Computation of Discrete Flows Over Networks via Constrained Wasserstein Barycenters

Ferran Arqué [*1] **César A. Uribe** [*2] **Carlos Ocampo-Martinez** [*1]

Abstract

We study a Wasserstein attraction approach for solving dynamic mass transport problems over networks. In the transport problem over networks, we start with a distribution over the set of nodes that needs to be "transported" to a target distribution accounting for the network topology. We exploit the specific structure of the problem, characterized by the computation of implicit gradient steps, and formulate an approach based on discretized flows. As a result, our proposed algorithm relies on the iterative computation of constrained Wasserstein barycenters. We show how the proposed method finds approximate solutions to the network transport problem, taking into account the topology of the network, the capacity of the communication channels, and the capacity of the individual nodes.

1. Introduction

Optimal transport (OT) theory has experienced increased interest over the last few years, due to its wide range of applications in both theoretical and applied fields of mathematics (Villani, 2008). In particular, the recent efforts to overcome the high computational cost of the associated linear programming problem (Cuturi, 2013), has made OT an attractive choice to tackle problems involving a large number of distributions or other high dimensional objects, and requiring a high desired accuracy.

Our work focuses on the discrete OT problem, where probability distributions are defined over the nodes of a finite graph. In classical OT approaches, it is assumed that mass (or a fraction of it) at each point in the support of one of the probability measures can be sent to any of the elements in the support of the other probability measure. As a result, the transport plan is executed effectively in one step. Contrary to existing literature, we seek to explicitly consider the topology of the underlying graph, which naturally imposes some transportation constraints. Adding the topology of the graph as a constraint means that there may not be a direct link between two points in the support, as the edges of the graph directly determine links. Therefore, our goal is to find a sequence of transport plans that move the mass from an initial distribution to a final one along the edges of a connected graph so that the cost of transportation is minimal while accounting for channel and node capacities.

Finding the amount of mass that needs to be sent through each edge, so that the total cost of transportation is minimal, is a well-known problem called the *minimum-cost flow problem*. This problem has been widely studied and different algorithms have been proposed to solve it (Ahuja et al., 1993). More importantly, the Wasserstein distance can be rewritten as a minimum-cost flow problem. However, classical methods to solve this problem do not have a system to discern between paths when the optimal flow is not unique. This nonuniqueness leads to unpredictability of the output from the solver since many paths can be indistinguishable in terms of costs. To avoid that case, some algorithms introduce an additional term to the objective function so that it becomes strongly convex. These regularized OT methods, like the well-known Sinkhorn algorithm (Cuturi, 2013), achieve uniqueness and significantly speed up the computation, compared to solving a large linear programming problem, but it is at the cost of finding an approximation of the solution to the original problem.

Our approach is based on the solution of the Wasserstein attraction problem, which requires the computation of a *Wasserstein barycenter* (WB) of two distributions at every iteration. Computing the WB yields an intermediate distribution, defined as the Fréchet mean of the two measures, which is the result of minimizing the sum of the distances between itself and each of the two distributions (Cuturi & Doucet, 2014). However, the support of this resulting distribution can include any of the graph nodes. We expand the definition of the WB problem by adding constraints that ensure the mean obtained has the appropriate support and each node does not receive more mass than the amount available

[*]Equal contribution [1]Automatic Control Department, Universitat Politècnica de Catalunya, Barcelona, Spain. [2]Department of Electrical and Computer Engineering, Rice University, Houston TX. Correspondence to: Ferran Arqué <ferran.arque@upc.edu>.

Proceedings of the LatinX in AI (LXAI) Research workshop at ICML 2021.

from its neighbors. The closest works to our setting with constrained WB are (Peyré, 2015; Cuturi & Peyré, 2016). The former presents a framework to approximate gradient flows for Wasserstein metrics by computing discrete entropy-regularized flows, which are computed as *JKO flows*. In particular, it introduces the concept of *Wasserstein attraction*, which is used in our work. The latter complements this previous work while focusing on the dual formulation of Wasserstein variational problems.

The main contributions of this paper are twofold: first, we propose the mathematical formulation of a Wasserstein attraction-like problem to solve mass transport problems over networks by writing them as the computation of a WB problem with additional constraints. And second, we present a methodology to find an approximation of optimal discrete flows over networks based on Dykstra's projection algorithm and the computation of proximal operators for the Kullback-Leibler divergence.

