

Transport Optimal

Théorie et Applications

Timothée Schmoderer

27 février 2018

INSA Rouen Normandie

Département Génie Mathématique

Introduction

Théorie du Transport Optimal

Définitions

Existence du transport Optimal

Formulation Duale

Cas quadratique dans \mathbb{R}^n

Formulation de Benamou et Brenier

Théorie des opérateurs proximaux

Résolution numérique

Exemple numériques

Exemple 1D

Exemple 2D

Conclusion

Introduction

1. 1781 - Gaspard Monge : *Mémoire sur la théorie des Déblais et des remblais*

Introduction

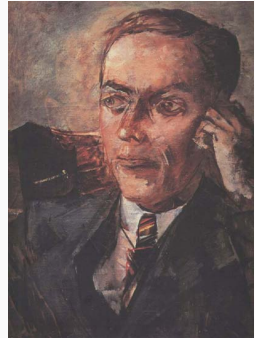
1. 1781 - Gaspard Monge : *Mémoire sur la théorie des Déblais et des remblais*
2. 1940 - Leonid Kantorovitch : Formulation moderne



(a) Gaspard Monge



(b) Mémoire sur la théorie des déblais et remblais



(c) Leonid Kantorovitch

Introduction

1. 1781 - Gaspard Monge : *Mémoire sur la théorie des Déblais et des remblais*
2. 1940 - Leonid Kantorovitch : Formulation moderne
3. 1999 - J.D. Benamou et Y. Brenier : Nouveaux développements

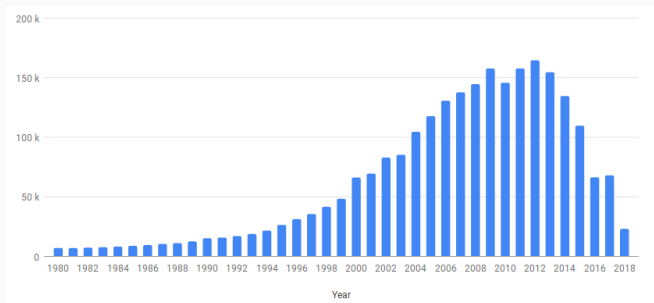


Figure 1 – Popularité des termes "Optimal Transport" ces 40 dernières années sur Google Scholar

1. Benamou, Jean-David et Brenier, Yann, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem.*
2. Papadakis, Nicolas et Peyré, Gabriel et Oudet, Edouard, *Optimal transport with proximal splitting*
3. Santambrogio, F., *Optimal Transport for Applied Mathematicians : Calculus of Variations, PDEs, and Modeling*
4. Combettes PL., Pesquet JC, *Proximal methods in signal processing*

Théorie du Transport Optimal

Le transport

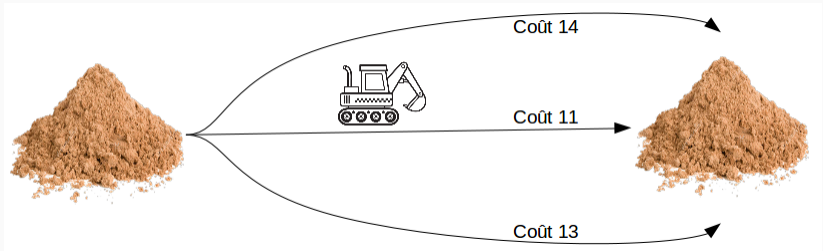


Figure 1 – Le transport optimal dans sa version vulgarisée

Definition

Soient deux mesures de probabilité μ sur $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ et ν sur $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ munis de leur tribu borélienne. Un **transport** est une application $T : \mathcal{X} \rightarrow \mathcal{Y}$ qui envoie la mesure μ sur la mesure ν . C'est à dire :

$$\forall B \in \mathcal{B}(\mathcal{Y}), \quad \mu(T^{-1}(B)) = \nu(B) \quad (1)$$

Notons $T_{\#}\mu = \nu$

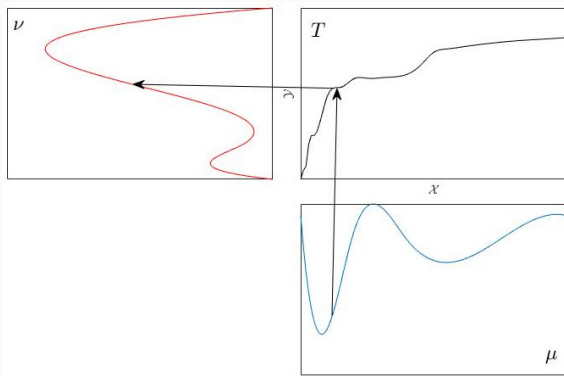


Figure 1 – Illustration de l'application de transport

Definition

Le coût est une application

$$C : \left\{ \begin{array}{ll} \mathcal{X} \times \mathcal{Y} & \longrightarrow [0, +\infty] \\ (x, y) & \longmapsto C(x, y) \end{array} \right.$$

