

Approximate Wasserstein Attraction Flows for Dynamic Mass Transport over Networks[★]

Ferran Arqué^a, César A. Uribe^b, Carlos Ocampo-Martinez^a

^a Automatic Control Department, Universitat Politècnica de Catalunya, Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Llorens i Artigas, 4-6, 08028, Barcelona, Spain

^b Department of Electrical and Computer Engineering, Rice University, Houston TX 77006

Abstract

This paper presents 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. Finally, we show the performance of this approach applied to large-scale water transportation networks.

Key words: Optimal transport, Wasserstein distance, constrained Wasserstein barycenter, discrete flow, network

1 Introduction

The formalization of the optimal transportation problem by Gaspard Monge in 1781 (Monge 1781) and Leonid Kantorovich (Kantorovitch 1958) evolved into a whole branch of mathematics (Ambrosio et al. 2008, Ambrosio & Gigli 2013, Villani 2008) called *optimal transport theory*. Optimal transport (OT) is based on the computation of the distance between two objects, probability distributions in many cases, commonly referred to as the *Wasserstein distance* (Vaserstein 1969). Optimal transport has a wide range of applications, e.g., image retrieval (Rubner et al. 2000), averaging atmospheric gas concentration data sets (Barré et al. 2020), segmentation and labeling of neurons (Nejatbakhsh et al. 2020), clustering patterns linking COVID-19 dynamics and human mobility (Nielsen et al. 2020), and different machine learning problems, such as generative adversarial networks (Arjovsky et al. 2017), low-rank matrix factor-

ization (Cuturi et al. 2020) or fair regression (Chzhen et al. 2020).

Despite the appealing properties of the Wasserstein distance, its computation requires solving an optimization problem. Such optimization problems become computationally prohibited on high dimensional objects, a large number of distributions or a high desired accuracy. However, the seminal work of Brenier (Brenier 1991) led to practical numerical algorithms that started the search for efficient algorithms to solve the OT problem (Peyré & Cuturi 2019).

Our work focuses on the discrete OT problem, where probability distributions are defined over the nodes of a graph, assumed to be finite. In traditional 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. However, 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

[★] This paper was not presented at any IFAC meeting. Corresponding author: Ferran Arqué. Tel. +34934015752. Fax +34934015750.

Email addresses: ferran.arque@upc.edu (Ferran Arqué), cauribe@rice.edu (César A. Uribe), carlos.ocampo@upc.edu (Carlos Ocampo-Martinez).

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 general 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 (Morton 1967, Ahuja et al. 1993, Hu et al. 2020), and different algorithms have been proposed to solve it (Kovács 2015). More importantly, the Wasserstein distance can be rewritten as a minimum-cost flow problem when considering a complete bipartite graph (Bassetti et al. 2020). This can be extended to more general graphs if one considers the shortest path distance as the cost of sending a resource unit from one node to the other. In this case, if we compute the minimum-cost flow, then the mass sent from one node to the other is known. Moreover, since the shortest path between them is also known, we can make the first step of the transportation by sending that mass through the first edge of the path. Following this approach, we obtain the desired sequence of transport plans. However, classical methods to solve this problem do not have a condition to discern between paths when the optimal flow is not unique (Essid & Solomon 2018). 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. Still, it is at the cost of finding an approximation of the solution to the original problem.

Our approach is based on the resolution of the *Wasserstein attraction* (WA) problem (Peyré 2015), 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 (Wasserstein) 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 from its neighbors. This approach resembles what is called *displacement interpolation* (Villani 2008, Solomon et al. 2015). However, displacement interpolation in the discrete-time case may require a small step size of the weight to ensure certain smoothness in the transportation (i.e., to avoid some of

the mass moving over more than one node in a single step), which may lead to having many more iterations than necessary. Furthermore, with this approach, there is the possibility that certain nodes receive more mass than the total obtainable from their neighboring nodes. In summary, the main differentiating factor between displacement interpolation and our proposal is the addition of the topology and capacity constraints imposed by a graph. In this regard, (Haasler et al. 2021) recently studied this problem in the context of traffic planning, where edge capacity constraints are taken into account, and proposed a framework based on the Lagrangian dual problem to solve it, which resembles the Sinkhorn algorithm.

