

# Transport optimal numérique et la physique quantique

Aristote Koen  
Thomas Chatrefou

Février 2020

## Introduction

Situons notre contexte à Paris et prenons en considération l'ensemble des boulangeries qui produisent des pains au chocolat, ces derniers devant être livrés chaque matin aux cafés dans lesquels les clients pourront les goûter. La production et la consommation de pains au chocolat sont décrites par  $\mu \in \mathbb{R}_+^N$ , ( $N \in \mathbb{N}$ ) et  $\nu \in \mathbb{R}_+^N$  respectivement et nous supposons que la production totale est égale à la consommation totale :

$$\sum_{i=1}^N \mu_i = \sum_{j=1}^N \nu_j$$

Le problème du transport optimal consiste alors à transporter la quantité de pains au chocolat produite par la boulangerie  $i \in I$  ( $I := \llbracket 1, N \rrbracket$  est l'ensemble des boulangeries) à un café  $j \in J$  ( $J := \llbracket 1, N \rrbracket$  est l'ensemble des cafés) de sorte que le coût de transport  $C_{ij} := c(x_i, y_j)$ , par exemple la distance entre la boulangerie et le café, soit minimal. On a noté  $x_i$  ( $y_j$ ) la position de la boulangerie  $i$  (café  $j$ ). Le problème du transport optimal ou problème de *Monge-Kantorovich* consiste à chercher un couplage optimal  $\gamma \in \mathbb{R}_+^{N \times N}$  qui nous indique comment la masse en  $i$  est répartie sur chaque  $j \in J$  (on peut imaginer que  $\gamma$  représente le camion qui transporte les pains au chocolat de la boulangerie  $i$  au café  $j$ ). Si le couplage optimal  $\gamma$  consiste à assigner le même café  $j$  à tous les pains au chocolat produits par la boulangerie  $i$ , alors on dit que le couplage est *déterministe*. Le problème de Monge-Kantorovich s'écrit sous la forme

$$\min \left\{ \sum_{1 \leq i, j \leq N} C_{ij} \gamma_{ij} \mid \gamma \in \Pi(\mu, \nu) \right\}, \quad (\mathcal{MK})$$

où

$$\Pi(\mu, \nu) := \left\{ \gamma \in \mathbb{R}_+^{N \times N} \mid \sum_{j=1}^N \gamma_{ij} = \mu_i, \forall i \in I \text{ et } \sum_{i=1}^N \gamma_{ij} = \nu_j, \forall j \in J \right\}$$

On remarque que  $(\mathcal{MK})$  est bien un problème d'optimisation sous contraintes : de positivité  $\gamma_{ij} \geq 0, \forall (i, j) \in I \times J$  et de marginales  $\sum_{j=1}^N \gamma_{ij} = \mu_i, \sum_{i=1}^N \gamma_{ij} = \nu_j$ . Le but de ce projet est d'étudier différents algorithmes pour résoudre  $(\mathcal{MK})$  et de les appliquer pour calculer l'énergie de répulsion entre les électrons.

# Problème de programmation linéaire

On définit un vecteur  $\mathbf{c} := (C_{11}, \dots, C_{1N}, \dots, C_{N1}, \dots, C_{NN}) \in \mathbb{R}^{N \times N}$ , un vecteur  $\mathbf{x} := (\gamma_{11}, \dots, \gamma_{1N}, \dots, \gamma_{N1}, \dots, \gamma_{NN}) \in \mathbb{R}^{N \times N}$  et un vecteur  $\mathbf{b} := (\mu_1, \dots, \mu_N, \nu_1, \dots, \nu_N) \in \mathbb{R}^{2N}$ . Alors avec le changement d'indice  $k = (i-1)N + j$  on a :

$$\sum_{1 \leq i, j \leq N} C_{ij} \gamma_{ij} = \sum_{k=1}^{N^2} c_k x_k = \langle \mathbf{c}, \mathbf{x} \rangle$$

avec les marginales qui se réécrivent de la façon suivante :

$$\forall p \in \llbracket 1, N \rrbracket, \quad b_p = \mu_p = \sum_{1 \leq i, j \leq N} \delta_{pi} \gamma_{ij} = \sum_{1 \leq i, j \leq N} A_{p, (i-1)N+j} \gamma_{ij} = \sum_{k=1}^{N^2} A_{pk} x_k$$

et

$$\forall p \in \llbracket N+1, 2N \rrbracket, \quad b_p = \nu_{p-N} = \sum_{1 \leq i, j \leq N} \delta_{j, p-N} \gamma_{ij} = \sum_{1 \leq i, j \leq N} A_{p, (i-1)N+j} \gamma_{ij} = \sum_{k=1}^{N^2} A_{pk} x_k$$

on a donc  $A\mathbf{x} = \mathbf{b}$  avec  $A \in \mathcal{M}_{2N, N^2}$  telle que

$$\forall p, i, j \in \llbracket 1, N \rrbracket, \quad A_{p, (i-1)N+j} = \begin{cases} \delta_{pi} & \text{si } p \leq N \\ \delta_{j, p-N} & \text{si } p > N \end{cases}$$

ce qui donne :

$$A = \begin{pmatrix} E_1 & E_2 & \dots & E_N \\ I_N & I_N & \dots & I_N \end{pmatrix}$$

où  $\forall k \in \llbracket 1, N \rrbracket, E_k \in \mathcal{M}_N(\mathbb{R})$  est la matrice dont la  $k$ -ième ligne est remplie de 1.

Afin de calculer la solution  $\gamma$  nous avons programmé une fonction Python `solveOT(mu, nu, C)` à travers l'algorithme du simplexe.

Voici les résultats obtenus pour  $\mu(x) = \chi_{[0,1]}(x), \nu(y) = \exp(-10(y-0.5)^2)$  sur l'intervale  $[0, 1]$  discrétisé en  $N = 20$  points et  $c(x, y) = |x - y|^2$

On observe la masse est principalement concentrée aux alentours de la diagonale, la solution n'est pas déterministe. Calculons le problème dual de  $(\mathcal{MK})$ . Le lagrangien du problème est le suivant :

$$l(\gamma, u, v) = \sum_{i=1}^N \sum_{j=1}^N C_{ij} \gamma_{ij} - \left\langle u, \sum_{j=1}^N \gamma_{i,j} - \mu \right\rangle - \left\langle v, \sum_{i=1}^N \gamma_{ij} - \nu \right\rangle$$

Ainsi le problème dual correspond à :

$$\begin{aligned} & \max_{u, v \in \mathbb{R}^N} \min_{\gamma \in \mathbb{R}^{N \times N}} \sum_{i=1}^N \sum_{j=1}^N C_{ij} \gamma_{ij} - \left\langle u, \sum_{j=1}^N \gamma_{i,j} - \mu \right\rangle - \left\langle v, \sum_{i=1}^N \gamma_{ij} - \nu \right\rangle \\ \iff & \max_{u, v \in \mathbb{R}^N} \langle u, \mu \rangle + \langle v, \nu \rangle + \min_{\gamma \in \mathbb{R}^{N \times N}} \sum_{i=1}^N \sum_{j=1}^N C_{ij} \gamma_{ij} - \left\langle u, \sum_{j=1}^N \gamma_{i,j} \right\rangle - \left\langle v, \sum_{i=1}^N \gamma_{ij} \right\rangle \end{aligned}$$