Le problème de transport Optimal

Étant données deux mesures de probabilités μ sur \mathcal{X} , ν sur \mathcal{Y} et une application coût C , trouver une application de transport T réalisant le

$$\inf \left\{ M(T) := \int_{\mathcal{X}} C(x, T(x)) \, d\mu(x), \quad T_{\#}\mu = \nu \right\} \quad (\text{MP})$$

Relaxation, Plan de transport optimal

Étant données deux mesures de probabilités μ et ν et un coût C , trouver la mesure π réalisant le

$$\inf \left\{ K(\pi) := \int_{\mathcal{X} \times \mathcal{Y}} C(x, y) d\pi(x, y), \quad \pi \in \Pi(\mu, \nu) \right\} \quad (\text{KP})$$

Où, $\Pi(\mu, \nu)$ est l'ensemble des plans de transports :

$$\Pi(\mu, \nu) = \left\{ \begin{array}{l} \pi \text{ une probabilité sur } \mathcal{X} \times \mathcal{Y} \\ \forall A \in \mathcal{B}(\mathcal{X}), \quad \pi(A \times \mathcal{Y}) = \mu(A) \\ \forall B \in \mathcal{B}(\mathcal{Y}), \quad \pi(\mathcal{X} \times B) = \nu(B), \end{array} \right. \quad (1)$$

Relaxation, Plan de transport optimal

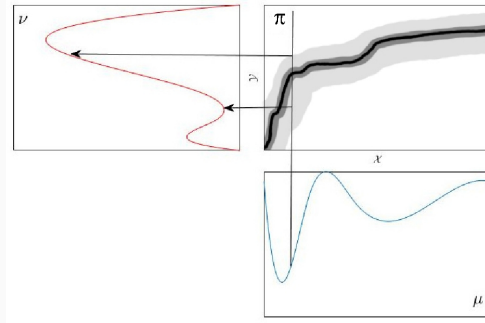


Figure 2 – Illustration de la notion de plan de transport

Théorème 2.1 :

Soit \mathcal{X} et \mathcal{Y} des espaces métriques compacts. Supposons que le coût $C : \mathcal{X} \times \mathcal{Y} \rightarrow [0, +\infty]$ soit continue. Alors le problème (KP) admet une solution.

Étant données deux mesures de probabilités μ et ν et un coût C , trouver deux fonctions continues, bornées ϕ et ψ réalisant le

$$\sup \left\{ \int_{\mathcal{X}} \phi \, d\mu + \int_{\mathcal{Y}} \psi \, d\nu, \quad (\phi, \psi) \in C_b(\mathcal{X}) \times C_b(\mathcal{Y}), \quad \phi \oplus \psi \leq C \right\}$$

(DKP)

Théorème 2.2 :

Supposons que \mathcal{X} et \mathcal{Y} soient compact et que C est continue. Alors il existe une solution (ϕ, ψ) au problème (DKP).

Théorème 2.3 :

Les problèmes (KP) et (DKP) sont bien duaux l'un de l'autre :

$$\min (KP) = \max (DKP) \quad (1)$$

Théorème 2.4 : *Existence du transport optimal*

Soient μ et ν des mesures de probabilités sur un domaine compact Ω de \mathbb{R}^n et un coût $C(x, y) = h(x - y)$, avec h strictement convexe.

Alors il existe une solution π au problème (KP).

De plus, π est unique, de la forme $(Id, T)_{\#}\mu$ si μ est absolument continue et $\partial\Omega$ est μ - négligeable.

Enfin, une solution de (DKP), ϕ qui est liée à T par :

$$T(x) = x - (\partial h)^{-1}(\nabla \phi(x)) \quad (2)$$

Prenons $C(x, y) = \frac{1}{2}\|x - y\|^2$.

Théorème 2.5 :

Soient μ et ν des probabilités sur \mathbb{R}^n et $C(x, y) = \frac{1}{2}\|x - y\|^2$.

Supposons que :

1. $\int \|x\|^2 d\mu, \int \|y\|^2 d\nu < +\infty$
2. μ ne donne pas de masse aux hypersurfaces de classe C^2 .

alors il existe une unique application de transport optimal T , donnée par $T = \nabla u$ pour une fonction convexe u .

