COMPUTATIONAL OPTIMAL TRANSPORT

—

QUADRATICALLY-REGULARIZED OPTIMAL TRANSPORT ON GRAPHS

Théophane Gregoir

Course directed by Gabriel Peyré

MATHÉMATIQUES
VISION
APPRENTISSAGE

école
normale
supérieure
paris–saclay

université
PARIS-SACLAY

ABSTRACT

Directed graphs represent one of the most general and intuitive framework for studying optimal transportation problems. In this article, authors focus on regularized transport and especially quadratic regularization instead of well-known entropic regularization. Indeed, the most classical regularization method for optimal transportation problem is entropic regularization which can be solved very efficiently thanks to Sinkhorn alternating projection algorithm. However, as exhibited in the Numerical Tours, this method can encounter major numerical issues in small regularization regime because of its strictly positive transportation solution. Therefore, unlike quadratic one, entropic regularization is not adapted to sparse transport, useful in many applications. In this work, authors demonstrate theoretical results providing control properties on quadratically regularized transport solution in low regularization regime and linking them to unregularized solutions. Besides, the dual problem of quadratically-regularized transport problem is stated and studied. Finally, this dual problem is used at the center of a new optimization algorithm for quadratic regularization on graph which is introduced and tested.

Contents

1 Introduction

General framework

Directed graphs represent a general framework for optimal transport as most of the discrete transportation problems can be formulated in a graph form : cloud of points, images... In particular, bipartite graph can model any transport between two distinct discrete sets. Therefore, results proven in this paper are in fact applicable to all transportation problems between discrete distributions.

Definition of the unregularized problem

Given a directed graph $G = (V, E)$ with n_v vertices, n_e oriented edges and $E \subseteq V \times V$, we consider a positive cost distributed on all edges $c \in \mathbb{R}_+^{n_e}$. Then, we introduce the incidence matrix $D \in \{-1, 0, 1\}^{n_e \times n_v}$ defined as :

$$D_{ev} = \begin{cases} -1 & \text{if } e = (v, .) \\ 1 & \text{if } e = (., v) \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Given a starting distribution ρ_0 and goal distribution ρ_1 in the probability simplex over V, we can formulate the optimal transport problem by searching for an optimal $J \in \mathbb{R}_+^{n_e}$. In optimal transportation theory, the minimum defined in (2) corresponds to the Wasserstein distance between ρ_0 and ρ_1 linked to cost c, denoted $\mathcal{W}_c(\rho_0, \rho_1)$.

$$\begin{aligned} \min_J \quad & c^t J \\ \text{s.t.} \quad & J \geq 0 \\ & D^t J = \rho_1 - \rho_0 \end{aligned} \tag{2}$$

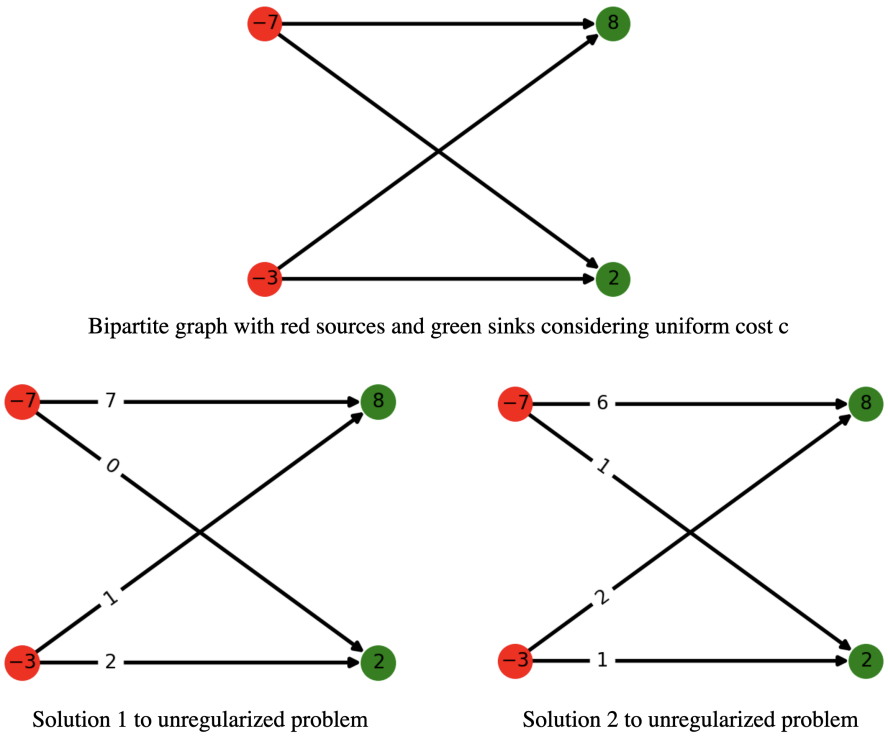
This formulation differs from the classical transportation representation as we are optimizing a vector J sized by the number of edges, n_e, instead of searching for an optimal transportation matrix living in $n_v \times n_v$. Using this framework is particularly relevant for memory and efficiency purposes in the case of very sparse problems (with a limited number of edges compared to the number of vertices).

Solving the unregularized problem

The most common way of solving the LP defined in (2) would be to use the Simplex algorithm introduced by Dantzig [1]. However, this method can be extremely costy, generally reaching polynomial complexity [9]. Furthermore, unregularized transport on directed graphs can have multiple optimal solutions (see Figure 1).