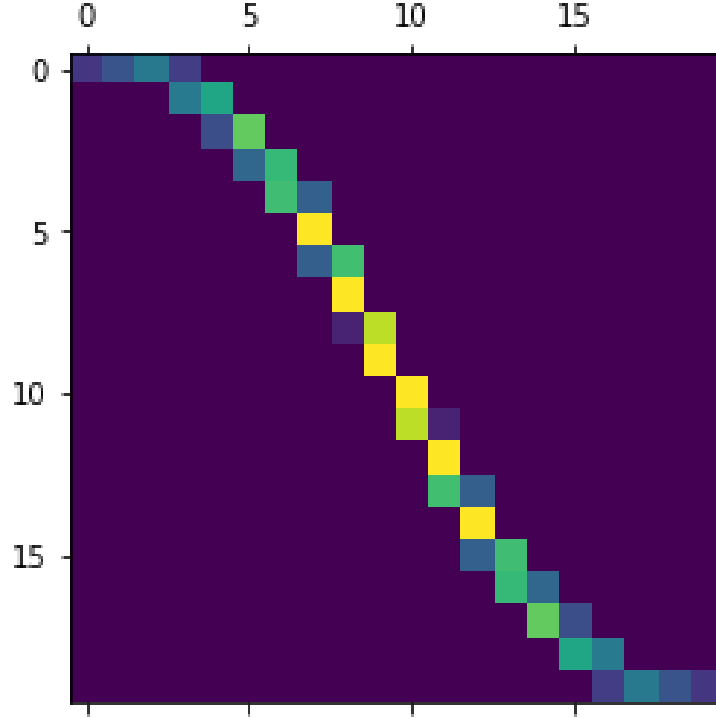


FIGURE 1 – Solution Gamma étoile

$$\Longleftrightarrow \max_{u,v \in \mathbb{R}^N} \langle u, \mu \rangle + \langle v, \nu \rangle + \min_{\gamma \in \mathbb{R}^{N \times N}} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} (C_{ij} - u_i - v_j)$$

Ainsi on a :

$$\min_{\gamma \in \mathbb{R}^{N \times N}} \sum_{i=1}^N \sum_{j=1}^N \gamma_{ij} (C_{ij} - u_i - v_j) = \begin{cases} 0 & \text{si } C_{ij} \geq u_i + v_j \\ -\infty & \text{sinon} \end{cases}$$

Et donc le problème dual s'écrit sous la forme :

$$\max_{u,v \in \mathbb{R}^N} \{ \langle u, \mu \rangle + \langle v, \nu \rangle \mid u, v \in \mathbb{R}^N, C_{ij} \geq u_i + v_j \forall (i, j) \in \llbracket 1 : N \rrbracket \}$$

Le problème dual décrit ci dessus peut être réécrit sous forme standard :

$$\max_{A^T y \leq C} \{ \langle y, b \rangle \}$$

Avec  $b = [\mu \ \nu]$  et  $y = [u \ v]$  les concaténations de  $\mu, \nu$  et  $u, v$  respectivement. Ainsi on peut appliquer l'algorithme du simplexe. On peut tracer les potentiels de Kantorovich  $u^*$  et  $v^*$  en faisant tourner la fonction `solveOTdual` (voir Figure ci-dessous). De plus on vérifie bien numériquement que les variables d'écart  $s$  obtenues à travers l'algorithme du simplexe vérifient bien  $x^* s = 0$ .

Afin de trouver les potentiels de Kantorovich à travers l'algorithme du simplexe nous implémentons une fonction Python `solveOTdual(mu, nu, C)`.

Voici le tracé des potentiels obtenus par la méthode du simplexe pour le problème dual `mu, nu, C`

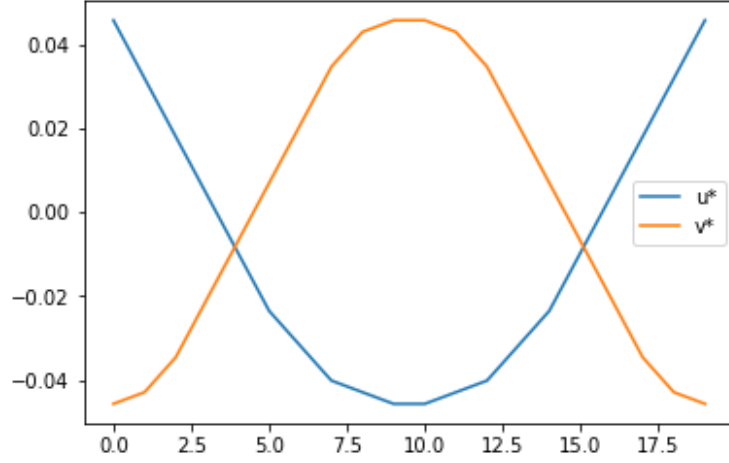


FIGURE 2 – Potentiels  $u, v$  via simplexe

Maintenant nous considérons le cas où  $\mu_i = \nu_j = 1/N \forall (i, j) \in \llbracket 1 : N \rrbracket$ . On sait qu'il existe une permutation optimale  $\sigma^*$  telle que :

$$\gamma_{ij}^* = \begin{cases} 1 & \text{si } j = \sigma^*(i) \\ 0 & \text{sinon} \end{cases}$$

En résolvant le problème de transport optimal pour  $N = 20$  et  $\{x_i\}_{i \geq 0}, \{y_j\}_{j \geq 0}$  générées aléatoirement on observe que la solution  $\gamma^*$  est une matrice de permutation

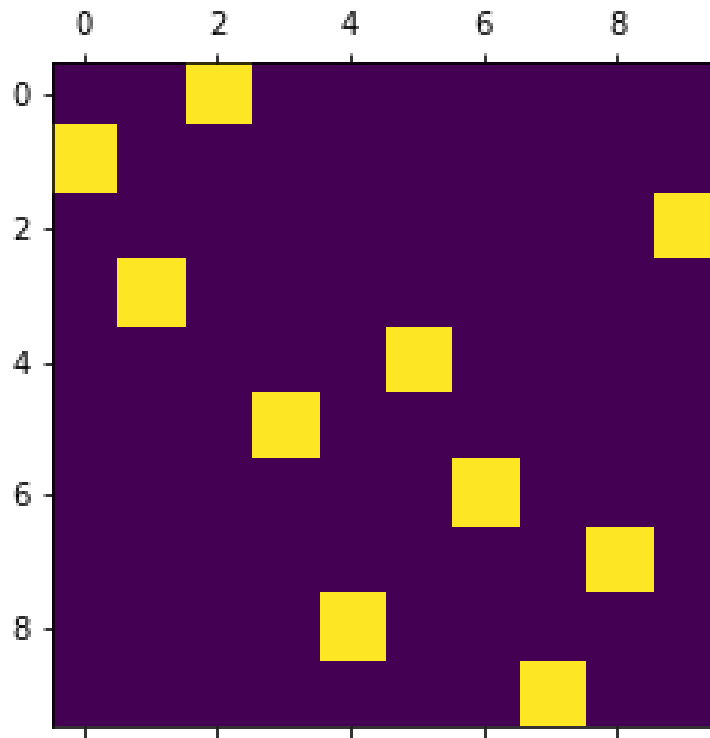


FIGURE 3 – Potentiels  $u, v$  via simplexe

On considère maintenant  $x_i, y_i \in \mathbb{R}^2$  et

$$C_{ij} = c(x_i, y_j) = |x_i^1 - y_j^1|^2 + |x_i^2 - y_j^2|^2$$

On résout de nouveau ce problème de transport avec  $N = 20$  et  $\{x_i\}_{i \geq 0}, \{y_j\}_{j \geq 0}$  générées aléatoirement selon une loi uniforme  $[0, 1]$  et une loi normale centrée réduite respectivement. On représente ci-dessous la solution  $\gamma^*$  puis les points dans le plan  $xy$  reliés entre eux.