Mettons nous dans \mathbb{R}^n , prenons $\mu = f_0 \, dx$ et $\nu = f_1 \, dx$.

$$\begin{aligned}\forall B \in \mathcal{B}(\mathbb{R}^n), \quad \int_B f_1(y) dy &= \int_{T^{-1}(B)} f_0(x) dx \\ &= \int_B \sum_{x \in T^{-1}(y)} \left(\frac{f_0(x)}{|\det \nabla T(x)|} \right) dy \\ \Rightarrow f_1(y) &= \sum_{x \in T^{-1}(y)} \left(\frac{f_0(x)}{|\det \nabla T(x)|} \right)\end{aligned}$$

Simplifications

Mettons nous dans \mathbb{R}^n , prenons $\mu = f_0 \, dx$ et $\nu = f_1 \, dx$.

$$\begin{aligned}\forall B \in \mathcal{B}(\mathbb{R}^n), \quad \int_B f_1(y) dy &= \int_{T^{-1}(B)} f_0(x) dx \\ &= \int_B \sum_{x \in T^{-1}(y)} \left(\frac{f_0(x)}{|\det \nabla T(x)|} \right) dy \\ \Rightarrow f_1(y) &= \sum_{x \in T^{-1}(y)} \left(\frac{f_0(x)}{|\det \nabla T(x)|} \right) \\ f_1(T(x)) &= \frac{f_0(x)}{|\det \nabla T(x)|} \quad (J)\end{aligned}$$

Mettons nous dans \mathbb{R}^n , prenons $\mu = f_0 \, dx$ et $\nu = f_1 \, dx$. Dans le cas quadratique :

$$f_1(\nabla u(x)) = \frac{f_0(x)}{\det Hu(x)} \quad (\text{MA})$$

Mettons nous dans \mathbb{R}^n , prenons $\mu = f_0 \, dx$ et $\nu = f_1 \, dx$. Notons $\mathcal{T}(f_0, f_1)$ l'ensemble des applications qui vérifient (??).

Definition

Métrique de Wasserstein Pour les coûts de transport de la forme $C(x, y) = \|x - y\|^p$, nous pouvons définir une métrique entre f_0 et f_1 par

$$W(f_0, f_1)^p = \min_{T \in \mathcal{T}(f_0, f_1)} \int \|x - T(x)\|^p f_0(x) \, dx$$

Formulation de Benamou et Brenier

Théorème 3.1 : Benamou, Brenier

Soient f_0 et f_1 deux densités de probabilité assez régulières. Alors,

$$\min_{T \in \mathcal{T}(f_0, f_1)} \int \|x - T(x)\|^2 f_0 dx = \min_{(f, v) \in C_v} \int_{\mathbb{R}^n} \int_0^1 f(t, x) \|v(t, x)\|^2 dt dx$$

Avec,

$f(t, x) : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$ la densité

$v(t, x) : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ champ de vecteurs vitesses

et

$$C_v = \left\{ (f, v) \mid \frac{\partial f}{\partial t} + \operatorname{div}_x(fv) = 0, f(0, \cdot) = f_0, f(1, \cdot) = f_1 \right\}$$

Théorie des opérateurs proximaux

Notation : Le domaine d'une fonction $f : \mathbb{R}^n \longrightarrow]-\infty, +\infty]$ est $\mathcal{D}(f) = \{x \in \mathbb{R}^n \mid f(x) < +\infty\}$.

Notons $\Gamma_0(\mathbb{R}^n)$ l'ensemble des fonctions convexes semi-continues inférieurement à valeurs dans $]-\infty, +\infty]$ telles que $\mathcal{D}(f) \neq \emptyset$.

Definition

Soit $f \in \Gamma_0(\mathbb{R}^n)$, le sous - différentiel de f est l'application suivante,

$$\partial f : \left\{ \begin{array}{ll} \mathbb{R}^n & \longrightarrow \mathcal{P}(\mathbb{R}^n) \\ x & \longmapsto \{u \in \mathbb{R}^n \mid \forall y \in \mathbb{R}^n, \langle u, y - x \rangle + f(x) \leq f(y)\} \end{array} \right.$$

Avec $\mathcal{P}(\mathbb{R}^n)$, l'ensemble des parties de \mathbb{R}^n .

Sous différentiel

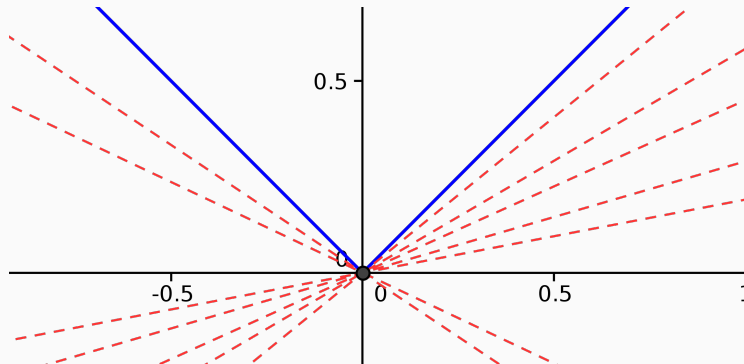


Figure 3 – Illustration du sous - différentiel de $|x|$ en 0

Propriété :