Moreover, our proposed approach can be reformulated as a discrete gradient flow problem. Several papers work on discrete gradient flows over graphs (or other discrete domains) (Chow et al. 2017, Erbar et al. 2020, Mielke 2013, Richemond & Maginnis 2017). However, such papers focus on the theoretical analysis of differential equations rather than the computational aspect with the regularized approximation of the Wasserstein metric (except for (Erbar et al. 2020) which provides a more in-depth discussion on the topic), and no additional constraints are considered on the elements of the graph. 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* (named after the authors in (Richard et al. 1998)). It introduces the concept of Wasserstein attraction, which is used in our work. We expand on this concept by observing that our particular problem formulation allows us to write each iteration of the WA problem as the computation of a WB, which unlocks the use of powerful computational tools found in the literature to solve this problem. Additionally, as previously mentioned, we further generalize the definition of this regularized flow by including the supplemental constraints of the topology of a network and the node and edge capacity bounds, which are features not considered in (Peyré 2015). The latter work, (Cuturi & Peyré 2016), complements (Peyré 2015) while focusing on the dual formulation of Wasserstein variational problems. In the context of applications of JKO flows in OT, (Bunne et al. 2021) recently proposed a novel procedure for the computation of JKO flows based on input convex neural networks. It is applied in the study of population dynamics, where it assumes that the dynamics of the model is parameterized by an energy function, which controls how the transport is executed at each step, from one state to the next. In our application, this role is performed by another Wasserstein distance function instead of an energy one (in addition to further constraints), which also allows for explicit computation of the JKO steps. Similar works such as (Chen et al. 2018, Guex et al. 2019) propose methods for optimal transportation over networks based on Markov processes. The authors

in (Chen et al. 2018) use the relative entropy as an index of closeness between measures and, doing so, they solve *Schrödinger’s bridge problem* (Schrödinger 1931) for the computation of transport plans in a fixed number of steps. This same entropy is used to measure how much the mass spreads in the transportation. Thus, they design a transport plan where the mass spreads as much as possible to guarantee robustness against failures in the paths of the network, while still ensuring a reasonably low total cost. In a similar fashion, (Guex et al. 2019) uses a bag-of-paths framework equivalent to solving either a standard or a relaxed entropy-regularized Wasserstein distance problem. Our approach allows topology changes as well, but it does so by solving a new problem at each step of the transportation, ignoring previous events, which increases the computational costs. However, the problem formulation allows us to consider node and edge capacity constraints explicitly. Moreover, flow speed can be adjusted with the weight parameter introduced when solving a WB problem.

The main contributions of this paper are threefold. 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. 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 JKO flow proximal operators for the Kullback-Leibler divergence and prove the convergence of these intermediate steps under certain assumptions. Finally, to the best of our knowledge, there are no works related to water management systems under the Wasserstein distance framework. Hence, we illustrate how this approach using WB can be implemented to model a supply-and-demand problem in the context of drinking water networks, where the network constraints are a crucial aspect inherent in their nature. In addition, we show how it can automatically adapt to dynamic changes on the network’s topology and agents. Furthermore, since there is no known method that can be used for fair comparison that can generate a flow that minimizes the Wasserstein distances and takes into account the network constraints, we have opted to compare the performance of our method with the commercial solver CPLEX with an explicit formulation of the constraints.

The remainder of this article is structured as follows. In Section 2, we provide the necessary background for our work, stating some basic definitions from discrete OT theory and present the formal statement of the problem we want to solve. In Section 3, we briefly review Dykstra’s projection algorithm in the setting of optimization problems involving the Kullback-Leibler divergence and how it can be used to solve the WB problem. Then, we show the additional steps needed on the algorithm to enforce support constraints and capacity bounds on the network’s links and nodes. With that, we present

our proposed approach. In Section 4, we provide some illustrative examples. We discuss our approach in the context of flow optimization on drinking water networks and give some remarks regarding the numerical implementation of the proposed algorithm. Finally, in Section 5, we provide some final comments and discuss future investigation directions.