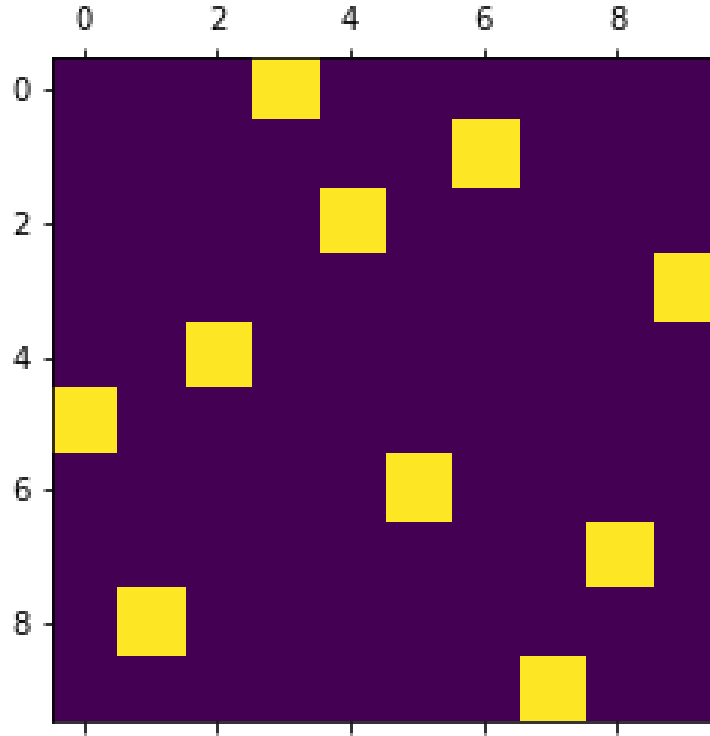


FIGURE 4 – Potentiels  $u, v$  via simplexe

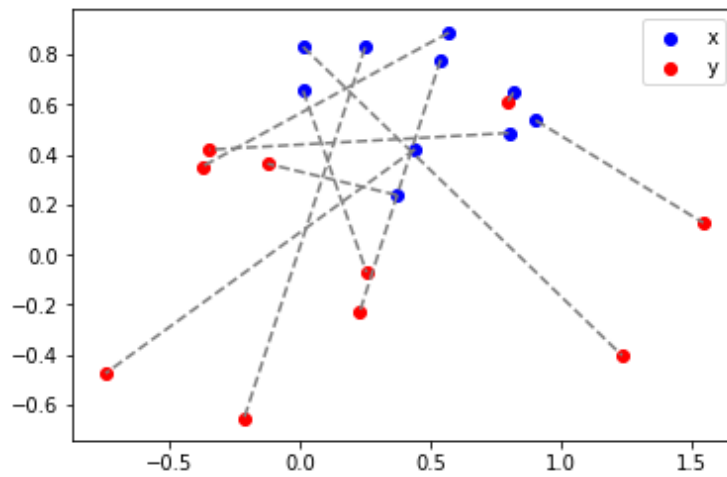


FIGURE 5 – Représentation et reliure des  $x_i$  et des  $y_i$

## Problème dual et fonction log-sum-exp

On considère la fonction log-sum-exp  $f : \mathbb{R}^N \rightarrow \mathbb{R}_+$  définie comme :

$$f(x) = \log \left( \sum_{i=1}^N \exp(x_i) \right)$$

Cette fonction est convexe. En effet, la hessienne  $H(x)$  de  $f$  est définie comme suit

$$H_{ij} = \frac{-e^{x_i} e^{x_j}}{\left( \sum_{k=1}^N e^{x_k} \right)^2} \quad H_{ii} = \frac{e^{x_i} \sum_{j \neq i}^N e^{x_j}}{\left( \sum_{k=1}^N e^{x_k} \right)^2}$$

Ainsi on observe que  $\forall i \in \llbracket 1 : N \rrbracket$  :

$$0 < H_{ii} = \sum_{j \neq i}^N |H_{ij}| := R_i$$

Ainsi les disques de Gerschgorin de la matrice  $H$   $D_i(H)$   $i \in \llbracket 1 : N \rrbracket$  définis comme :

$$D_i(H) = \left\{ x \in \mathbb{R} \mid |H_{ii} - x| \leq \sum_{j \neq i} |H_{ij}| \right\} = D(H_{ii}, R_i) \subset \mathbb{R}_+$$

$$\implies \bigcup_{i=1}^N D_i(H) \subset \mathbb{R}_+$$

Et donc par le théorème de Gerschgorin on sait que le spectre de la hessienne est inclus dans  $\mathbb{R}_+$ , donc elle est semi-définie positive et donc  $f$  est convexe.

On peut aussi remarquer que :

$$\max_i x_i \leq \log \left( \sum_{i=1}^N \exp(x_i) \right) \leq \max_i x_i + \log N$$

En effet :

$$\begin{aligned} x_i &\leq \max_i(x_i) \\ \implies \exp(x_i) &\leq \exp(\max_i(x_i)) \\ \implies \sum_{i=1}^N \exp(x_i) &\leq N \exp(\max_i x_i) \\ \implies \log \left( \sum_{i=1}^N \exp(x_i) \right) &\leq \max_i(x_i) + \log N \end{aligned}$$

Et :

$$\begin{aligned} \exp \left( \max_i(x_i) \right) &\leq \sum_{i=1}^N \exp(x_i) \\ \implies \max_i(x_i) &\leq \log \left( \sum_{i=1}^N \exp(x_i) \right) \end{aligned}$$

Ainsi on a :

$$\begin{aligned}\varepsilon \max_i \left( \frac{x_i}{\varepsilon} \right) &\leq \varepsilon \log \left( \sum_{i=1}^N \exp \left( \frac{x_i}{\varepsilon} \right) \right) \leq \varepsilon \left( \max_i \left( \frac{x_i}{\varepsilon} \right) + \log N \right) \\ \iff \max_i (x_i) &\leq \varepsilon \log \left( \sum_{i=1}^N \exp \left( \frac{x_i}{\varepsilon} \right) \right) \leq \max_i (x_i) + \varepsilon \log N \\ \implies f_\varepsilon &\rightarrow \max_i (x_i) \quad \varepsilon \rightarrow 0\end{aligned}$$

En se focalisant sur le problème dual on peut remarquer que

$$v_j = \min_i (C_{ij} - u_i) \quad \forall j \in J$$

Et donc on peut régulariser le problème dual, en effet :

$$\sum_i u_i \mu_i + \sum_j v_j \nu_j = \sum_i u_i \mu_i + \sum_j \min_k (C_{kj} - u_k) \nu_j$$

Ainsi on considère

$$\sum_i u_i \mu_i - \sum_j \varepsilon \log \left( \sum_{k=1}^N \exp \left( \frac{u_k - C_{kj}}{\varepsilon} \right) \right) \nu_j = \sum_i u_i \mu_i - \sum_j f_\varepsilon(u - C_{:j}) \nu_j$$

Ainsi lorsque  $\varepsilon \rightarrow 0$

$$\begin{aligned}\sum_i u_i \mu_i - \sum_j \varepsilon \log \left( \sum_{k=1}^N \exp \left( \frac{u_k - C_{kj}}{\varepsilon} \right) \right) \nu_j &\rightarrow \sum_i u_i \mu_i - \sum_j \max_k (-C_{kj} + u_k) \nu_j \\ &= \sum_i u_i \mu_i + \sum_j \min_k (C_{kj} - u_k) \nu_j = \sum_i u_i \mu_i + \sum_j v_j \nu_j\end{aligned}$$