Soit $f \in \Gamma_0(\mathbb{R}^n)$. Si f est différentiable sur son domaine, alors

$$\forall x \in \mathcal{D}(f)^\circ, \partial f(x) = \{\nabla f(x)\}$$

Théorème 4.1 :

Soit $f \in \Gamma_0(\mathbb{R}^n)$. Le point $x^* \in \mathbb{R}^n$ est un point de minimum global de f si et seulement si

$$0 \in \partial f(x^*)$$

Definition

Soit $f \in \Gamma_0(\mathbb{R}^n)$. L'opérateur proximal de f , noté Prox_f , est défini pour tout $x \in \mathbb{R}^n$ par :

$$\text{Prox}_f(x) = \arg \min_{y \in \mathbb{R}^n} f(y) + \frac{1}{2} \|x - y\|^2$$

Opérateur proximal

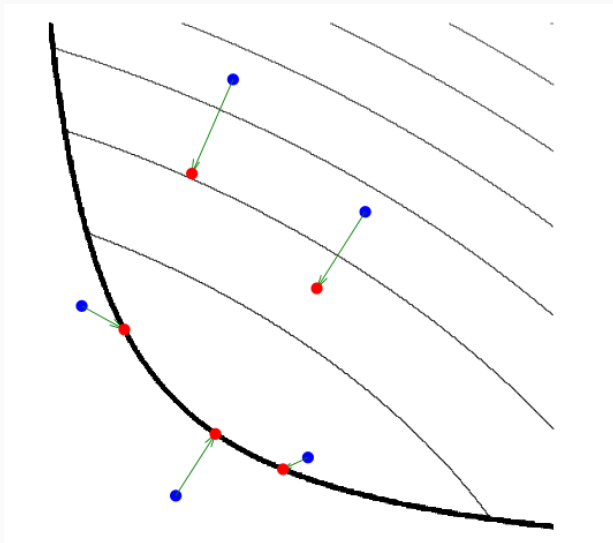


Figure 3 – Illustration de l'opérateur proximal

Propriété :

Soit une fonction $f \in \Gamma_0(\mathbb{R}^n)$. Alors, $\text{Prox}_f(x)$ existe et est unique pour tout $x \in \mathbb{R}^n$.

De plus, il est caractérisé par :

$$\forall (x, p) \in \mathbb{R}^n \times \mathbb{R}^n \quad p = \text{Prox}_f x \iff x - p \in \partial f(p)$$

En particulier, si f est différentiable alors :

$$\forall (x, p) \in \mathbb{R}^n \times \mathbb{R}^n \quad p = \text{Prox}_f x \iff x - p = \nabla f(p)$$

Propriété :

Le point x^* minimise f si et seulement si

$$x^* = \text{Prox}_f(x^*)$$

Soient f_1 et f_2 deux fonctions de $\Gamma_0(\mathbb{R}^n)$ telles que $f_1 + f_2 \in \Gamma_0(\mathbb{R}^n)$, en particulier le support de $f_1 + f_2$ est non vide. Supposons que

$$\lim_{\|x\| \rightarrow +\infty} f_1(x) + f_2(x) = +\infty$$

Considérons alors le problème de minimisation, Trouver $x \in \mathbb{R}^n$ qui réalise le

$$\min_{x \in \mathbb{R}^n} f_1(x) + f_2(x) \tag{3}$$

Propriété :

Le problème 3 admet une solution et pour tout $\gamma > 0$, elle est caractérisée par :

$$\begin{cases} x^* &= \text{Prox}_{\gamma f_2} y \\ \text{Prox}_{\gamma f_2} y &= \text{Prox}_{\gamma f_1} (2\text{Prox}_{\gamma f_2} y - y) \end{cases}$$

Algorithme de Douglas Rachford

Soient $\gamma > 0$ et $y_0 \in \mathbb{R}^n$

Pour $n = 0, 1, \dots$

$$x_n = \text{Prox}_{\gamma f_2} y_n$$

$$\lambda_n \in]0, 2[$$

$$y_{n+1} = y_n + \lambda_n (\text{Prox}_{\gamma f_1} (2x_n - y_n) - x_n)$$

Fin

Résolution numérique

Posons le changement de variable $m = fv$.

$$\min_{(f,m) \in \mathcal{C}_v} \int_{[0,1]^n} \int_0^1 \frac{\|m(t,x)\|^2}{f(t,x)} dt dx$$

Avec,

$$\begin{aligned} \mathcal{C}_v = \{ (f, m) \mid & \frac{\partial f}{\partial t} + \operatorname{div}_x(m) = 0, \\ & f(0, \cdot) = f_0, \quad f(1, \cdot) = f_1, \\ & m(\cdot, 0) = m(\cdot, 1) = 0 \} \end{aligned}$$