The role of regularization

Therefore, one of the most common method to overcome the previous limits consists in adding a regularization term to the objective. Transforming the objective into a strongly convex one, regularization ensures uniqueness of the solution and often enables the introduction of efficient solving algorithms.



Bipartite graph with red sources and green sinks considering uniform cost c

Solution 1 to unregularized problem

Solution 2 to unregularized problem

Figure 1: Illustration of non-uniqueness of solutions for unregularized optimal tranport problem on directed graph

Entropic regularization

Coming back to the classical formulation with a transportation plan variable T in $n_v \times n_v$, entropic regularization can be used by adding a term $\epsilon \sum_{v,w \in V} T_{vw} \ln(T_{vw})$ to the objective with $\epsilon > 0$. Besides, entropic regularization benefits from the existence of a powerful solving algorithm : Sinkhorn alternating projection algorithm based on dual analysis [8]. However, the entropic term enforces all elements in the transportation plan to be strictly positive ruling out sparse solutions and generating numerical issues in low regularization regime (cf. Numerical Tour 2 with ϵ too small). Furthermore, the non-optimality of the formulation of the entropic regularization in the special case of graph will be discussed in the core report.

Quadratic regularization

In this work, authors decided to explore a sparse-friendly regularization by studying the properties of quadratic regularization and its solutions. Given $\alpha > 0$, quadratically-regularized optimal transport on directed graphs can be formulated as follow.

$$\begin{aligned} \min_{J} \quad & c^t J + \frac{\alpha}{2} J^t J \\ \text{s.t.} \quad & J \geq 0 \\ & D^t J = \rho_1 - \rho_0 \end{aligned} \tag{3}$$

Previous works using quadratic regularization

In Biostatistics, quadratic regularization was introduced to overcome computational difficulties when fitting models with data including a significant number of variables and very few samples [5]. By promoting sparse solutions, quadratic regularizers drastically improved convergence, especially in gene expression arrays.

In a more standard set-up, quadratic regularizer was introduced in L_1 minimization problem [10]. The quadratic regularizer permitted to ensure sparsity of solutions of linear equations through the linearized Bregman method.

Then, in image analysis, quadratic regularization was introduced for pixel transport using Earth Mover's Distance (alias for Wasserstein distance in image analysis) [6]. In this work, the quadratic regularizer permits to guarantee uniqueness of the solution and integrates a primal-dual solving algorithm.

Contributions

In the article studied in this report [4], authors adapt quadratic regularizer to optimal transport theory in directed graphs. First, a study of the dual problem is conducted. Then, properties to estimate the deviation between regularized and unregularized solutions are stated and proven. Finally, authors propose a solving algorithm for quadratically regularized optimal transport based on dual problem resolution.

Subsequent works

More recently, quadratic regularization applied to optimal transport was further investigated [7]. In particular, a comparison between entropic and quadratic regularizations was presented.

2 Analysis and Numerics on Quadratically-Regularized Optimal Transport on Graphs

2.1 Dual Analysis

NB : Dual analysis is briefly reviewed in the article, encouraging us to solve it properly. The notations and tools used in this part are directly extracted from Alexandre d'Aspremont's course on Convex Optimization delivered during MVA first semester.

2.1.1 Building the dual problem

We start by stating the primal problem of quadratically regularized optimal transport as introduced in the introduction with $f = \rho_1 - \rho_0$.

$$\begin{aligned} \min_{J} \quad & c^t J + \frac{\alpha}{2} J^t J \\ \text{s.t.} \quad & J \geq 0 \\ & D^t J = f \end{aligned} \tag{4}$$

The objective being convex and the constraints being linear, this optimization problem is a convex problem. Given a strictly feasible point, Slater's condition ensures strong duality between primal and dual, motivating the dual analysis.

We consider the Lagrangian L and its dual variables : $\lambda \in \mathbb{R}^{n_e}$ and $\mu \in \mathbb{R}^{n_v}$.

$$\begin{aligned} L(J, \lambda, \mu) &= c^t J + \frac{\alpha}{2} J^t J - \lambda^t J + \mu^t (f - D^t J) \\ &= \frac{\alpha}{2} J^t J + (c - \lambda - D\mu)^t J + \mu^t f \end{aligned} \tag{5}$$

We introduce the dual function g.

$$g(\lambda, \mu) = \inf_J L(J, \lambda, \mu) \tag{6}$$

By cancelling the gradient, we minimize the convex function h defined by :

$$\begin{aligned} h(J) &= \frac{\alpha}{2} J^t J + (c - \lambda - D\mu)^t J \\ \nabla h(J) = 0 \iff J &= \frac{D\mu - c - \lambda}{\alpha} \end{aligned} \tag{7}$$

Therefore, we obtain the following dual problem.

$$\begin{aligned} \max_{\lambda, \mu} \quad & \mu^t f - \frac{1}{2\alpha} (D\mu - c + \lambda)^t (D\mu - c + \lambda) \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned} \tag{8}$$

By developing the scalar product in the dual objective, we can isolate a concave function in λ noted $q(\lambda)$. This function q can be maximized on the domain $\lambda \geq 0$.

$$\begin{aligned} q(\lambda) &= -\frac{1}{2\alpha}(2\lambda^t(D\mu - c) + \lambda^t\lambda) \\ \nabla q(\lambda) = 0 &\iff \lambda = -(D\mu - c) \end{aligned} \tag{9}$$