Ce qui est équivalent au critère du problème dual. De plus pour les contraintes

$$u_i + v_j \leq C_{ij} \iff u_i + \min_i (C_{ij} - u_i) \leq C_{ij} \implies \min_i (C_{ij} - u_i) \leq C_{ij} - u_i \quad \forall j \in J$$

Ainsi les contraintes étant toujours vérifiées on obtient un problème régularisé en fonction de seule la variable  $u$  et sans contraintes qui converge bien vers le problème dual précédent pour  $\varepsilon \rightarrow 0$

$$\begin{aligned}\max_{u \in \mathbb{R}^N} \sum_i u_i \mu_i - \sum_j \varepsilon \log \left( \sum_{k=1}^N \exp \left( \frac{u_k - C_{kj}}{\varepsilon} \right) \right) \nu_j \\ v_j = \min_i (C_{ij} - u_i) \quad \forall j \in J\end{aligned}$$

On calcule alors le gradient du critère :

$$(\nabla_{\text{Critere}})_i = \mu_i - \sum_j \frac{\nu_j \exp \left( \frac{u_i - C_{ij}}{\varepsilon} \right)}{\sum_{k=1}^N \exp \left( \frac{u_k - C_{kj}}{\varepsilon} \right)}$$

On programme ainsi une fonction python `GradDualReg(mu, nu, C, eps, t)` afin d'appliquer une descente de gradient à pas fixe sur le critère qui retourne les potentiels de Kantorovich.

Nous appliquons cette méthode pour différents `eps` afin de montrer la convergence de la méthode vers les solutions obtenues à travers l'algorithme du simplexe quand  $\epsilon \rightarrow 0$ . Voici les résultats obtenus pour  $\epsilon = 1, 0.5, 0.1, 0.05, 0.01, 0.005$  :

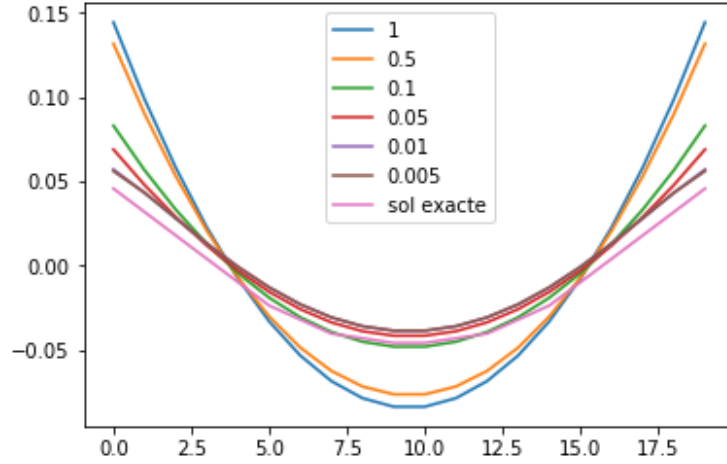


FIGURE 6 – Potentiels  $u$

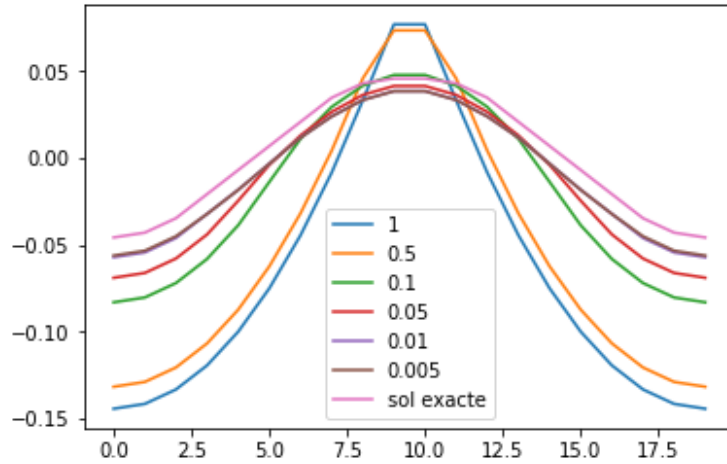


FIGURE 7 – Potentiels  $v$

On observe bien une convergence vers la solution du simplexe.

## Régularisation entropique

On s'intéresse maintenant à la régularisation du problème primal en pénalisant la contrainte de positivité en ajoutant un terme d'entropie : on va minimiser le critère suivant

$$F(\gamma) := \sum_{ij} C_{ij} \gamma_{ij} + \varepsilon \text{Ent}(\gamma)$$



où  $\varepsilon > 0$  est un paramètre de régularisation et  $\text{Ent} : \mathbb{R}_+^{N \times N} \rightarrow \mathbb{R}$  est l'entropie qu'on définit comme

$$\text{Ent}(\eta) = \sum_{ij} \eta_{ij} (\log(\eta_{ij}) - 1)$$

Montrons que  $F$  est convexe : d'abord le terme  $\sum_{ij} C_{ij} \gamma_{ij}$  est linéaire en  $\gamma$  donc il est convexe, ensuite réécrivons la fonction  $\text{Ent}$  avec le changement d'indice  $k = (i - 1)N + j \in \llbracket 1, N^2 \rrbracket$  :

$$\text{Ent}(\gamma) = \sum_k x_k (\log(x_k) - 1)$$

Supposons de plus que  $\gamma$  est strictement positive, c'est-à-dire que tous les coefficients de  $\mathbf{x}$  sont strictement positifs.  $\text{Ent}$  étant deux fois différentiable, on a son gradient :

$$(\nabla \text{Ent}(\gamma))_i = \partial_i \text{Ent}(\gamma) = \sum_k \delta_{ik} \log(x_k) = (\log(x_i))_i$$

et sa Hessienne :

$$(D^2 \text{Ent}(\gamma))_{ij} = \partial_j \partial_i \text{Ent}(\gamma) = \sum_k \delta_{jk} \delta_{ik} \frac{1}{x_k} = \left( \frac{\delta_{ij}}{x_i} \right)_{ij}$$

$D^2 \text{Ent}(\gamma)$  est diagonale et positive donc définie positive.

Enfin en supposant qu'il existe  $k \in \llbracket 1, N^2 \rrbracket$  tel que  $x_k = 0$ , alors par croissance comparée de  $x \mapsto x$  et  $x \mapsto \log x$ , on a que  $x_k \log(x_k)$  existe et vaut 0 ce qui réduit le  $k$ -ième du gradient et de la diagonale de la Hessienne à 0 et donc ne modifie pas son caractère définie positive.

$\text{Ent}$  est donc convexe et comme  $\varepsilon > 0$ , on obtient que  $F$  est convexe.