Grilles de discrétisations

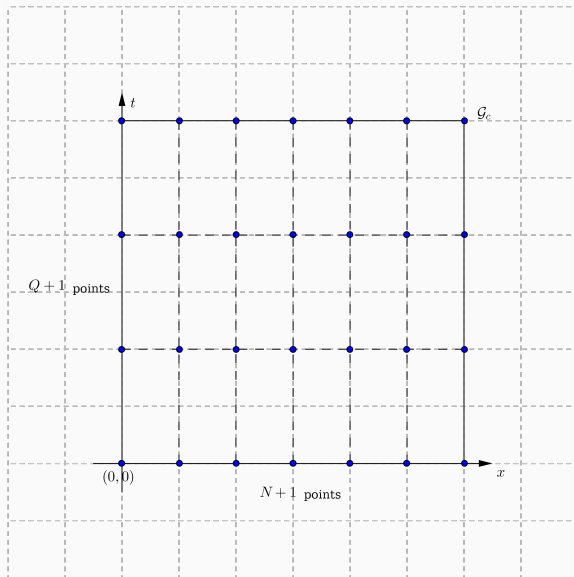
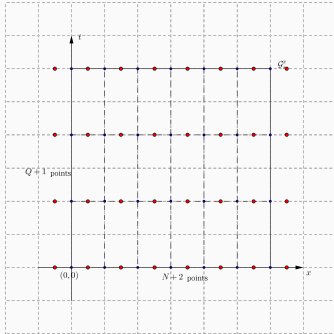
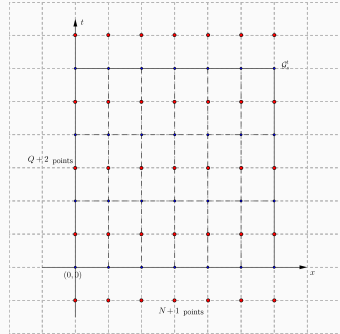


Figure 3 – Grille centrée

Grilles de discrétisations



(a) Grille décentrée en espace



(b) Grille décentrée en temps

Figure 3 – Grilles de discrétisation pour $N = 6$ et $Q = 3$

Notons

$$V = (m_{ij}, f_{ij})_{0 \leq i \leq N}^{0 \leq j \leq Q}$$

et

$$U = (\bar{m}, \bar{f}) = \left((\bar{m}_{ij})_{-1 \leq i \leq N}^{0 \leq j \leq Q}, (\bar{f}_{ij})_{0 \leq i \leq N}^{-1 \leq j \leq Q} \right)$$

Interpolation :

$$\mathcal{I} : \left\{ \begin{array}{lcl} \mathcal{E}_s & \longrightarrow & \mathcal{E}_c \\ (\bar{m}, \bar{f}) & \longmapsto & \begin{cases} m_{i,j} = (\bar{m}_{i-1,j} + \bar{m}_{i,j})/2 \\ f_{i,j} = (\bar{f}_{i,j-1} + \bar{f}_{i,j})/2 \end{cases} \end{array} \right.$$

Opérateur de divergence :

$$\text{div} : \left\{ \begin{array}{ll} \mathcal{E}_s & \longrightarrow \mathbb{R}^{\mathcal{G}_c} \\ U & \longmapsto \text{div}(U)_{i,j} = N(\bar{m}_{i,j} - \bar{m}_{i-1,j}) + Q(\bar{f}_{i,j} - \bar{f}_{i,j-1}) \end{array} \right.$$

Extraction des frontières

$$b : \left\{ \begin{array}{lcl} \mathcal{E}_s & \longrightarrow & \mathbb{R}^{Q+1} \times \mathbb{R}^{Q+1} \times \mathbb{R}^{N+1} \times \mathbb{R}^{N+1} \\ U & \longmapsto & \left((\bar{m}_{-1,j}, \bar{m}_{N,j})_{j=0}^Q, (\bar{f}_{i,-1}, \bar{f}_{i,Q})_{i=0}^N \right) \end{array} \right.$$

Problème discrétisé

$$\min_{(U,V) \in \mathcal{E}_s \times \mathcal{E}_c} \mathcal{J}(V) + \iota_{\mathcal{C}}(U) + \iota_{\mathcal{C}_s}(U, V)$$

Ici, l'indicatrice d'un ensemble convexe est définie de la façon suivante ; $\iota_{\mathcal{C}} = \begin{cases} 0 & \text{si } U \in \mathcal{C} \\ +\infty & \text{sinon} \end{cases}$.

$$\mathcal{J}(V) = \sum_{k \in \mathcal{G}_c} \frac{\|m_k\|^2}{f_k}$$

$$\mathcal{C} = \{U \in \mathcal{E}_s \mid \operatorname{div}(U) = 0, \ b(U) = b_0 = (0, 0, f_0, f_1)\}$$

$$\mathcal{C}_s = \{(U, V) \in \mathcal{E}_s \times \mathcal{E}_c \mid V = \mathcal{I}(U)\}$$