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 $x \in \mathbb{R}^n$, $\|x\|$ stands for its Euclidean norm. 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

$$\text{KL}(\pi|\xi) = \sum_{i,j=1}^n \pi_{ij} \ln \left(\frac{\pi_{ij}}{\xi_{ij}} \right) - \pi_{ij} + \xi_{ij},$$

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 $V = \{1, \dots, n\}$, 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 $c : E \rightarrow \mathbb{R}_+$ 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 $[C]_{ji} = d(j, i)$ 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 1-Wasserstein distance between two probability distributions μ and ν on \mathcal{G} as

$$W_1(\mu, \nu) = \min_{\pi \in \Pi(\mu, \nu)} \sum_{x, y \in V} d(x, y) \pi(x, y),$$

where the minimizer is computed over all couplings on $V \times V$ with marginals μ and ν , i.e., the set of *optimal transport plans* $\Pi(\mu, \nu) = \{\pi \in \mathbb{R}_+^{n \times n} \mid \pi \mathbf{1} = \mu, \pi^\top \mathbf{1} = \nu\}$.

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 $\rho_0 = \mu$, $\rho_{t+1} = \pi_t \mathbf{1}$, $\rho_t = \pi_t^\top \mathbf{1}$ and $\lim_{t \rightarrow \infty} \rho_t = \nu$. Moreover, the transport cost at each iteration should be minimized.

Furthermore, the desired sequence of transport plans is required to satisfy the following constraints imposed by the network:

- (a) A node can only send mass to its neighbors, i.e.,

$$[\pi_t]_{ij} > 0 \text{ if } [\rho_t]_j > 0 \text{ and } (j, i) \in E. \quad (1)$$

In other words, the flow should follow the sparsity pattern induced by the graph topology. Intuitively, a flow can only be assigned between two nodes if and only if there is an edge connecting them. Hence, for a transport plan π_t it must hold that $\text{SUPP}(\rho_{t+1}) \subseteq \{\text{SUPP}(\rho_t) \cup \{j \mid (j, i) \in E\}\}$.

- (b) The mass sent over an edge cannot be greater than the associated edge capacity, i.e.,

$$\pi_t \leq \tilde{C}, \quad (2)$$

for a matrix of capacities $\tilde{C} \in \mathbb{R}_+^{n \times n}$, where $[\tilde{C}]_{ij}$ is the capacity of the edge $(j, i) \in E$ (the inequality should be understood element-wise).

- (c) The mass at a node i at some time instant $t \geq 0$ must not exceed its local storage capacity, i.e.,

$$\rho_t \leq \rho, \quad (3)$$

for a vector of storage capacities $\rho \in \mathbb{R}_+^n$ (again, the inequality is understood entry-wise).

- (d) The mass transported from a node j to a node i cannot exceed the mass held at node j , i.e.,

$$[\pi_t]_{ij} \leq [\rho_t]_j.$$

2.2 Wasserstein Attraction Flows

We formulate the dynamic transport problem described in Section 2.1 as a constrained *Wasserstein attraction* (WA) problem (Peyré 2015, Section 5.2). Our main technical tool will be the JKO flow proximal operators which

we introduce next. We first present the JKO flow proximal operator with respect to a functional f . For all $q \in \text{Prob}(V)$,

$$\text{Prox}_{\tau, f}^{W_1}(q) \triangleq \underset{p \in \text{Prob}(V)}{\text{argmin}} \{W_1(p, q) + \tau f(p)\},$$

where τ is a step-size. Thus, starting from an initial distribution $\rho_0 = \mu$, the discrete JKO flow with respect to f is defined as

$$\rho_{t+1} \triangleq \text{Prox}_{\tau, f}^{W_1}(\rho_t). \quad (4)$$

Wasserstein attraction refers to the flow generated by the implicit gradient steps in (4), known as *JKO stepping*, with respect to the potential function defined as $W_1(\rho_t, \nu)$ for some fixed distribution ν . Informally, the potential function drives the flow to minimize its Wasserstein distance to a *target* distribution. Thus, we define the WA discrete flow as

$$\begin{aligned} \rho_{t+1} &= \text{Prox}_{\tau, W_1(\cdot, \nu)}^{W_1}(\rho_t) \\ &= \underset{p \in \text{Prob}(V)}{\text{argmin}} \{W_1(p, \rho_t) + \tau W_1(p, \nu)\}. \end{aligned} \quad (5)$$