Maintenant posons  $\bar{\gamma}_{ij} = \exp\left(\frac{-C_{ij}}{\varepsilon}\right)$ , la fonction  $F$  peut donc se réécrire

$$\begin{aligned} F(\gamma) &= \sum_{ij} \gamma_{ij} \left( -\varepsilon \frac{-C_{ij}}{\varepsilon} + \varepsilon (\log(\gamma_{ij}) - 1) \right) \\ &= \sum_{ij} \varepsilon \gamma_{ij} (-\log(\bar{\gamma}_{ij}) + \log(\gamma_{ij}) - 1) \end{aligned}$$

et ainsi se mettre sous la forme suivante :

$$F(\gamma) = \sum_{ij} \varepsilon \gamma_{ij} \left( \log\left(\frac{\gamma_{ij}}{\bar{\gamma}_{ij}}\right) - 1 \right)$$

dont le gradient s'écrit

$$\nabla F(\gamma)_{ij} = \varepsilon \log\left(\frac{\gamma_{ij}}{\bar{\gamma}_{ij}}\right)$$

Le problème  $(\mathcal{MK})$  devient

$$\min \left\{ F(\gamma) \mid \sum_{j=1}^N \gamma_{ij} = \mu_i, \sum_{i=1}^N \gamma_{ij} = \nu_j \right\} \quad (\text{H})$$

Le Lagrangien de ce problème s'écrit

$$l(\gamma, \mathbf{u}, \mathbf{v}) = F(\gamma) - \sum_{i=1}^N u_i \left( \sum_{j=1}^N \gamma_{ij} - \mu_i \right) - \sum_{j=1}^N v_j \left( \sum_{i=1}^N \gamma_{ij} - \nu_j \right)$$

où  $u_i$  et  $v_j$  sont les multiplicateurs de Lagrange. On a plus simplement

$$l(\gamma, \mathbf{u}, \mathbf{v}) = F(\gamma) + \langle \mathbf{u}, \mu \rangle + \langle \mathbf{v}, \nu \rangle - \sum_{ij} \gamma_{ij} (u_i + v_j)$$

On écrit alors la condition d'optimalité d'ordre 1 du problème primal :

$$\nabla_{\gamma} l(\gamma, \mathbf{u}, \mathbf{v}) = \nabla F(\gamma) - (\mathbf{u} + \mathbf{v}) = 0$$

dont on déduit que la solution  $\gamma_{ij}^*$  vérifie

$$\varepsilon \log \left( \frac{\gamma_{ij}^*}{\bar{\gamma}_{ij}} \right) = u_i + v_j$$

et en notant  $a_i = \exp \left( \frac{u_i}{\varepsilon} \right)$  et  $b_j = \exp \left( \frac{v_j}{\varepsilon} \right)$ , on a que pour tout  $i, j$  :

$$\gamma_{ij}^* = a_i b_j \bar{\gamma}_{ij}$$

En imposant de plus que  $\gamma^*$  doit satisfaire les marginales, on obtient

$$a_i \sum_j b_j \bar{\gamma}_{ij} = \mu_i \quad \text{et} \quad b_j \sum_i a_i \bar{\gamma}_{ij} = \nu_j$$

soit les équations connues sous le nom d'équations de Bernstein-Schrödinger :

$$a_i = \frac{\mu_i}{\sum_j b_j \bar{\gamma}_{ij}} \quad \text{et} \quad b_j = \frac{\nu_j}{\sum_i a_i \bar{\gamma}_{ij}}$$

Étant donné l'expression du Lagrangien  $l(\gamma, \mathbf{u}, \mathbf{v})$  donnée précédemment, le problème dual s'écrit :

$$\max_{\mathbf{u}, \mathbf{v} \in \mathbb{R}^N} \min_{\gamma \in \mathbb{R}_+^{N \times N}} l(\gamma, \mathbf{u}, \mathbf{v})$$

Le Lagrangien étant convexe pour la variable  $\gamma$ , son minimum est alors atteint lorsque son gradient par rapport à  $\gamma$  s'annule, soit en  $\gamma^*$  donc le problème dual se note encore :

$$\max_{\mathbf{u}, \mathbf{v}} F(\gamma^*)$$

or d'après l'expression de  $\gamma^*$  et des  $a_i$  et  $b_j$  :

$$\begin{aligned} F(\gamma^*) &= \sum_{ij} \varepsilon a_i b_j \bar{\gamma}_{ij} \left( \log \left( \frac{a_i b_j \bar{\gamma}_{ij}}{\bar{\gamma}_{ij}} \right) - 1 \right) \\ &= \sum_{ij} a_i b_j \bar{\gamma}_{ij} (u_i + v_j - \varepsilon) \\ &= \sum_i u_i a_i \sum_j b_j \bar{\gamma}_{ij} + \sum_j v_j b_j \sum_i a_i \bar{\gamma}_{ij} - \varepsilon \sum_{ij} \bar{\gamma}_{ij} \\ &= \sum_i u_i \mu_i + \sum_j v_j \nu_j - \varepsilon \sum_{ij} e^{(u_i + v_j - C_{ij})/\varepsilon} \end{aligned}$$

d'après les équations de Bernstein-Schrödinger. On a donc :

$$\max_{\mathbf{u}, \mathbf{v}} F(\gamma^*) = \max_{\mathbf{u}, \mathbf{v}} \left[ \langle \mathbf{u}, \mu \rangle + \langle \mathbf{v}, \nu \rangle - \varepsilon \sum_{ij} e^{(u_i + v_j - C_{ij})/\varepsilon} \right]$$

$$\begin{cases} = +\infty & \text{si } \exists i, \mu_i \leq 0 \text{ ou } \exists j, \nu_j \leq 0 \\ \in \mathbb{R} & \text{si } \mu \geq 0 \text{ et } \nu \geq 0 \end{cases}$$

Ainsi, on donne le problème dual suivant :

$$\max_{\mathbf{u}, \mathbf{v}} \left\{ \langle \mathbf{u}, \mu \rangle + \langle \mathbf{v}, \nu \rangle - \varepsilon \sum_{ij} e^{(u_i + v_j - C_{ij})/\varepsilon} \mid \mu \geq 0 \text{ et } \nu \geq 0 \right\}$$

On applique maintenant un algorithme de descente de gradient pour résoudre le problème dual. On programme une fonction `EntDual(mu, nu, C, eps)` retournant les potentiels de Kantorovich pour le paramètre de régularisation  $\varepsilon$ .

Voici les résultats obtenus :

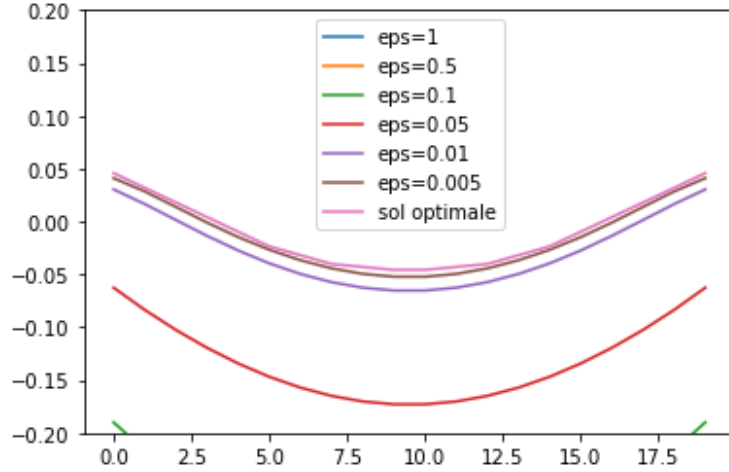


FIGURE 8 – Potentiels  $u$

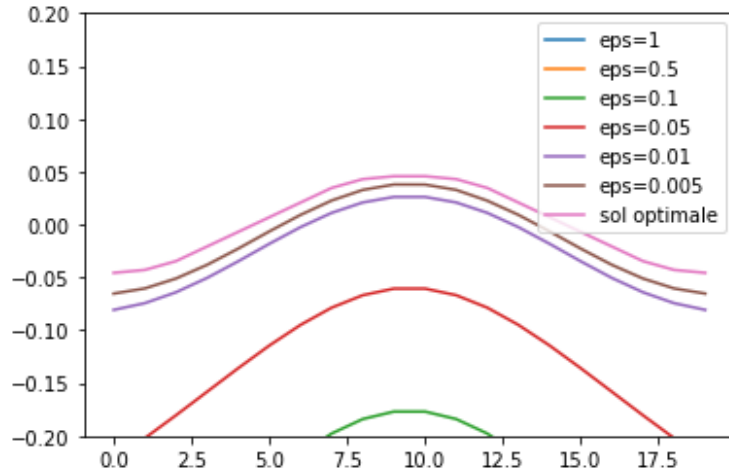


FIGURE 9 – Potentiels  $v$

On observe bien la convergence vers la solution optimale du simplexe pour des  $\varepsilon$  de plus en plus petits.