Validating $\lambda \geq 0$, we obtain the explicit definition of λ^*.

$$\lambda_i^* = \begin{cases} -(D\mu - c)_i & \text{if } (D\mu - c)_i \leq 0 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

Finally, we obtain the following unconstrained dual problem :

$$\begin{aligned} \max_{\mu} \quad & \mu^t f - \frac{1}{2\alpha}\|(D\mu - c)_+\|_2^2 \\ & \text{where } a_{+i} = \max(0, a_i), \forall i \in [1, n_e] \end{aligned} \tag{11}$$

2.1.2 Going back to primal problem

As strong duality holds, we can operate complementary slackness from KKT conditions and merge (7) with (10) in order to obtain a closed form for quadratically-regularized optimal transport solution J_α using dual solution μ_α.

$$J_\alpha = \frac{(D\mu_\alpha - c)_+}{\alpha} \tag{12}$$

2.2 Linking quadratically regularized and unregularized solutions

Using quadratic regularization ensures uniqueness of solution and gives access to more efficient solving algorithm. However, for this regularization to be useful, it is necessary to check how close quadratically regularized solutions can be to unregularized solutions. Therefore, the goal of this section is to provide insights on the proof of the following proposition, main result from the article.

Proposition 1. *There exists a constant $\alpha_\infty > 0$ depending on the graph G and data $f = \rho_1 - \rho_0$ such that for all $\alpha \in [0, \alpha_\infty]$, the solution J_α of quadratically regularized (4) is also a solution of unregularized (2).*

2.2.1 Illustration of the proposition

In this first part, a case study is presented to illustrate Proposition 1. A simple transport of mass through a directed graph from one point to the other is illustrated in Figure 2.

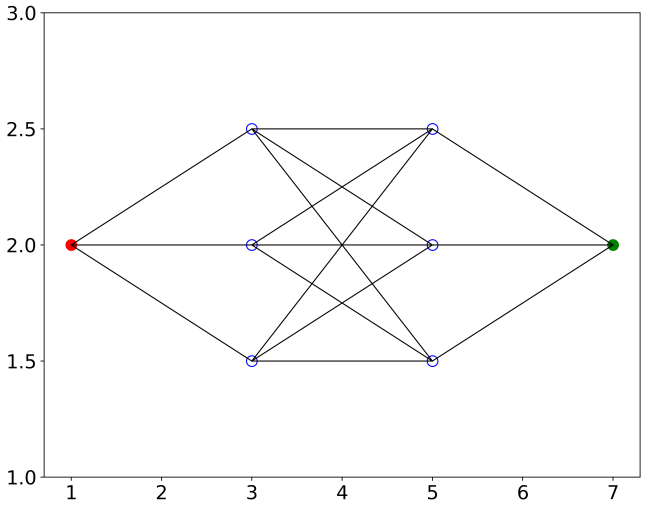


Figure 2: Illustration of initial situation of the graph. Edges are directed from left to right and their associated costs correspond to squared euclidean distances between nodes. The red node holds an initial mass of 1 while all other have a 0 initial mass. The green node holds a final mass of 1 while all other have a 0 final mass.

For this experiment, the unregularized optimal transport is clearly a unique central path for the entire mass. As α decreases, the quadratically-regularized optimal transport converges towards this unique central path. As shown in Figure 3, for $\alpha < 0.1$, J_α is also a solution of the unregularized optimal transport. Therefore, according to Proposition 1, for this particular problem, $\alpha_\infty \geq 0.1$. Moreover, another quadratic characteristic is clearly visible, as α decreases, J_α tends to be sparser.