Notation: The column vector of all ones is denoted by $\mathbf{1}$ and I is the identity matrix. The adjacency matrix of a graph is denoted by A, and we will write $\bar{A} = A + I$ when considering the connection of one node to itself. $\mathbb{R}_+$ and $\mathbb{R}_{++}$ refer to non-negative and strictly positive real values respectively. Given two matrices $A, B \in \mathbb{R}^{n\times m}$, $\langle A, B\rangle = \sum_{i,j} A_{ij}B_{ij}$. We define the support of a function (or vector) ρ as $\text{SUPP}(\rho) = \{i \mid \rho(i) > 0\}$. We denote $\text{KL}(\pi|\xi)$ as the *Kullback-Leibler divergence* between $\pi \in \mathbb{R}^{n\times n}_+$ and $\xi \in \mathbb{R}^{n\times n}_{++}$, defined as $\text{KL}(\pi|\xi) = \sum_{i,j=1}^n \pi_{ij}\ln(\pi_{ij}/\xi_{ij}) - \pi_{ij} + \xi_{ij}$, with the convention $0\ln(0) = 0$. Finally, the indicator function of a set $\mathcal{C}$ is defined as $\iota_{\mathcal{C}}(x) = 0$ if $x \in \mathcal{C}$, and $\iota_{\mathcal{C}}(x) = +\infty$ otherwise.

2. Problem Statement: Discrete Flows and Wasserstein Attraction on Graphs

2.1. Discrete Flows on Graphs

Consider a *discrete*, *finite*, *fixed* and *connected* graph $\mathcal{G} = (V, E)$, where V is a set of n nodes $V = (1, \cdots, n)$, and E is a set of directed edges such that $E \subseteq V \times V$, where $(j, i) \in E$ if and only if there is an edge between the node $j \in V$ and node $i \in V$. Denote the probability simplex on V as $\text{Prob}(V) = \{\mu \in \mathbb{R}^n_+ \mid \sum_{x\in V} \mu(x) = 1\}$. The set of edges E has an associated weight function $c : E \to \mathbb{R}_+$ where each edge $e \in E$ has a corresponding weight $c_e = c(e)$, i.e., the cost of sending a unit of mass using the edge e. Furthermore, endow the graph $\mathcal{G}$ with its natural metric d which measures the total weight of the shortest path between any two nodes in $\mathcal{G}$.

We study the discrete flow (i.e., discretization in time) problem of optimally transporting an initial mass distribution $\mu \in \text{Prob}(V)$ to a target mass distribution $\nu \in \text{Prob}(V)$ using the graph $\mathcal{G}$. The associated weight of each edge allows us to define a cost matrix $C \in \mathbb{R}^{n\times n}_+$, where $[C]_{ji} = \mathsf{d}(j, i)$ indicates the cost of transporting a unit mass from node j to node i. Moreover, we endow the space $\text{Prob}(V)$ of probability measures on V with the entropy-regularized 1-Wasserstein distance between two probability distributions μ and ν on $\mathcal{G}$, defined as

$$\mathcal{W}_\gamma(\mu, \nu) = \min_{\pi\in\Pi(\mu,\nu)} \langle C, \pi\rangle + \gamma H(\pi), \tag{1}$$

where $H(\pi) = \sum \pi_{ij}(\ln \pi_{ij} - 1) = \langle \pi, \ln \pi - \mathbf{1}\mathbf{1}^\top\rangle$ is the the negative entropy and $\gamma > 0$ is the regularization parameter. Here the minimizer (defined as the *regularized optimal transport plan*) is computed over all couplings on $V \times V$ with marginals μ and ν, i.e., $\Pi(\mu, \nu) = \left\{\pi \in \mathbb{R}^{n\times n}_+ \;\middle|\; \pi\mathbf{1} = \mu,\ \pi^\intercal\mathbf{1} = \nu\right\}$.

Our objective is to design a discrete flow $\{\rho_t\}_{t\geq 0}$ on $\mathcal{G}$, where $\rho_t \in \text{Prob}(V)$, by constructing a sequence of transport plans $\{\pi_t\}_{t\geq 0}$ such that $\rho_0{=}\mu$, $\rho_{t+1}{=}\pi_t\mathbf{1}$, $\rho_t{=}\pi_t^\intercal\mathbf{1}$, $\lim_{t\to\infty}\rho_t{=}\nu$ and the transport cost at each iteration is minimized. Moreover, the desired sequence of transport plans is required to satisfy the constraints imposed by the network, namely that a node can only send mass to its neighbors (in other words, the flow should follow the sparsity pattern induced by the graph topology), the mass sent over an edge cannot be greater than the associated edge capacity and the mass at a node must not exceed its local storage capacity.