Ci dessous la solution  $\gamma$  obtenue pour  $\varepsilon = 0.005$

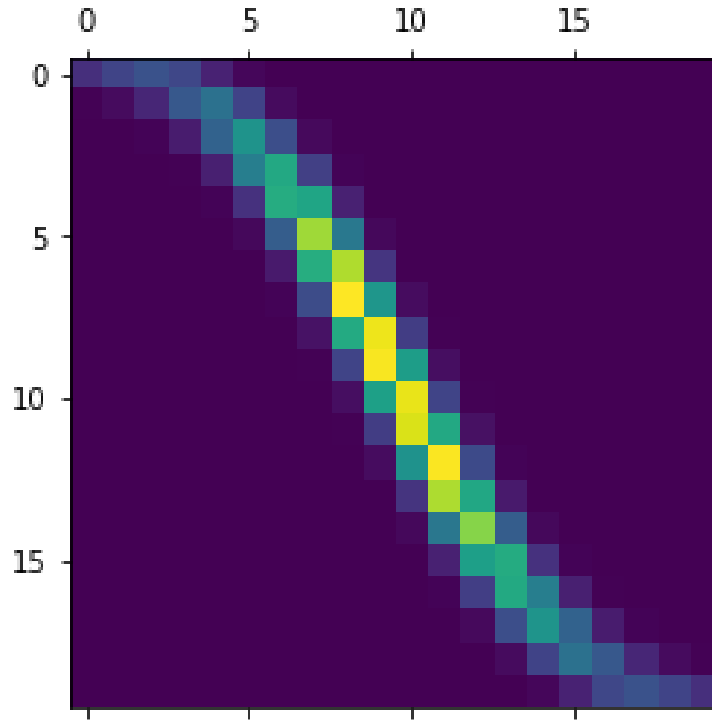


FIGURE 10 – Solution primale  $\gamma$

Finalement nous testons une méthode alternative : l'algorithme de Sinkhorn suivant :

$$a_i^{n+1} = \frac{\mu_i}{\sum_j b_j^n \gamma_{ij}} \quad b_j^{n+1} = \frac{\nu_j}{\sum_i a_i^{n+1} \gamma_{ij}}$$

Nous le programmons à travers la fonction Python `Sinkhorn(mu, nu, C, eps)` et retourne les potentiels  $u, v, \gamma^*$ . Voici les résultats obtenu :

On observe bien la convergence de l'algorithme de Sinkhorn vers les potentiels optimaux du simplexe pour un  $\varepsilon$  tendant vers 0

En conclusion on peut observer en calculant l'erreur entre la solution de chaque algorithme des problèmes régularisés et celle du simplexe que on a bien convergence lorsque epsilon tend vers zero. En effet en traçant ces erreurs on obtient :

On observe que plus epsilon est petit plus l'erreur entre les solutions des differents problèmes régularisés par rapport à la solution du simplexe diminue.

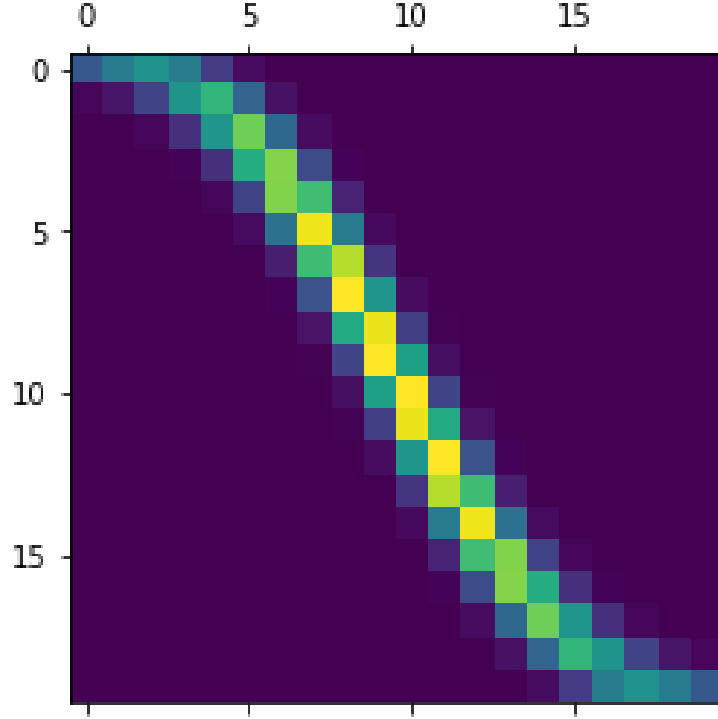


FIGURE 11 – Solution primale  $\gamma$  avec Sinkhorn

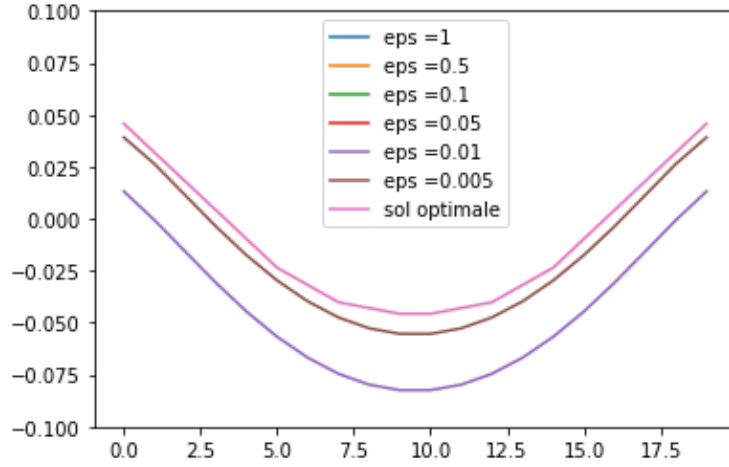


FIGURE 12 – Solution  $u^*$  avec Sinkhorn

## Transport multi-marges et l'interaction électron-électron

On traite maintenant le cas du problème du transport optimal multi-marges. Ne considérons plus uniquement les boulangeries et les cafés, mais prenons également en compte les hôtels, les restaurants, etc. et cherchons un couplage  $\gamma \in \otimes^K \mathbb{R}_+^K$  ( $\gamma$  est maintenant un tenseur  $K$  dimensionnel et on notera les composantes comme  $\gamma_{i_1, \dots, i_K}$ ) nous indiquant la quantité de pains au chocolat envoyée par la boulangerie  $i_1$  au café  $i_2$  à l'hôtel  $i_3$  etc. Cet exemple peut paraître simple, mais nous pouvons remarquer que le transport optimal multi-marges apparaît plus général que le transport optimal classique, car il nous permet de modéliser l'interaction entre les boulangeries et tous les clients possibles.