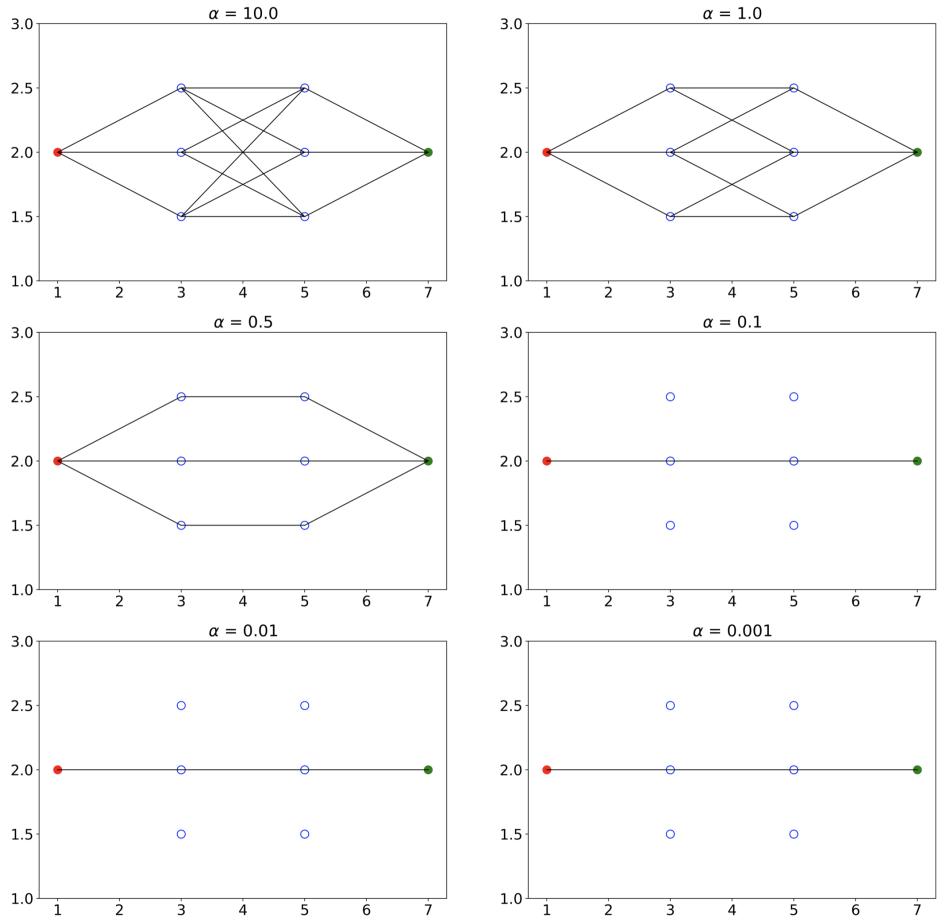


Figure 3: Representations of the edges used by quadratically-regularized optimal transport for decreasing parameter α. CVXPY QP solver was used to solve the regularized problems. Edges with a flow below 10^{-6} are considered to be 0.

2.2.2 Quadratically regularized solutions behaviour

In order to prove Proposition 1, we first need to explore which links can be established between quadratically regularized solutions corresponding to different parameters α. In the objective function of (4), it is possible to separate a L_1 term linked to the cost from the L_2 regularizer. Studying this separation, we can state Proposition 2 which is simple to prove thanks to the strict inequality for strongly convex function applied to the objective function.

Proposition 2. *Let $0 < \alpha_1 < \alpha_2$ and J_α the solution of (4) for parameter α. If $J_{\alpha_1} \neq J_{\alpha_2}$, then $c^t J_{\alpha_1} < c^t J_{\alpha_2}$ and $J_{\alpha_1}^t J_{\alpha_1} > J_{\alpha_2}^t J_{\alpha_2}$.*

In order to visualize this behaviour, numerical computations are provided in Figure 4 for a simple bipartite graph with two components of 5 vertices each.

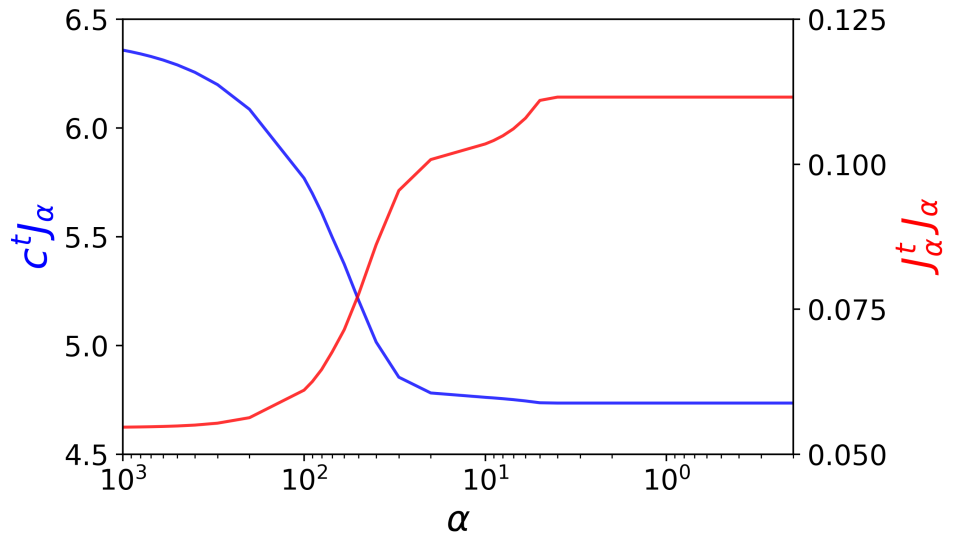


Figure 4: Given a bipartite graph with a source component and target component containing 5 vertices each, we observe the two parts of the objective function of (4) evaluated at their optimal solution J_α. The cost assigned to each edge is a random number between 1 and 10, the source ρ_0 and target ρ_1 are random and normalized. We can clearly see that the cost part is increasing with α reducing while the quadratic part follows an opposite behaviour.

2.2.3 Divergence-free decomposition

After highlighting that an optimal J_α cannot contain any cycle, the main concept introduced for the demonstration is divergence-free decomposition for two flows without cycles. Conceptually, it corresponds to the algorithmic creation of a path composed of elements R_k respecting $D^t R_k = 0$ between two flow vectors J_1 and J_2 both validating the equality constraint $D^t J = f$ from (4). We can note :

$$J_1 = J_2 + \sum_{k=1}^{n} \epsilon_k R_k \tag{13}$$

The algorithm implements an iterative construction of the decomposition.

Algorithm 1: Divergence-Free Decomposition Algorithm

for *k = 1,2,...* **do**
  Step 1 : Starting from a path where $J_1 \neq J_2$, build R_k so that it removes mass where $J_2 > J_1$ and it adds the same mass on a portion where $J_1 > J_2$ (ensured thanks to mass conservation).
  Step 2 : Set ϵ_k as being the minimum difference between flows on newly created paths. The path corresponding to this minimum difference will become an equality path for J_1 and J_2 after J_2 update.
  Step 3 : Update J_2 by adding $\epsilon_k R_k$.
  Step 4 : Stop if $J_1 = J_2$.
end

The algorithm progresses while keeping the divergence property on each R_k. At each step, the flows J_1 and J_2 are equal on one more path ensuring convergence.