The WA defined in (5) has a precise optimization structure. However, the computation of each proximal operation is computationally intense (Peyré 2015). Moreover, the constraints imposed by the graph are not taken into account. In the next subsection, we describe our proposed approach for the efficient computation of the discrete WA, taking into account the constraints imposed by the network.

2.3 Approximate Wasserstein Attraction Flow on Graphs

Initially, we present the entropy regularized discrete JKO flow for the WA problem following the ideas introduced in (Peyré 2015). The main contribution in (Peyré 2015) is to replace the Wasserstein distance functions with their entropy regularized versions. The use of entropic regularization has been shown useful for the design of computational approaches for OT (Cuturi 2013).

Definition 1 Given a cost matrix $C \in \mathbb{R}_+^{n \times n}$, the **discrete entropy-regularized Wasserstein distance** between $\mu, \nu \in \text{Prob}(V)$ is defined as

$$\mathcal{W}_\gamma(\mu, \nu) = \min_{\pi \in \Pi(\mu, \nu)} \langle C, \pi \rangle + \gamma H(\pi), \quad (6)$$

where $H(\pi) = \sum \pi_{ij} (\ln \pi_{ij} - 1) = \langle \pi, \ln \pi - \mathbf{1} \mathbf{1}^\top \rangle$ is the negative entropy and $\gamma \geq 0$ is the regularization parameter.

Now, we can define the approximate entropy-regularized WA flow as

$$\begin{aligned}\rho_{t+1} &= \text{Prox}_{\tau, \mathcal{W}_\gamma(\cdot, \nu)}^{\mathcal{W}_\gamma}(\rho_t) \\ &= \underset{p \in \text{Prob}(V)}{\text{argmin}} \{ \mathcal{W}_\gamma(p, \rho_t) + \tau \mathcal{W}_\gamma(p, \nu) \}. \quad (7)\end{aligned}$$

Note $\mathcal{W}_\gamma(\cdot, \cdot)$ is a strictly convex and coercive function, therefore the operator in (7) is uniquely defined.

Next, we state one important observation about the entropy-regularized WA flow in (7). Without loss of generality, one can multiply the argument in the optimization problem (7) by a constant $\omega = 1/(1 + \tau)$. Thus, we obtain

$$\rho_{t+1} = \underset{p \in \text{Prob}(V)}{\text{argmin}} \{ \omega \mathcal{W}_\gamma(p, \rho_t) + (1 - \omega) \mathcal{W}_\gamma(p, \nu) \}, \quad (8)$$

which is precisely the entropy-regularized Wasserstein barycenter between ρ_t and ν (Cuturi & Doucet 2014). Recall that for a finite set of probability distributions $\{\mu_i\}_{i=1}^m$ where $\mu_i \in \text{Prob}(V)$, the entropy-regularized Wasserstein barycenter is defined as

$$\mu \triangleq \underset{p \in \text{Prob}(V)}{\text{argmin}} \sum_{i=1}^m \omega_i \mathcal{W}_\gamma(p, \mu_i),$$

where $\omega_i \geq 0$ and $\sum_{i=1}^m \omega_i = 1$.

We interpret the Wasserstein attraction problem as the sequential computation of Wasserstein barycenters. This introduces an additional weight parameter that can be modified to give preference to one measure or the other. Such parameter consequently alters how the mass is transported across the graph, as we illustrate further along this paper.

Note that the barycenter is not restricted to only two distributions but as many as one may need. This means that the solution proposed here could be extended for problems akin to ours but involving more than two distributions, and in turn, we would have several weight parameters to customize the solutions obtained (Tupitsa et al. 2020).

The method that we propose uses *Dijkstra's projection algorithm* (Dijkstra 1983). In our setting, much like Sinkhorn's algorithm, it is easier to implement than more traditional schemes designed to solve mathematical programs.