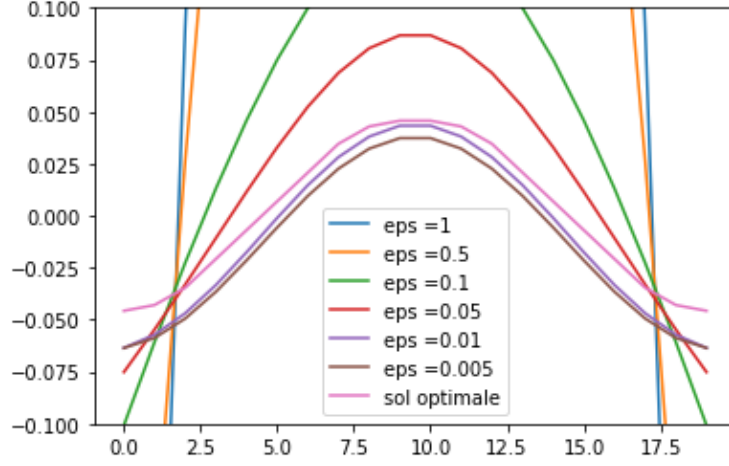


FIGURE 13 – Solution primale  $v^*$

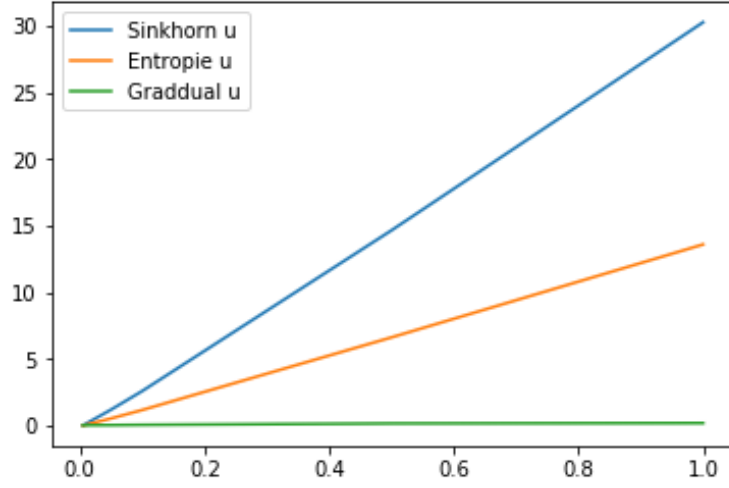


FIGURE 14 – Erreur par rapport à la solution du simplexe pour différents epsilon pour le potentiel  $u$

Dans le cas où on prend comme coût l'interaction de Coulomb

$$C_{i_1, \dots, i_K} = c(x_{i_1}, \dots, x_{i_K}) = \sum_{i \neq j}^K \frac{1}{|x_{i_i} - x_{i_j}|}$$

et  $K$  marginales  $\mu^k(x_{i_k})$ ,  $k \in \llbracket 1, K \rrbracket$  sont toutes égales à  $\rho(x_i)$  (qui dénote la densité électronique : la probabilité de trouver un électron en  $x_i$ ) alors le problème du transport multi-marge nous donne l'énergie de répulsion entre les électrons et le couplage optimal  $\gamma_{i_1, \dots, i_K}^*$  la probabilité de trouver l'électron  $k$  à la position  $i_k$ .

On se place dans le cas  $K = 3$ , le problème de Monge-Kantorovich s'écrit sous la forme :

$$\min \left\{ \sum_{i_1, i_2, i_3} C_{i_1, i_2, i_3} \gamma_{i_1, i_2, i_3} \mid \gamma \in \Pi(\rho(x_{i_k})) \right\}, \quad (\mathcal{MK})$$

où

$$\Pi(\rho(x_{i_k})) := \left\{ \gamma \in \mathbb{R}_+^{N^3} \mid \sum_{i_2, i_3} \gamma_{i_1, i_2, i_3} = \rho(x_{i_1}), \sum_{i_1, i_3} \gamma_{i_1, i_2, i_3} = \rho(x_{i_2}), \text{ et } \sum_{i_1, i_2} \gamma_{i_1, i_2, i_3} = \rho(x_{i_3}) \right\}$$

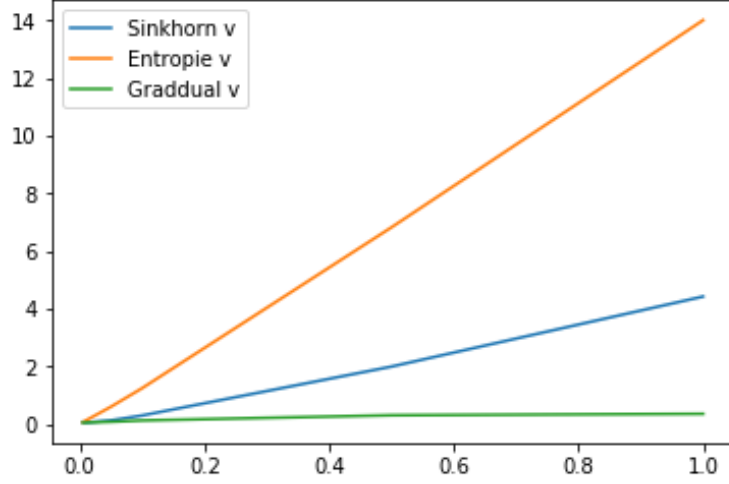


FIGURE 15 – Erreur par rapport à la solution du simplexe pour différents epsilon pour le potentiel v

Le Lagrangien de ce problème s'écrit

$$l(\gamma, u_{i_1}, u_{i_2}, u_{i_3}) = \sum_{i_1, i_2, i_3} C_{i_1, i_2, i_3} \gamma_{i_1, i_2, i_3} - \sum_{k=1}^3 u_{i_k} \left( \sum_{i \neq j \neq k} \gamma_{i, i_j, i_k} - \rho(x_{i_k}) \right)$$

ainsi

$$\begin{aligned} \min_{\gamma \in \mathbb{R}_+^{N^3}} l(\gamma, u_{i_1}, u_{i_2}, u_{i_3}) &= \min_{\gamma \geq 0} \left[ \sum_{i_1, i_2, i_3} \gamma_{i_1, i_2, i_3} (C_{i_1, i_2, i_3} - (u_{i_1} + u_{i_2} + u_{i_3})) \right] + \sum_{k=1}^3 u_{i_k} \rho(x_{i_k}) \\ &= \begin{cases} -\infty & \text{si } C_{i_1, i_2, i_3} \leq u_{i_1} + u_{i_2} + u_{i_3} \\ \sum_{k=1}^3 u_{i_k} \rho(x_{i_k}) & \text{si } C_{i_1, i_2, i_3} \geq u_{i_1} + u_{i_2} + u_{i_3} \end{cases} \end{aligned}$$

et donc le problème dual s'écrit

$$\max_{u_{i_k} \in \mathbb{R}^N} \left\{ \sum_{k=1}^3 u_{i_k} \rho(x_{i_k}) \mid \forall (i_1, i_2, i_3) \in \llbracket 1, N \rrbracket^3, u_{i_1} + u_{i_2} + u_{i_3} \leq C_{i_1, i_2, i_3} \right\}$$

On s'intéresse maintenant à la régularisation du problème primal : on va minimiser le critère suivant

$$F(\gamma) := \sum_{i_1, i_2, i_3} C_{i_1, i_2, i_3} \gamma_{i_1, i_2, i_3} + \varepsilon \text{Ent}(\gamma)$$

où  $\varepsilon > 0$  et  $\text{Ent} : \mathbb{R}_+^{N^3} \rightarrow \mathbb{R}$  est telle que

$$\text{Ent}(\gamma) = \sum_{i_1, i_2, i_3} \gamma_{i_1, i_2, i_3} (\log(\gamma_{i_1, i_2, i_3}) - 1)$$

On montre de la même façon que dans la Partie précédente que  $F$  est convexe et qu'en notant