2.2.4 Structure of the proof

Proposition 1 is proven by contradiction. It amounts to show that the following proposition is false :

$$\forall \alpha > 0, \exists\, \alpha_0 \in [0,\ \alpha] : J_{\alpha_0} \text{ solves (4) without solving (2).} \tag{14}$$

Given a graph G and a difference of distributions f, we start by choosing a $\alpha_0 \in [0,\ \alpha]$ for an α which we will select a posteriori. We also note J_0 a solution of unregularized (2) and J_{α_0} the solution of (4) for parameter α_0.

The contradiction proof is built in 4 main steps :

- Introduce the Divergence-Free Decomposition between J_{α_0} and J_0. We note : $J_0 = J_{\alpha_0} + \sum_{k=1}^{n} \epsilon_k R_k$.
- Perturb the flow J_{α_0} into $J^{'}$ by adding a chosen element R_k verifying $c^t R_k < 0$ (exists otherwise J_{α_0} solves (2)).
- Show that $J^{'}$ decreases the L_1 cost part of the objective.
- Choose α small enough to ensure $J^{'}$ is linked to a smaller minimizer than the supposed optimum linked to J_{α_0}. This choice is made so that L_1 cost dominates over the quadratic part. The contradiction is yielded.

2.3 Comparison between entropic and quadratic regularizations

2.3.1 Why quadratic regularization is more adapted than entropic one in the graph framework ?

Generally, entropic regularization is applied directly to the transportation plan $T \in \mathbb{R}^{n_v \times n_v}$. In order to apply this method in the graph framework, a matrix $C \in \mathbb{R}_+^{n_v \times n_v}$ of lowest cost path between each pair of vertices has to be calculated. This process can be done thanks to Dijkstra's algorithm [3] but can also become quite costy. Once done, the entropic regularization problem can be stated for the graph case given $\epsilon > 0$.

$$\begin{aligned} \min_{T \in \mathbb{R}^{n_v \times n_v}} \quad & \mathrm{Tr}(C^t T) + \epsilon \sum_{(v,w) \in V^2, v \neq w} T_{vw} \ln(T_{vw}) \\ \text{s.t.} \quad & T \geq 0 \\ & T\,1 = \rho_0 \\ & T^t\,1 = \rho_1 \end{aligned} \tag{15}$$

In fact, the formulation of the entropic regularization on $J \in \mathbb{R}_+^{n_e}$ would be as simple as the quadratic one. However, Sinkhorn algorithm being only suitable to transportation plans, this formulation would be clearly less serviceable.

Therefore, as quadratic regularization on $J \in \mathbb{R}_+^{n_e}$ corresponds to a classical QP problem (solved thanks to barrier method for example), graph framework appears to be more adapted for this regularization especially in the case of very sparse edges in a large graph ($n_e \ll n_v^2$).

2.3.2 Finding a metric to compare α and ϵ parameters

In order to analyse the two regularizations, we need to compare two approximately similar regimes of regularization. The most intuitive idea would be to measure the distance between the two differently regularized solutions and an unregularized solution. However, as discussed in the introduction, unregularized optimal transport can accept multiple solutions. Therefore, although regularized solutions may be converging towards an unregularized solution as the regularized term is reduced, it is not ensured it is the selected unregularized solution.

The most appropriate choice of metric seems to be the absolute difference between the optimal value of unregularized optimal transport and the evaluation of the unregularized objective function at the given regularized solution, J_α or J_ϵ.

$$d(\alpha) = |\, c^t J_\alpha \,-\, c^t J_0 \,| \text{ and } d(\epsilon) = |\, c^t J_\epsilon \,-\, c^t J_0 \,|$$
where J_0 is an unregularized optimal solution, J_α the α quadratic one and J_ϵ the ϵ entropic one. (16)

This choice of metric is confirmed by Figure 5. Indeed, the metric d defined above enables to measure a distance to the regularized problem which appears to be rel-