Another feature of the proposed approach is that, unlike in the computation of the Wasserstein distance (or, for that matter, solving the minimum-cost flow problem), we do not compute the complete flow in a single step, which would also entail having to store the shortest path between each node (or at least the first step

of each path). In this regard, our method not only does not need to store this additional information, but it is also *memoryless* in the sense that, at each step, the algorithm solves a new problem with initial and final distributions. This is advantageous since these measures do not need to be the same as in the previous steps (even the parameters, such as the weights, can be changed). This adaptability is the main difference between the flow we compute, a discrete one, and the one found by solving a minimum-cost flow problem, which is continuous. These aspects might take importance in future works where this method could be adapted in the context of decentralized or distributed optimization, where the available information at each node is limited (Krawtschenko et al. 2020, Dvurechenskii et al. 2018).

Approximate solutions to problems of the form (8) can be efficiently computed by reformulating the entropy-regularized OT problem (6) as

$$W_\gamma(\mu, \nu) = \min_{\pi \in \Pi(\mu, \nu)} \text{KL}(\pi | \xi), \quad (9)$$

where $\xi = e^{-C/\gamma}$ (entry-wise exponential) (Benamou et al. 2015). Note that (9) can be extended for higher dimensional arrays (such as the tuples $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m)$ introduced in the definition of the WB) by summing over the indices (i, j, k, \dots) . Thus, following (Benamou et al. 2015), we can rewrite Problem (8) as

$$\min_{\boldsymbol{\pi} \in \mathcal{C}_f \cap \mathcal{C}_e} \text{KL}_\omega(\boldsymbol{\pi} | \xi) = \omega \text{KL}(\pi_1 | \xi) + (1 - \omega) \text{KL}(\pi_2 | \xi), \quad (10)$$

where

$$\mathcal{C}_f = \{ \pi_1, \pi_2 \mid \pi_1 \mathbf{1} = \rho_t, \pi_2 \mathbf{1} = \nu \}, \quad (11)$$

$$\mathcal{C}_e = \{ \pi_1, \pi_2 \mid \pi_1^\top \mathbf{1} = \pi_2^\top \mathbf{1} = p \}. \quad (12)$$

Finally, taking into account the constraints in (1), (2) and (3) in Problem (10), we can state our main contribution regarding the design of the entropy-regularized discrete WA flow.

Problem 2 Consider a discrete, finite, fixed and connected graph with n vertices, $\tilde{C} \in \mathbb{R}_+^{n \times n}$ the capacity matrix, and $\mu, \nu \in \text{Prob}(V)$ the initial and final distributions respectively. We design the sequence of probability measures $\{\rho_t\}_{t \geq 0}$ by finding, for each $t \geq 0$, the transport plan that solves the optimization problem

$$\{\pi_t\} = \underset{\substack{\boldsymbol{\pi} \in \mathcal{C}_f \cap \mathcal{C}_e \\ \boldsymbol{\pi} \in \mathcal{C}_1 \cap \mathcal{C}_2 \cap \mathcal{C}_3}}{\text{argmin}} \omega \text{KL}(\pi_1 | \xi) + (1 - \omega) \text{KL}(\pi_2 | \xi), \quad (13a)$$

where

$$\mathcal{C}_f = \{\pi \in \mathbb{R}_+^{n \times n} \times \mathbb{R}_+^{n \times n} \mid \pi_1 \mathbf{1} = \rho_t, \pi_2 \mathbf{1} = \nu\} \quad (13b)$$

$$\mathcal{C}_e = \{\pi \in \mathbb{R}_+^{n \times n} \times \mathbb{R}_+^{n \times n} \mid \pi_1^\top \mathbf{1} = \pi_2^\top \mathbf{1} = p\} \quad (13c)$$

$$\mathcal{C}_1 = \{\pi \in \mathbb{R}_+^{n \times n} \times \mathbb{R}_+^{n \times n} \mid \pi_1 \leq \tilde{C}\} \quad (13d)$$

$$\mathcal{C}_2 = \{\pi \in \mathbb{R}_+^{n \times n} \times \mathbb{R}_+^{n \times n} \mid \pi_1^\top \mathbf{1} \leq \rho, \pi_2^\top \mathbf{1} \leq \rho\} \quad (13e)$$

$$\mathcal{C}_3 = \left\{ \pi \in \mathbb{R}_+^{n \times n} \times \mathbb{R}_+^{n \times n} \mid [\pi_1^\top \mathbf{1}]_i \leq \sum_{j:(j,i) \in E} [\rho_t]_j \right\} \quad (13f)$$