2.2. Wasserstein Attraction Flows on Graphs

We formulate the dynamic transport problem described in Section 2.1 as a constrained *Wasserstein attraction* (WA) problem (Peyré, 2015). Starting from an initial distribution $\rho_0 = \mu$, the approximate entropy-regularized WA flow is defined as

$$\begin{aligned}\rho_{t+1} &= \text{Prox}^{\mathcal{W}_\gamma}_{\tau,\mathcal{W}_\gamma(\cdot,\nu)}(\rho_t)\\ &= \operatorname*{argmin}_{p\in\text{Prob}(V)} \left\{\mathcal{W}_\gamma(p, \rho_t) + \tau\mathcal{W}_\gamma(p, \nu)\right\},\end{aligned} \tag{2}$$

where τ is a step size. Note $\mathcal{W}_\gamma(\cdot,\cdot)$ is a strictly convex and coercive function, therefore the operator in (2) is uniquely defined.

Next, we state one important observation about the entropy-regularized WA flow in (2). Without loss of generality, one can multiply the argument in the optimization problem (2) by a constant $\omega = 1/(1 + \tau)$. Thus, we obtain

$$\rho_{t+1} = \operatorname*{argmin}_{p\in\text{Prob}(V)} \left\{\omega\mathcal{W}_\gamma(p, \rho_t) + (1 - \omega)\mathcal{W}_\gamma(p, \nu)\right\}, \tag{3}$$

which is precisely the entropy-regularized WB between ρ_t and ν (Cuturi & Doucet, 2014). Hence, we interpret the WA as the sequential computation of Wasserstein barycenters, which can be more intuitive and furthermore offers the possibility to use theoretical results found in the literature

for both the WA and WB problems. This introduces an additional weight parameter that can be modified to give preference to one measure or the other, which consequently alters how the mass is transported across the graph. With this, we state in Problem 1 our main contribution regarding the design of the entropy-regularized discrete WA flow.

Problem 1 *Let $A \in \{0,1\}^{n\times n}$ be the adjacency matrix of a discrete, finite, fixed and connected graph with n vertices, $C \in \mathbb{R}_+^{n\times n}$ the cost matrix, and ρ_0, $\nu \in Prob(V)$ the initial and final distributions respectively. Design a sequence of probability measures $\{\rho_t\}_{t\geq 0}$ by finding, for each $t \geq 0$, the transport plan that solves the optimization problem*

$$\{\pi_t\} = \underset{\substack{\boldsymbol{\pi}\in\mathcal{C}_f\cap\mathcal{C}_e \\ \boldsymbol{\pi}\in\mathcal{C}_1\cap\mathcal{C}_2\cap\mathcal{C}_3}}{\operatorname{argmin}} \ \omega KL(\pi_1|\xi)+(1-\omega)KL(\pi_2|\xi), \tag{4a}$$

where

$$\mathcal{C}_f = \left\{\pi_1,\pi_2 \in \mathbb{R}_+^{n\times n} \mid \pi_1^\intercal \mathbf{1} = \rho_t, \pi_2^\intercal \mathbf{1} = \nu\right\} \tag{4b}$$

$$\mathcal{C}_e = \left\{\pi_1,\pi_2 \in \mathbb{R}_+^{n\times n} \mid \pi_1 \mathbf{1} = \pi_2 \mathbf{1} = p\right\} \tag{4c}$$

$$\mathcal{C}_1 = \left\{\pi_1,\pi_2 \in \mathbb{R}_+^{n\times n} \mid \pi_1 \leq \tilde{C}\right\} \tag{4d}$$

$$\mathcal{C}_2 = \left\{\pi_1,\pi_2 \in \mathbb{R}_+^{n\times n} \mid \pi_1 \mathbf{1} \leq \rho\right\} \tag{4e}$$

$$\mathcal{C}_3 = \left\{\pi_1,\pi_2 \in \mathbb{R}_+^{n\times n} \mid [\pi_1\mathbf{1}]_i \leq \sum_{j:(j,i)\in E} [\rho_t]_j\right\} \tag{4f}$$