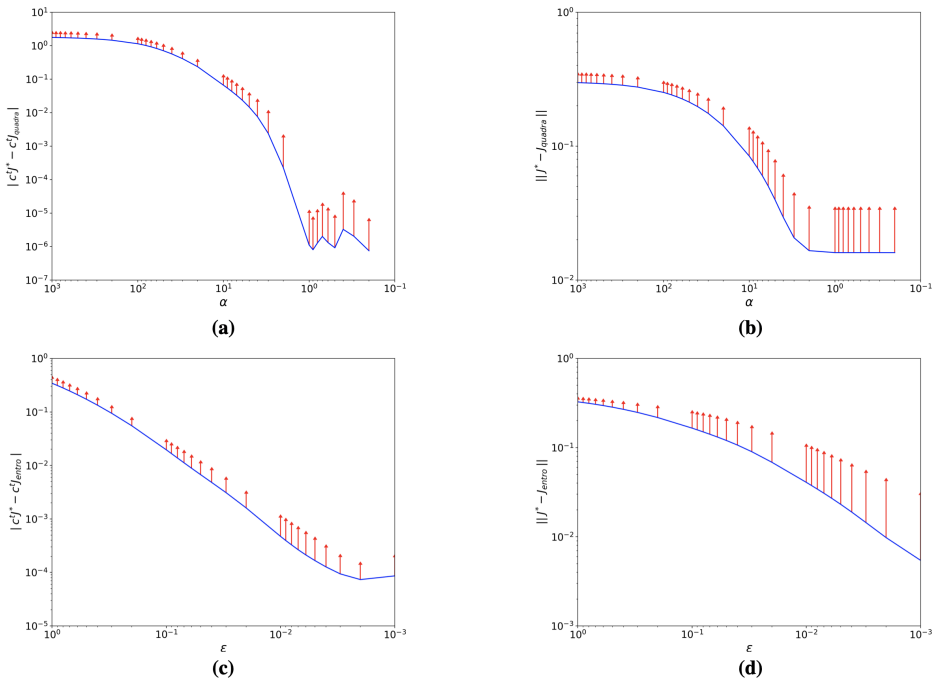


Figure 5: On the 4 figures, 100 experiments are conducted for each value of the parameter. For each experiment, a bipartite graph with 5 vertices in each component is created with a random cost attributed to edges and given source and target distribution. The entropic regularized problem on T is solved thanks to advanced Sinkhorn (using log-sum-exp trick) while the quadratic regularized problem on J is solved as classic QP thanks to CVXPY (interior point method). The plot of the mean of the metric over the 100 experiments is in blue while the standard deviations are indicated by red arrows.

evant as it converges to 0 in a reasonably uniform way (small standard deviations on graphs **(a)** and **(c)**). On the other hand, the distance to an unregularized optimal transport plan cannot be considered as an adequate metric, as highlighted on graphs **(b)** and **(d)** where the mean distance is floored and the standard deviations are significant. Indeed, in part of the experiments, regularized solutions converge towards an unregularized solution which is not the one selected.

2.3.3 Case study

NB : This case study is inspired from a figure concluding a recently published article [7].

In this part, entropic regularization and quadratic regularization are compared on a problem of optimal transport between two clouds of points. This configuration can be assimilated to a bipartite graph with a cost per edge corresponding to the euclidian distance between the points. The two clouds of points are generated as follows :

- **source** : 100 points randomly selected in the square of side $s = 1$ centered in $C = (0, 0)$

- **target** : 150 points randomly selected in the ring of center $C = (0, 0)$, radius $R = 2$ and width $w = 0.4$

Finally, the source and target distributions are defined as uniform.

Then, we select an entropic parameter $\epsilon = 0.001$ and a quadratic parameter $\alpha = 0.5$ which are comparable according to metric d introduced in the previous part. Here, the measured distances to unregularized transport are $d(\epsilon) = 5 \times 10^{-5}$ and $d(\alpha) = 1 \times 10^{-5}$.

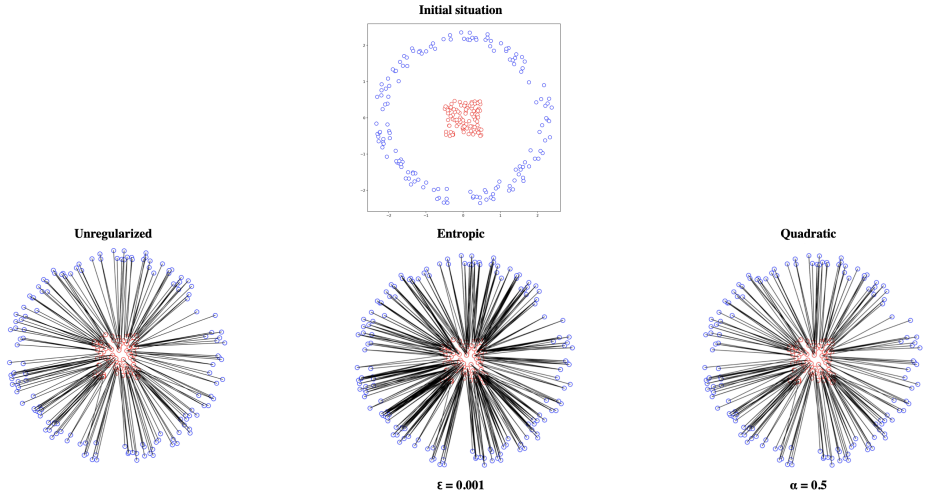


Figure 6: Each black line represents a strictly positive value in the transportation plan represented. The entropic regularized problem on T is solved thanks to advanced Sinkhorn (using log-sum-exp trick) while the quadratic regularized problem on J is solved as classic QP thanks to CVXPY (interior point method). Transportation plan values lower than 10^{-6} are assimilated to 0 in these figures.

In Figure 6, entropic solution generates more connections while quadratic and unregularized appears to be more sparse. This visible difference is supported by the transportation plans and their histograms presented in Figure 7. In the histograms of the quadratic and unregularized transportation plan, 3 main peaks are distinguishable whereas the entropic histogram illustrates a more homogeneous flow between the distributions.

Thanks to this case study, the main advantage of quadratic regularization is highlighted. Quadratic regularizer permits to obtain sparse solutions which can be almost as sparse as the unregularized ones. On the contrary, entropic solution by enforcing the strict positivity the transportation plan tends to reach more homogeneous solutions.