$\bar{\gamma}_{i_1, i_2, i_3} = \exp\left(\frac{-C_{i_1, i_2, i_3}}{\varepsilon}\right)$ ,  $F$  peut se mettre sous la forme suivante :

$$F(\gamma) = \sum_{i_1, i_2, i_3} \varepsilon \gamma_{i_1, i_2, i_3} \left( \log\left(\frac{\gamma_{i_1, i_2, i_3}}{\bar{\gamma}_{i_1, i_2, i_3}}\right) - 1 \right)$$

dont le gradient s'écrit

$$\nabla F(\gamma)_{i_1, i_2, i_3} = \varepsilon \log \left( \frac{\gamma_{i_1, i_2, i_3}}{\bar{\gamma}_{i_1, i_2, i_3}} \right)$$

Le problème régularisé est alors

$$\min_{\gamma} \left\{ F(\gamma) \mid \sum_{i_2, i_3} \gamma_{ij} = \rho(x_{i_1}), \sum_{i_1, i_3} \gamma_{i_1, i_2, i_3} = \rho(x_{i_2}), \sum_{i_1, i_2} \gamma_{i_1, i_2, i_3} = \rho(x_{i_3}) \right\} \quad (\text{H})$$

Le Lagrangien de ce problème s'écrit

$$l(\gamma, u_{i_k}) = F(\gamma) - \sum_{k=1}^3 u_{i_k} \left( \sum_{i \neq j \neq k} \gamma_{i, i_j, i_k} - \rho(x_{i_k}) \right)$$

où  $u_{i_k, 1 \leq k \leq 3}$  sont les multiplicateurs de Lagrange. On écrit alors la condition d'optimalité d'ordre 1 du problème primal :

$$\nabla_{\gamma} l(\gamma, u_{i_k}) = \nabla F(\gamma) - (u_{i_1} + u_{i_2} + u_{i_3}) = 0$$

dont on déduit la solution  $\gamma_{i_1, i_2, i_3}^*$  en fonction de  $a_{i_1} = \exp\left(\frac{u_{i_1}}{\varepsilon}\right)$ ,  $b_{i_2} = \exp\left(\frac{u_{i_2}}{\varepsilon}\right)$  et  $c_{i_3} = \exp\left(\frac{u_{i_3}}{\varepsilon}\right)$  :

$$\gamma_{i_1, i_2, i_3}^* = a_{i_1} b_{i_2} c_{i_3} \bar{\gamma}_{i_1, i_2, i_3}$$

En imposant de plus que  $\gamma^*$  doit satisfaire les marginales, on obtient

$$a_{i_1} \sum_{i_2, i_3} b_{i_2} c_{i_3} \bar{\gamma}_{i_1, i_2, i_3} = \rho_{i_1}, \quad b_{i_2} \sum_{i_1, i_3} a_{i_1} c_{i_3} \bar{\gamma}_{i_1, i_2, i_3} = \rho_{i_2}, \quad c_{i_3} \sum_{i_1, i_2} a_{i_1} b_{i_2} \bar{\gamma}_{i_1, i_2, i_3} = \rho_{i_3}$$

soit les équations suivantes :

$$a_{i_1} = \frac{\rho_{i_1}}{\sum_{i_2, i_3} b_{i_2} c_{i_3} \bar{\gamma}_{i_1, i_2, i_3}}, \quad b_{i_2} = \frac{\rho_{i_2}}{\sum_{i_1, i_3} a_{i_1} c_{i_3} \bar{\gamma}_{i_1, i_2, i_3}}, \quad c_{i_3} = \frac{\rho_{i_3}}{\sum_{i_1, i_2} a_{i_1} b_{i_2} \bar{\gamma}_{i_1, i_2, i_3}}$$

Étant donné l'expression du Lagrangien  $l(\gamma, u_{i_k})$  donnée précédemment, le problème dual s'écrit :

$$\max_{u_{i_k}} \min_{\gamma \geq 0} l(\gamma, u_{i_k})$$

Le Lagrangien étant convexe pour la variable  $\gamma$ , son minimum est alors atteint lorsque son gradient par rapport à  $\gamma$  s'annule, soit en  $\gamma^*$  donc le problème dual se note encore :

$$\max_{u_{i_k}} F(\gamma^*)$$

puis avec un calcul similaire à celui de la Partie précédente et les expressions de  $\gamma^*$ ,  $a_{i_1}$ ,  $b_{i_2}$  et  $c_{i_3}$ , on obtient :

$$\begin{aligned} \max_{u_{i_k}} F(\gamma^*) &= \max_{u_{i_k}} \left[ \sum_{k=1}^3 u_{i_k} \rho(x_{i_k}) - \varepsilon \sum_{i_1, i_2, i_3} e^{(u_{i_1} + u_{i_2} + u_{i_3} - C_{i_1, i_2, i_3})/\varepsilon} \right] \\ &\begin{cases} = +\infty & \text{si } \exists k, \exists i_k \rho(x_{i_k}) \leq 0 \\ \in \mathbb{R} & \text{si } \forall k, \rho(x_{i_k}) \geq 0 \end{cases} \end{aligned}$$



Ainsi, on donne le problème dual suivant :

$$\max_{u_{i_k}} \left\{ \sum_{k=1}^3 u_{i_k} \rho(x_{i_k}) - \varepsilon \sum_{i_1, i_2, i_3} e^{(u_{i_1} + u_{i_2} + u_{i_3} - C_{i_1, i_2, i_3})/\varepsilon} \mid \forall k, \rho(x_{i_k}) \geq 0 \right\}$$

Enfin, d'après les équations précédentes, l'algorithme de Sinkhorn pour ce problème s'écrit avec l'itération pour  $n \in \mathbb{N}$  :

$$\begin{cases} a_{i_1}^{n+1} = \frac{\rho_{i_1}}{\sum_{i_2, i_3} b_{i_2}^n c_{i_3}^n \bar{\gamma}_{i_1, i_2, i_3}} \\ b_{i_2}^{n+1} = \frac{\rho_{i_2}}{\sum_{i_1, i_3} a_{i_1}^{n+1} c_{i_3}^n \bar{\gamma}_{i_1, i_2, i_3}} \\ c_{i_3}^{n+1} = \frac{\rho_{i_3}}{\sum_{i_1, i_2} a_{i_1}^{n+1} b_{i_2}^{n+1} \bar{\gamma}_{i_1, i_2, i_3}} \end{cases}$$

Cherchons maintenant à décomposer  $\bar{\gamma}_{i_1, i_2, i_3}$  tel que les sommes dans les équations de Bernstein-Schrödinger peuvent décrire comme des produits matrices-vecteurs : i.e.  $a\tilde{\gamma}b\tilde{\gamma}$  où  $\tilde{\gamma} \in \mathbb{R}^{N \times N}$ . On a pour  $i_1 = i$  fixé :

$$\sum_{i_2, i_3} b_{i_2} c_{i_3} \bar{\gamma}_{i_1, i_2, i_3} = \sum_{i_2, i_3} b_{i_2} \exp\left(-\frac{e_{i_2, i_3}}{\varepsilon}\right) c_{i_3} \exp\left(-\frac{e_{i_2, i_3}}{\varepsilon}\right)$$

où  $e_{i_2, i_3} = \left( \frac{1}{|x_i - x_{i_2}|} + \frac{1}{|x_i - x_{i_3}|} + \frac{1}{|x_{i_2} - x_{i_3}|} \right)$ .

On peut penser que  $\tilde{\gamma} = \exp\left(\frac{e_{i_2, i_3}}{\varepsilon}\right)$  fonctionne mais le terme en  $\exp\left(-\frac{1}{\varepsilon|x_{i_2} - x_{i_3}|}\right)$  empêche la séparation des sommes sur les indices  $i_2$  et  $i_3$ .