Algorithme de Douglas Rachford pour le problème de transport optimal

$$\text{Prox}_{\gamma f_1}(U, V) = (\text{Proj}_{\mathcal{C}}(U), \text{Prox}_{\gamma \mathcal{J}}(V))$$

$$\text{Prox}_{\gamma f_2}(U, V) = \text{Proj}_{\mathcal{C}_s}(U, V)$$

Soit une suite $(x_n, y_n) \in (\mathcal{E}_s \times \mathcal{E}_c)^2$. A partir, d'une condition initiale y_0 ,

$$x_n = \text{Proj}_{\mathcal{C}_s}(y_n)$$

$$y_{n+1} = y_n + \lambda_n (\text{Prox}_{\gamma f_1}(2x_n - y_n) - x_n)$$

Alors $x_n \rightarrow x^*$ une solution du problème.

Opérateur proximal de la fonction de coût

Proposition 5.1 :

Nous avons

$$\forall w \in \mathcal{G}, \text{Prox}_{\gamma \mathcal{J}}(w) = (\text{Prox}_{\gamma J}(w_k))_{k \in \mathcal{G}}$$

Où, pour tout $w_k = (m_k, f_k) \in \mathbb{R}^n \times \mathbb{R}$,

$$\text{Prox}_{\gamma J}(m_k, f_k) = \begin{cases} \left(\frac{f_k^* m_k}{f_k^* + \gamma}, f_k^* \right) & \text{si } f_k^* > 0 \\ (0, 0) & \text{sinon} \end{cases}$$

et f_k^* est la plus grande racine réelle du polynôme de degré 3 :

$$P(X) = (X - f_k)(X + \gamma)^2 - \frac{\gamma}{2} \|m_k\|^2 = 0 \quad (3)$$

Écrivons,

$$\mathcal{C} = \{U \in \mathcal{E}_s \mid AU = y\} \quad A = \begin{pmatrix} \text{div} \\ b \end{pmatrix}, \quad y = \begin{pmatrix} 0 \\ b_0 \end{pmatrix}$$

C'est la projection sur un ensemble convexe sous contraintes.

$$\text{Proj}_{\mathcal{C}}(U) = U - A^*(AA^*)^{-1}(y - AU)$$

$$\mathcal{C}_s = \{(U, V) \in \mathcal{E}_s \times \mathcal{E}_c \mid V = \mathcal{I}(U)\}$$

$$\text{Proj}_{\mathcal{C}_s}(U, V) = (\tilde{U}, \mathcal{I}(\tilde{U}))$$

$$\tilde{U} = (Id + \mathcal{I}^* \mathcal{I})^{-1}(U + \mathcal{I}^*(V))$$

$$\min_{U \in \mathcal{E}_s} \mathcal{J}_\omega^\beta(\mathcal{I}(U)) + \iota_{\mathcal{C}}(U) \quad (4)$$

Où, $\beta \in [0, 1]$, et le vecteur de poids $\omega = (\omega_k)_{k \in \mathcal{G}_c}$ vérifie $1 \leq \omega_k \leq +\infty$. La fonctionnelle généralisée est définie par :

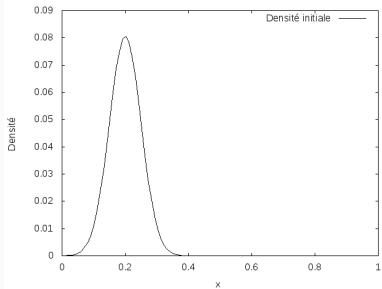
$$\mathcal{J}_\omega^\beta(V) = \sum_{k \in \mathcal{G}_c} \omega_k J_\beta(m_k, f_k) \quad (5)$$

$$\forall (m, f) \in \mathbb{R} \times \mathbb{R}, \quad J_\beta = \begin{cases} \frac{\|m\|^2}{2f^\beta} & \text{si } f > 0 \\ 0 & \text{si } (m, f) = (0, 0) \\ +\infty & \text{sinon} \end{cases} \quad (6)$$

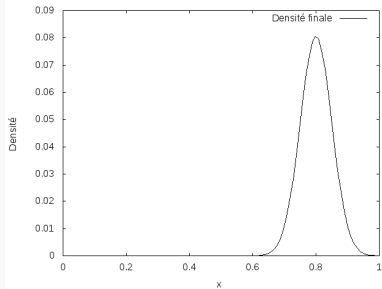
Remarques :

1. Le coefficient $\beta \in [0, 1]$ agit comme un interrupteur entre l'interpolation L^2 ($\beta = 0$) entre les deux densités et le calcul du transport optimal ($\beta = 1$), i.e. le calcul de la distance L^2 de Wasserstein.
2. Les poids ω_k pondèrent certaines régions de l'espace temps et permettent d'autoriser plus ou moins le passage d'une densité dans une région de l'espace ; $\omega_k = +\infty$ correspond à une zone interdite.