We note that the constraint set \mathcal{C}_3 is redundant if in \mathcal{C}_1 we consider $[\tilde{C}]_{ij} = 0$ when nodes i and j are not connected. This is, in fact, what we propose for our procedure in Section 3.3. Nevertheless, we write it explicitly in the problem formulation since it is a necessary constraint that could be imposed differently in other methodologies.

Figure 1 shows a simple example to illustrate the steps we obtain by solving Problem 2. The transport is computed over a path graph, and it starts with an initial distribution (top left) with its mass located in the central nodes, and the final mass (bottom right) is distributed closer to the extremes of the path. At each iteration, we show the resulting distribution found by solving (13) considering the previous solution as the initial measure. We have also considered a storage capacity of 0.3 for the third-to-last node, resulting in partially sending the mass in the fourth iteration. We see how the mass is transported from the initial setting until the final distribution is eventually covered while verifying the constraints imposed in the problem statement addressed here.

3 Iterative Projections for the Computation of Transport Plans

Now that we have the necessary background on discrete OT and have introduced the problem we want to solve, we describe the approach that we propose. We will solve the regularized version of the WB problem, with the additional constraints (1), (2) and (3). To do so, we use a well-known algorithm for solving regularized OT problems called Dykstra's projection algorithm (Dykstra 1983), which, in our setting, is a generalization of the widely used Iterative Bregman Projections (IBP) algorithm (Benamou et al. 2015). We use Dykstra's method because the convergence of IBP cannot be guaranteed in the presence of inequality constraints.

In Section 3.1, we give some background on how this algorithm is used to compute the regularized WB. In Section 3.2, we show how one can modify the algorithm to compute the WB with the added constraints, and finally, in Section 3.3, we move on to the description of the proposed algorithm.

3.1 Computation of the WB using Dykstra's projection algorithm

Dykstra's projection algorithm can be used to solve problems of the form

$$\min_{\pi \in \cap_i \mathcal{C}_i} \text{KL}(\pi \mid \xi),$$

much like Problem 2 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

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

with initial values $\pi^{(0)} = \xi$ and $q^{(0)} = q^{(-1)} = \dots = q^{(-L+1)} = \mathbf{1}\mathbf{1}^\top$. 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 (7) by $\omega = 1/(1+\tau)$, as noted in Section 2.

The next proposition states how we can compute in closed form the proximal operator corresponding to each constraint in the WB problem (10).

Proposition 3 (Proposition 1 in (Benamou et al. 2015))

The proximal operator of the indicator function $\iota_{\mathcal{C}_f}$, corresponding to the constraint set \mathcal{C}_f in (11), has the closed form

$$\left[\text{Prox}_{\iota_{\mathcal{C}_f}}^{\text{KL}\omega}(\pi) \right]_l = \text{Prox}_{\iota_{\{\pi_l \mathbf{1} = P_l\}}}^{\text{KL}}(\pi_l) = \text{diag} \left(\frac{P_l}{\pi_l \mathbf{1}} \right) \pi_l, \quad (14)$$

where $l = 1, 2$ and $P_1 = \rho_t, P_2 = \nu$.

Remark 4 For set \mathcal{C}_f , since the constraint is imposed to each transport plan independently from the rest, we can compute the proximal operator $\text{Prox}_{\iota_{\mathcal{C}_f}}^{\text{KL}\omega}(\pi)$ the same way as with the Wasserstein distance in (Benamou et al. 2015), but individually for each π_l .