3. Iterative Projections for the Computation of Transport Plans

Dykstra's projection algorithm can be used to solve problems of the form $\min_{\pi\in\cap_i\mathcal{C}_i} \mathrm{KL}(\pi|\xi)$, much like Problem 1 defined in Section 2. It is based on the computation of the proximal operators for the KL divergence. This is done iteratively, cycling through each constraint set $\mathcal{C}_i$, and since $\mathcal{C} = \cap_i \mathcal{C}_i$ is a finite intersection of L sets, we shall define, for every index i, $\mathcal{C}_{i+L} = \mathcal{C}_i$. Then, for each $k > 0$ we compute

$$\pi^{(k)} = \mathrm{Prox}^{\mathrm{KL}}_{\iota_{\mathcal{C}_k}}\left(\pi^{(k-1)}\cdot q^{(k-L)}\right), \ q^{(k)} = q^{(k-L)}\frac{\pi^{(k-1)}}{\pi^{(k)}},$$

with initial values $\pi^{(0)} = \xi$ and $q^{(0)} = q^{(-1)} = \ldots = q^{(-L+1)} = \mathbf{1}\mathbf{1}^\intercal$. The product and division of matrices are considered element-wise. We slightly abuse notation by omitting the step-size τ in the definition of the proximal operator, since we are multiplying the argument in the optimization problem (2) by $\omega = 1/(1+\tau)$, as noted in Section 2.

The next propositions state how we can compute in closed form the proximal operators corresponding to each constraint in Problem 1.

Proposition 1 (Proposition 1 in (Benamou et al., 2015)) *The proximal operator of the indicator function $\iota_{\mathcal{C}_1}$ has the closed form*

$$\left[\mathrm{Prox}^{\mathrm{KL}_\omega}_{\iota_{\mathcal{C}_f}}(\boldsymbol{\pi})\right]_l = \mathrm{Prox}^{\mathrm{KL}}_{\iota_{\{\pi_l\mathbf{1}=P_l\}}}(\pi_l) = diag\left(\frac{P_l}{\pi_l\mathbf{1}}\right)\pi_l. \tag{5}$$

Proposition 2 (Proposition 2 in (Benamou et al., 2015)) *The proximal operator of the indicator function $\iota_{\mathcal{C}_e}$ has the closed form*

$$\left[\mathrm{Prox}^{\mathrm{KL}_\omega}_{\iota_{\mathcal{C}_e}}(\boldsymbol{\pi})\right]_l = \pi_l diag\left(\frac{p}{\mathbf{1}^\intercal\pi_l}\right), \tag{6}$$

where $p = \prod_{l=1}^m (\mathbf{1}^\intercal\pi_l)^{\omega_l}$ (the products and exponentiation are considered element-wise), and $m = 2$ in our case.

Proposition 3 (Section 5.2 in (Benamou et al., 2015)) *The proximal map for the function $\iota_{\mathcal{C}_1}$ is defined as*

$$\left[\mathrm{Prox}^{\mathrm{KL}_\omega}_{\iota_{\mathcal{C}_1}}(\boldsymbol{\pi})\right]_l = \mathrm{Prox}^{\mathrm{KL}}_{\iota_{\{\pi_l\leq\tilde{C}\}}}(\pi_l) = \min\left(\pi_l, \tilde{C}\right), \tag{7}$$

with the minimum computed element-wise.

Proposition 4 (Proposition 5 in (Benamou et al., 2015)) *For the the indicator function $\iota_{\mathcal{C}_2}$ one has*

$$\begin{aligned}\left[\mathrm{Prox}^{\mathrm{KL}_\omega}_{\iota_{\mathcal{C}_2}}(\boldsymbol{\pi})\right]_l &= \mathrm{Prox}^{\mathrm{KL}}_{\iota_{\{\pi_l^\intercal\mathbf{1}\leq\rho\}}}(\pi_l) \\ &= \pi_l diag\left(\min\left(\frac{\rho}{\pi_l^\intercal\mathbf{1}}, \mathbf{1}\right)\right),\end{aligned} \tag{8}$$

with minimum and division of vectors done element-wise.

Nodes in $\mathrm{SUPP}(\rho_t)$ can send mass to non-neighboring nodes. To fix this issue, we adapt constraint (4d). We redefine the capacity matrix $\tilde{C}$ for the transport plan π_1 from ρ_t to ρ_{t+1}, such that for the nodes in the support of ρ_t, if there is no connection between one of them and another node, the "link" between them has zero capacity, i.e.,

$$[\tilde{C}]_{ij} = \begin{cases} 0 & \text{if } j \in \mathrm{SUPP}(\rho_t) \text{ and } \bar{A}_{ij} = 0, \\ [\tilde{C}]_{ij} & \text{otherwise.}\end{cases} \tag{9}$$

Algorithm 1 summarizes the proposed method. It is important to remark that our entropy-regularized approach does not allow the scheme to converge exactly to the target distribution ν, since the additional entropy term forces every node to send a small amount of mass to the rest. Moreover, we cannot guarantee the convergence of Algorithm 1 for a fixed weight ω, and to our knowledge, there is no proof for it as of yet. However, if instead of taking fixed values for both γ and ω we consider, at each step t, $\gamma(t)$, $\omega(t)$ such that $\gamma(t)$, $\omega(t) \to 0$ as $t \to +\infty$, we can ensure its convergence (Benamou et al., 2015; Peyré, 2015). Despite that, in the