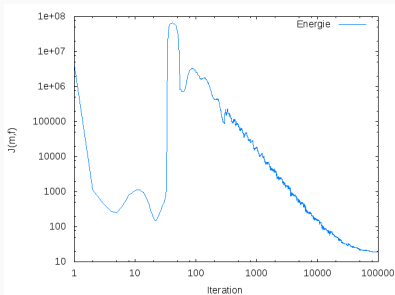
Exemple numériques



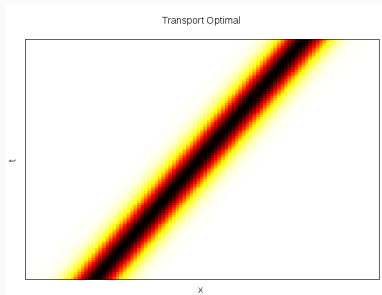
(a) Densité initiale



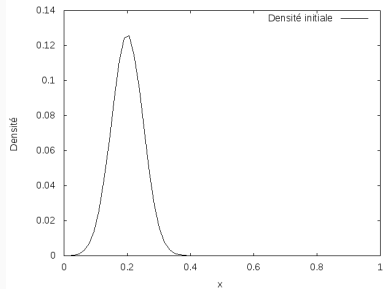
(b) Densité finale



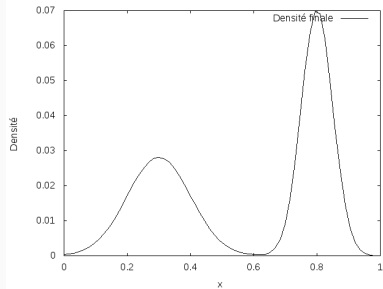
(c) $\mathcal{J}(m, f)$



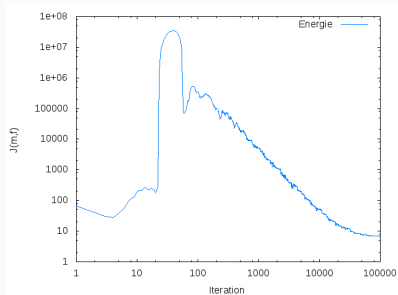
(d) Le Transport



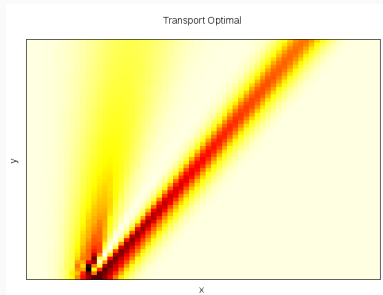
(a) Densité initiale



(b) Densité finale

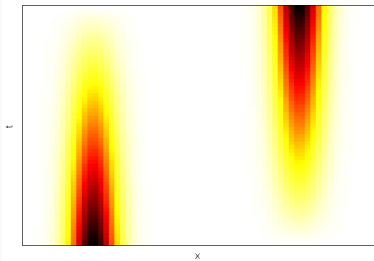


(c) $\mathcal{J}(m, f)$



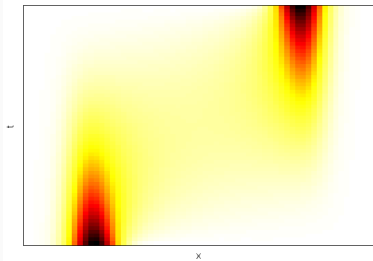
(d) Le Transport

Interpolation L2



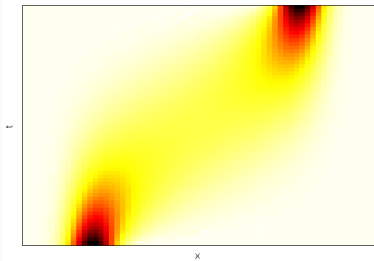
(a) $\beta = 0$

Transport Optimal Généralisé



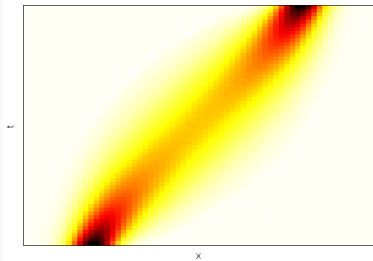
(b) $\beta = 1/4$

Transport Optimal Généralisé

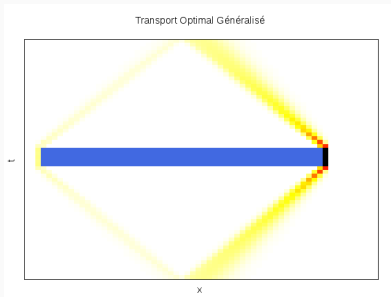


(c) $\beta = 1/2$

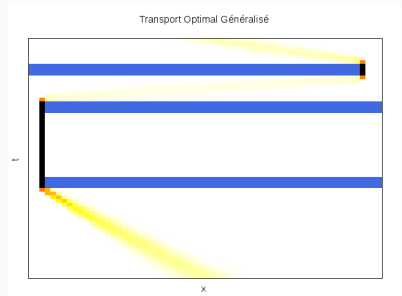
Transport Optimal Généralisé



(d) $\beta = 3/4$

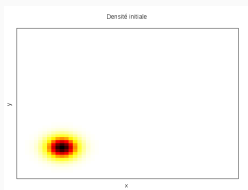


(a) Un obstacle au milieu du passage

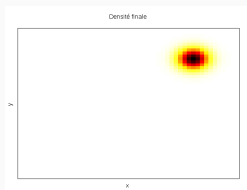