Proposition 5 (Proposition 2 in (Benamou et al. 2015))

The proximal operator of the indicator function $\iota_{\mathcal{C}_e}$, corresponding to the constraint set \mathcal{C}_e in (12), has the closed form

$$\left[\text{Prox}_{\iota_{\mathcal{C}_e}}^{\text{KL}\omega}(\pi) \right]_l = \pi_l \text{diag} \left(\frac{p}{\mathbf{1}^\top \pi_l} \right), \quad (15)$$

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

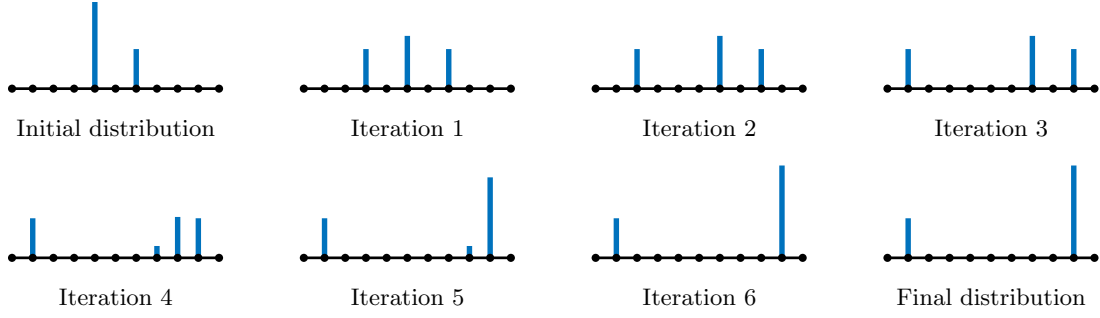


Fig. 1. Illustrative example showing the steps obtained by solving problem 2. The top left plot shows the initial distribution, and on the bottom right, we have the target distribution, which is reached at iteration 6. Note that there is a capacity constraint on the third-to-last node.

3.2 Capacity and support constrained WB

In the context of networks, it is reasonable to restrict how much mass can be sent from one node to another, i.e., to add a capacity to the edges connecting the nodes. This constraint is imposed on each transport plan by defining a capacity matrix $\tilde{C} \in \mathbb{R}^{n \times n}$ such that $[\tilde{C}]_{ij}$ is the maximum mass that can be sent from node i to node j .

Similarly to Proposition 3, since this capacity constraint is imposed on each transport plan independently of the rest, the projection is done individually for each transport plan. The following proposition concerns the computation of the proximal operator for the set \mathcal{C}_1 in (13d).

Proposition 6 (Section 5.2 in (Benamou et al. 2015))
The proximal map for the function $\iota_{\{\pi_1 \leq \tilde{C}\}}$ is defined as

$$\text{Prox}_{\iota_{\{\pi_1 \leq \tilde{C}\}}}^{\text{KL}}(\pi_1) = \min(\pi_1, \tilde{C}), \quad (16)$$

with the minimum computed element-wise.

We can also have capacity limits on some of the nodes, meaning that even though the optimal solution might send a certain amount of mass to one of these nodes, it may not be possible to hold that much quantity. This corresponds to the constraint set \mathcal{C}_2 in (13e). This set is, in fact, the same as one of the sets defined to solve partial transport problems, as seen in (Benamou et al. 2015) (except for having ρ instead of one of the marginals). From that, we get the following result for the computation of the projection on this set in closed form.

Proposition 7 (Proposition 5 in (Benamou et al. 2015))

For the indicator function $\iota_{\mathcal{C}_2}$, corresponding to the constraint set \mathcal{C}_2 in (13e), one has

$$\begin{aligned} [\text{Prox}_{\iota_{\mathcal{C}_2}}^{\text{KL}\omega}(\pi)]_l &= \text{Prox}_{\iota_{\{\pi_l^\top \mathbf{1} \leq \rho\}}}^{\text{KL}}(\pi_l) \\ &= \pi_l \text{diag} \left(\min \left(\frac{\rho}{\pi_l^\top \mathbf{1}}, \mathbf{1} \right) \right), \end{aligned} \quad (17)$$

where the minimum and division of vectors are considered element-wise.