Algorithm 1 Conceptual procedure of the approach

Input: Initial and final distribution vectors ρ_0 and ν, adjacency matrix A, cost matrix C, storage capacity vector ρ, regularization parameter $\gamma(t)$ and weight $\omega(t)$ such that $\gamma(t), \omega(t) \to 0$ as $t \to +\infty$, accuracy parameter ε

$t = 0$
while $\frac{1}{2} \|\nu - \rho_t\|_1 > \varepsilon$ **do**
  Define the capacity matrix $\tilde{C}$ as seen in (9)
  Compute the WB ρ_{t+1} with weights $\omega_1 = \omega(t)$ and $\omega_2 = 1 - \omega(t)$ and the additional support and capacity constraints by using Dykstra's projection algorithm with initial conditions $\pi_1^{(0)} = \pi_2^{(0)} = e^{-\frac{C}{\gamma(t)}}$ and the proximal operators defined on (5), (6) and (8) for both transport plans, and (7) only for transport plan π_1 to enforce the capacity constraint (4d) with capacity matrix $\tilde{C}$
  $t \leftarrow t + 1$
end while

Output: $\{\rho_t\}_t$

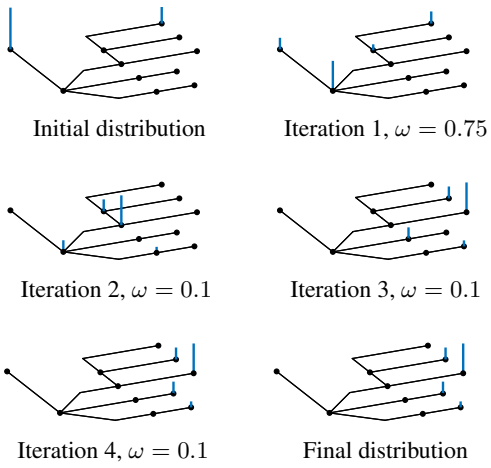

Figure 1. Steps obtained with Algorithm 1 on a small network. In the first step, $\omega = 0.75$ is used in favor of the initial distribution so that the transportation is done more gradually. In the following steps, the weight is reduced to $\omega = 0.1$, so that the target is covered much faster.

simulations carried out, we considered the weight ω to be both fixed and tending to zero and we observed how, for a constant $\omega < 1/2$, the mass reaches the target distribution as well. Regarding the convergence of the computation of each intermediate distribution in the discrete flow, we state the following lemma.

Lemma 1 *For each step t, let $\tilde{C}$ be the capacity matrix defined in (9) such that it verifies $\tilde{C}^\intercal \mathbf{1} > \rho_t$, and let ρ be the retention capacity vector in the constraint set $\mathcal{C}_2$ such that $\rho_t < \rho$ (both inequalities are considered element-wise). Then, the iterative computation of the proximal steps defined in Propositions 1, 2, 3 and 4 converges to the solution of (4a).*

Finally, Figure 1 shows a simple case study of a small network to illustrate the effectiveness of the proposed approach.

4. Conclusions and Future Work

In this paper, we have presented a mathematical formulation to resolve discrete optimal flows over networks, based on the computation of constrained Wasserstein Barycenters. Using the entropically regularized approximation of the Wasserstein metric allows us to make use of Dykstra's projection algorithm, which in itself is easy to implement and is competitive in terms of performance speed since it only requires elementary operations such as matrix and vector products. Moreover, with this methodology, the solution obtained is unique. The scheme presented can be extended to consider more than two distributions (by the definition of the Wasserstein barycenter) and is able to adapt to different changes, thus, a new line of investigation could be to use the proposed approach to tackle problems involving decentralized or distributed models, where not all the information is available for every agent.

References

Ahuja, R., Magnanti, T., and Orlin, J. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.

Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., and Peyré, G. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pp. 2292–2300, 2013.

Cuturi, M. and Doucet, A. Fast computation of Wasserstein barycenters. *Journal of Machine Learning Research*, 2014.

Cuturi, M. and Peyré, G. A smoothed dual approach for variational Wasserstein problems. *SIAM J. Imaging Sci.*, 9:320–343, 2016.

Peyré, G. Entropic approximation of Wasserstein gradient flows. *SIAM Journal on Imaging Sciences*, 8(4):2323–2351, 2015.

Villani, C. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.