(b) Un "labyrinthe"

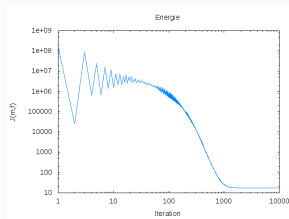
Figure 6 – Transport de deux gaussienne avec des obstacles



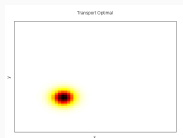
(a) Densité initiale



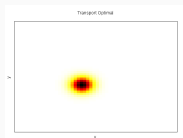
(b) Densité finale



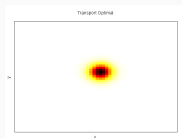
(c) Energie



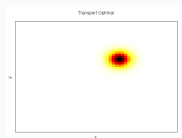
(d) $t = 0.2$



(e) $t = 0.4$

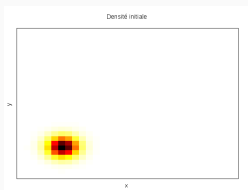


(f) $t = 0.6$

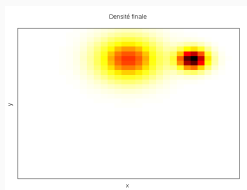


(g) $t = 0.8$

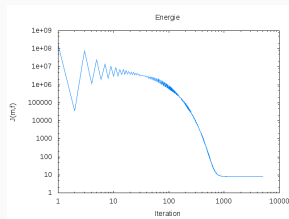
Figure 7 – Transport de deux gaussienne $N = P = Q = 50$



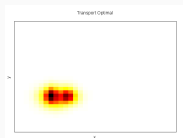
(a) Densité initiale



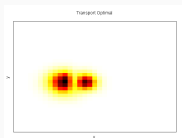
(b) Densité finale



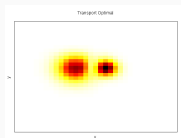
(c) Energie



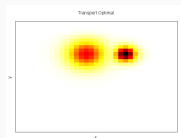
(d) $t = 0.2$



(e) $t = 0.4$

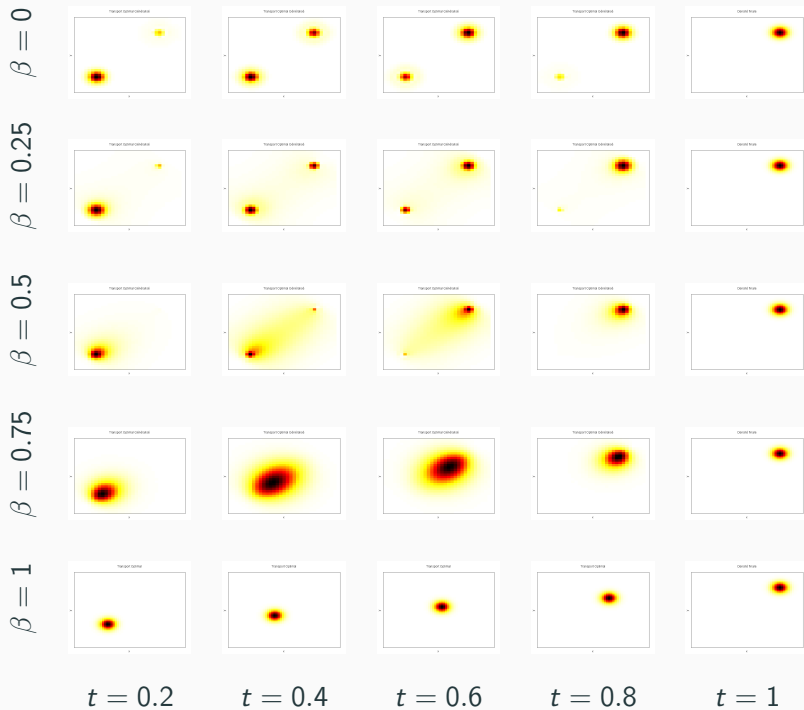


(f) $t = 0.6$



(g) $t = 0.8$

Figure 8 – Transport de deux gaussienne $N = P = Q = 31$



Vidéo

Vidéo

Conclusion