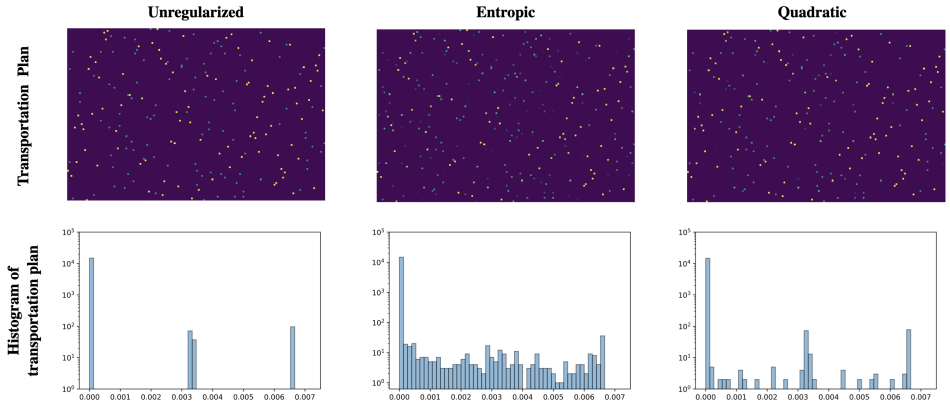


Figure 7: On the first line, the transportation plans are presented in a raw manner thanks to a colormap from yellow for the maximum of the matrix to dark purple for 0. On the second line, histograms of the values of the matrix are plotted. Transportation plan values lower than 10^{-6} are assimilated to 0 in these figures.

2.4 Original solving algorithm

2.4.1 Principles

In order to solve quadratically regularized transport on a graph (4), an original algorithm using the dual problem (11) is proposed. The main concept of this algorithm is to combine optimization methods with classical graph knowledge for enhanced calculation management.

First, a diagonal indicator of activated edges $M(\mu)$ (edges used by the flow J) is introduced and defined as follows.

$$M(\mu)_{ee} = \begin{cases} 1 & \text{if } (D\mu - c)_e > 0 \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

Then, the dual obtained in the first section is reformulated by introducing $M(\mu)$.

$$\max_{\mu} \quad \alpha \mu^t f - \frac{1}{2}(D\mu - c)^t \, M(\mu) \, (D\mu - c) \tag{18}$$

Then, the method follows a common gradient descent format : direction-search, line-search for step size and update of variable until gradient vanishes.

The main difficulty lies in the fact that the diagonal indicator M is considered as constant at each optimization step. This hypothesis leads to closed forms for the gradient and hessian of objective function in (4), noted g.

$$\begin{aligned} \nabla g(\mu) &= \alpha f - D^t M (D\mu - c) \\ \nabla^2 g(\mu) &= -D^t M D = L \end{aligned} \tag{19}$$

In more details, here are the algorithms steps :

- **Direction-search** : The method alternates between gradient and pseudo-Newton method (with Moore Penrose pseudoinverse of L) to obtain direction s.
- **Line-search** : Considering M constant, finding optimal step t is a parabola optimization which gives a closed form, noted t_{quadra}. However, to ensure M constant, the step size needs to be capped by the lowest t which modifies M, noted $t_{active\ set}$.
  Therefore, we define $t = \min(t_{quadra}, t_{active\ set})$.
- **Variable update** : $\mu = \mu + t\, s$.
- **Laplacian update** : the Laplacian L is turned into a full rank matrix by adding NN^t where N is an orthogonal basis of Ker(L).
  M being idempotent, we have : $L + NN^t = \begin{pmatrix} MD \\ N^t \end{pmatrix}^t \begin{pmatrix} MD \\ N^t \end{pmatrix} = W^t W$
  By using QR decomposition of W, we get : $L + NN^t = R^t R$.

Therefore, we simply need to update R thanks to Cholesky updates depending on the impact of added or deleted active edges. For example, for a supplementary active edge e, $D_e D_e^t$ is added to L (D_e row of D corresponding to e). If this new active edge merges two connected components then N also needs to be modified accordingly.

2.4.2 Implementation

The algorithm was implemented in Matlab in the article. The choice was made to implement it in Python to test its robustness.

Three main difficulties were encountered in the Python implementation :

- **Cholesky updates** : There is no official Python library implementing robust Cholesky updates. Therefore, a few stabilization problems were encountered.
- **Graph connected components** : N needs to be modified if there is a change in the connected components. Nevertheless, this step was not explained in the article although it is not trivial, especially in the case of deletion of an edge. Here, Breadth-First Search algorithm is used to define connected components of the graph after each deletion and its time complexity is substantial : $O(n_v + n_e)$.
- $t_{active\ set}$ **convergence towards** 0 : for some configurations, especially non bipartite ones, $t_{active\ set}$ converged to 0 before the algorithm converges preventing the solving algorithm to update μ. In the article, only one sentence concerns this problem : "We never find a case where $t_{active\ set} = 0$ before convergence."

2.4.3 Performance comparisons

To measure performances, two types of graphs with varying number of vertices were used :

- **Bipartite graphs** : Vertices are randomly split into two components and a random distribution is attributed to each component. Each vertex of component A is connected to all vertices in component B. A random floating number between 1 and 10 is attributed to each edge as its cost.
- **Fully connected graphs** : Every vertex is connected to all other vertices. Identically, a random floating number between 1 and 10 is attributed to each edge as its cost.

First, due to the problem of $t_{active\ set}$ convergence towards 0, the failure rate was evaluated for both bipartite and fully connected graphs (see Table 2). A failure corresponds to a number of iterations exceeding 3000. The failure rate seems to be lower in the case of bipartite graphs and it might be because of a lower number of edges to handle. Besides, over 10 vertices, the failure is almost of 100% for every α. Furthermore, for all failures, the algorithm is stuck at a given point because of $t_{active\ set}$ being extremely small (usually below 10^{-10}).

