la même mesure. La racine carrée W_2 est une distance, et $W_2(\mu, \nu) = 0$ exactement lorsque $\mu = \nu$. Son carré ne vérifie généralement pas l'inégalité triangulaire : pour des masses de Dirac en 0, 1 et 2, les distances au carré correspondantes valent 1, 1 et 4.

Le théorème de Brenier précise une condition sous laquelle le couplage optimal est une application [3]. Si la source μ est absolument continue par rapport à la mesure de Lebesgue et si les deux mesures ont des moments d'ordre deux finis, le problème quadratique admet un unique couplage optimal. Il est induit par une application $T = \nabla \varphi$, où φ est convexe, et T est unique μ -presque partout. L'absolue continuité et le choix du coût sont essentiels ; on ne peut pas affirmer cette conclusion pour des mesures et des coûts arbitraires.

4.4 Applications

4.4.1 Transférer une palette de couleurs

Pour deux images ayant le même nombre n de pixels, notons $x_i, y_j \in \mathbb{R}^3$ leurs couleurs RVB. Une affectation minimisant $\sum_i |x_i - y_{\sigma(i)}|^2$ donne une image dont la couleur à la position source i est $y_{\sigma(i)}$. Chaque couleur cible est utilisée une fois : la sortie possède donc exactement la distribution empirique des couleurs de la cible. L'optimisation limite les changements de couleur, mais ne comporte aucun terme de régularité spatiale ; la préservation des contours et des textures n'est donc pas garantie.



FIGURE 4.6. Transfert de palette. Colonnes : source, cible et source recolorée. En haut, les couleurs sur la grille de l'image ; en bas, les nuages de points correspondants dans l'espace RVB. Une permutation des couleurs cibles donne le même nuage dans les deux dernières colonnes.

Cette distinction sépare le transport de couleurs du déplacement des pixels dans le plan de l'image. Ici, les positions transportées sont des points de l'espace des couleurs ; la grille d'affichage reste fixe. Pour des images de tailles différentes ou des groupes de couleurs pondérés, on utilise un couplage et une règle précisant

sa conversion en couleurs de sortie. Une moyenne barycentrique des couleurs couplées ne préserve pas nécessairement l'histogramme cible exact.

4.4.2 Barycentres de formes

La moyenne euclidienne pondérée de trois points minimise $\alpha|x - r|^2 + \beta|y - r|^2 + \gamma|z - r|^2$, où $\alpha, \beta, \gamma \geq 0$ et $\alpha + \beta + \gamma = 1$. Elle vaut $r = \alpha x + \beta y + \gamma z$. Remplacer les distances euclidiennes au carré par des distances de transport au carré définit un barycentre de Wasserstein [1] :

$$\min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} \alpha W_2(\mu, \rho)^2 + \beta W_2(\nu, \rho)^2 + \gamma W_2(\eta, \rho)^2. \quad (4.10)$$

On note $\mathcal{P}_2$ l'ensemble des mesures de probabilité à moment d'ordre deux fini. Dans un calcul sur une grille, on restreint plutôt ρ à des poids de probabilité sur cette grille. Une forme se représente par une densité uniforme à l'intérieur, normalisée pour avoir une masse totale égale à un. Faire varier les trois poids produit une famille de distributions interpolées (figure 4.7). Ces distributions ne sont pas nécessairement uniformes à l'intérieur d'une forme au contour net, et leur unicité demande des hypothèses supplémentaires.

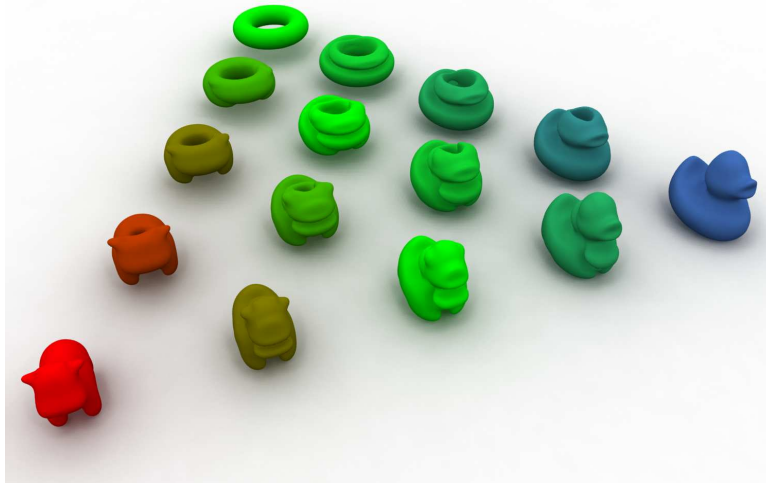


FIGURE 4.7. Interpolation barycentrique numérique de formes tridimensionnelles représentées par des distributions de probabilité. Les trois formes de référence occupent les sommets du triangle ; les points intérieurs correspondent à différents poids dans (4.10).

4.4.3 Comparer des textes

Normaliser les nombres d'occurrences des mots d'un texte donne une distribution de probabilité. Le transport compare deux telles distributions à l'aide d'un coût décrivant la dissimilarité entre mots, par exemple une distance entre leurs plongements vectoriels. Le choix de ce coût détermine quelles substitutions sont peu coûteuses ; on peut aussi l'apprendre pour une tâche donnée [5]. Les histogrammes de mots oublient leur ordre et ne peuvent donc pas, à eux seuls, saisir toutes les distinctions grammaticales ou sémantiques.



FIGURE 4.8. Nuages des mots les plus fréquents de deux textes. Leur taille indique leur fréquence relative ; ces illustrations ne constituent pas un histogramme numérique complet.

Pour aller plus loin. Pour la théorie, on pourra consulter [9, 10]. Les méthodes numériques et les applications sont développées dans [8]. Le cours [Optimal Transport for Machine Learning \(OT4ML\)](#) propose une autre introduction à ces idées.

Remerciements. Je remercie Vincent Beck, Gwenn Guichaoua et Marie-Noëlle Peyré pour leurs relectures attentives.

Références

- [1] Martial Agueh and Guillaume Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2) :904–924, 2011.
- [2] Garrett Birkhoff. Tres observaciones sobre el álgebra lineal. *Universidad Nacional de Tucumán Revista Series A*, 5 :147–151, 1946.
- [3] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4) :375–417, 1991.
- [4] George B. Dantzig. Application of the simplex method to a transportation problem. In Tjalling C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 359–373. Wiley, 1951.
- [5] Gao Huang, Chuan Guo, Matt J Kusner, Yu Sun, Fei Sha, and Kilian Q Weinberger. Supervised word mover’s distance. In *Advances in Neural Information Processing Systems*, pages 4862–4870, 2016.
- [6] Leonid Kantorovich. On the transfer of masses (in russian). *Doklady Akademii Nauk*, 37(2) :227–229, 1942.
- [7] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l’Académie Royale des Sciences*, pages 666–704, 1781.
- [8] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6) :355–607, 2019. URL <https://arxiv.org/abs/1803.00567>.
- [9] Filippo Santambrogio. *Optimal transport for applied mathematicians*. Birkhäuser, 2015.
- [10] Cédric Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics Series. American Mathematical Society, 2003. ISBN 9780821833124.
- [11] 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.