	4	**6**	**8**	**10**
0.05	45 / 30	88 / 62	99 / 90	100 / 98
0.1	41 / 32	85 / 60	95 / 85	100 / 96
0.5	41 / 21	91 / 54	99 / 80	100 / 92
1	42 / 26	79 / 60	93 / 81	98 / 88
5	36 / 21	69 / 40	83 / 68	95 / 73
10	43 / 22	60 / 43	87 / 62	96 / 69
50	36 / 19	66 / 33	83 / 52	92 / 65

Table 1: Table of failure rates of the algorithm due to $t_{active\ set}$ convergence towards 0. For each combination of parameters, 100 experiments were conducted for bipartite and fully connected graphs to observe the failure rate in each situation. Each column corresponds to a n_v and each line to a value of α. Results are in $\%$ and are presented as follows : $\text{failure}_{\text{fully connected}}$ / $\text{failure}_{\text{bipartite}}$.

In case of absence of failure, a comparison with the QP solver from CVXPY [2] applied to the primal problem was conducted in terms of runtime. In order to have enough viable experiments, only low number of vertices for biparite graphs are considered (see Table 2). In the very limited cases observed, the algorithm seems to be competitive with the QP solver from CVXPY. Furthermore, for $\alpha > 5$, performances are even better than the one from the CVXPY library.

	4	**6**
0.05	10.6 / 6.5	91.6 / 6.4
0.1	9.8 / 6.5	57.6 / 6.7
0.5	4.8 / 6.5	17.3 / 6.6
1	4.3 / 6.4	11.4 / 6.5
5	3.3 / 6.4	6.3 / 6.7
10	2.2 / 6.4	4.9 / 6.5
50	1.8 / 6.4	3.5 / 6.5

Table 2: Table of median runtimes for CVXPY and the developed solving algorithm. For each combination of parameters, 100 experiments were conducted for bipartite. The runtime is kept only if both algorithms don't fail. Each column corresponds to a n_v and each line to a value of α. Results are in ms and are presented as follows : $\text{runtime}_{\text{new}}$ / $\text{runtime}_{\text{CVXPY}}$.

3 Conclusion and perspectives

First and foremost, the article discussed in this report brings major theoretical guarantees on the link between quadratically-regularized and unregularized solutions applicable to all discrete transport frameworks. Indeed, the main result (Proposition 1) permits to understand the behaviour of quadratically regularized optimal transport in low regularization regime.

Then, our comparison between quadratic and entropic regularizers in the case of graphs highlighted how quadratic regularization can enable us to reach very sparse solutions, close to unregularized ones. Also more adapted to the framework of graphs, quadratic regularizers appear to be an important tool despite being often systematically replaced by entropic regularizer.

Nevertheless, as underlined by our study of the proposed solving algorithm, the main obstacle to a wider use of quadratic regularization is the absence of simple and efficient solving algorithms. The main advantage of entropic regularization comes from its extremely simple solving Sinkhorn algorithm. Here, the algorithm proposed seems to be complicated to adapt to different languages while keeping reasonable robustness and performances.

To go further, a quantification of the distance (cost difference) between regularized and unregularized solutions could be explored. In a similar vein, α_∞ could be more explicitly defined based on G and f.

References

[1] George B. Dantzig. "Origins of the simplex method". In: *Technical Report SOL* 87-5 (1987). URL: https://apps.dtic.mil/dtic/tr/fulltext/u2/a182708.pdf.

[2] Steven Diamond and Stephen Boyd. *CVXPY: A Python-Embedded Modeling Language for Convex Optimization*. 2016. arXiv: 1603.00943 [math.OC].

[3] E. Dijkstra. "A note on two problems in connexion with graphs". In: *Numerische Mathematik* 1 (1959), pp. 269–271.

[4] Montacer Essid and Justin Solomon. *Quadratically-Regularized Optimal Transport on Graphs*. 2018. arXiv: 1704.08200 [math.OC].

[5] Trevor Hastie and Robert Tibshirani. "Efficient quadratic regularization for expression arrays". In: *Biostatistics (Oxford, England)* 5 (Aug. 2004), pp. 329–40. DOI: 10.1093/biostatistics/5.3.329.

[6] Wuchen Li et al. "A Parallel Method for Earth Mover's Distance". In: *Journal of Scientific Computing* 75 (Apr. 2018). DOI: 10.1007/s10915-017-0529-1.

[7] Dirk A. Lorenz, Paul Manns, and Christian Meyer. *Quadratically regularized optimal transport*. 2019. arXiv: 1903.01112 [math.OC].

[8] Richard Sinkhorn and Paul Knopp. "Concerning nonnegative matrices and doubly stochastic matrices." In: *Pacific J. Math.* 21.2 (1967), pp. 343–348. URL: https://projecteuclid.org:443/euclid.pjm/1102992505.

[9] Daniel A. Spielman and Shang-Hua Teng. "Smoothed Analysis of Algorithms: Why the Simplex Algorithm Usually Takes Polynomial Time". In: *CoRR* cs.DS/0111050 (2001). URL: https://arxiv.org/abs/cs/0111050.

[10] Wotao Yin. "Analysis and Generalizations of the Linearized Bregman Method". In: *SIAM J. Imaging Sciences* 3 (Jan. 2010), pp. 856–877. DOI: 10.1137/090760350.