In addition to the capacity constraints (13d) and (13e), we want to restrict the barycenter domain to a smaller set of nodes, rather than the whole graph, since we can only send mass to the nodes in the support of ρ_t and their neighbors. In this case, we would obtain a vector of dimension $n^* \leq n$, where each element corresponds to the mass at one node of the subset. It is clear from the second constraint of the WB problem, $\pi_l^\top \mathbf{1} = p \forall l$ for some measure p , that by resizing p to have dimension n^* , now $\pi_l \in \mathbb{R}^{n \times n^*}$ and, subsequently, for the cost and capacity matrices, we should only take the columns corresponding to the subset of nodes (thus $C_l, \tilde{C}_l \in \mathbb{R}^{n \times n^*}$). Therefore, the dimensions of the arrays in the computation of the projections still agree. However, let us go through the deduction of the computation of the projection on \mathcal{C}_e shown in (15) to see that it is well defined and it still holds for this support constraint ($\text{Prox}_{\iota_{\mathcal{C}_f}}^{\text{KL}\omega}$ is similar and $\text{Prox}_{\iota_{\mathcal{C}_1}}^{\text{KL}\omega}$ and $\text{Prox}_{\iota_{\mathcal{C}_2}}^{\text{KL}\omega}$ are straightforward).

Proposition 8 The computation of the proximal operator $\text{Prox}_{\iota_{\mathcal{C}_e}}^{\text{KL}\omega}(\pi)$ in (15) still holds for $n \times n^*$ dimensional matrix inputs, where $n^* \leq n$.

PROOF. Given $\pi_l^{(k-1)} \in \mathbb{R}^{n \times n^*}$, computing the projection on the set \mathcal{C}_e consists in solving the optimization problem

$$\min_{\pi^{(k)} \in \mathcal{C}_e} \sum_{l=1}^m \omega_l \text{KL} \left(\pi_l^{(k)} \middle| \pi_l^{(k-1)} \right).$$

For the sake of notation, we define $\pi_l := \pi_l^{(k)}$, $\bar{\pi}_l := \pi_l^{(k-1)}$ and, with that, expanding the problem leaves us

with

$$\begin{aligned} \min_{\pi, p} \quad & \sum_{l, i, j} \omega_l \pi_{l, ij} \left(\ln \frac{\pi_{l, ij}}{\bar{\pi}_{l, ij}} - 1 \right) \\ \text{s.t.} \quad & \pi_l^\top \mathbf{1} = p, \quad l = 1, \dots, m. \end{aligned}$$

The Lagrangian of this problem is

$$\sum_l \sum_{i, j} \omega_l \pi_{l, ij} \left(\ln \frac{\pi_{l, ij}}{\bar{\pi}_{l, ij}} - 1 \right) + \lambda_l^\top (\pi_l^\top \mathbf{1} - p),$$

where $\lambda_l \in \mathbb{R}^{n^*} \forall l$ are the Lagrange multipliers.

On one hand, if we differentiate the Lagrangian with respect to $\pi_{l, ij}$ and equate to zero, we get

$$\omega_l \left(\ln \frac{\pi_{l, ij}}{\bar{\pi}_{l, ij}} \right) + \lambda_{l, j} = 0. \quad (18)$$

On the other hand, differentiating with respect to p_j yields

$$-\sum_l \lambda_{l, j} = 0. \quad (19)$$

Isolating $\pi_{l, ij}$ from (18) yields $\pi_{l, ij} = \bar{\pi}_{l, ij} e^{-\frac{\lambda_{l, j}}{\omega_l}}$, thus

$$\pi_l = \bar{\pi}_l \text{diag} \left(e^{-\frac{\lambda_{l, 1}}{\omega_l}}, \dots, e^{-\frac{\lambda_{l, n^*}}{\omega_l}} \right). \quad (20)$$

Then, combining (20) together with the constraint $\pi_l^\top \mathbf{1} = p$, we obtain

$$\text{diag} \left(e^{-\frac{\lambda_{l, 1}}{\omega_l}}, \dots, e^{-\frac{\lambda_{l, n^*}}{\omega_l}} \right) \bar{\pi}_l^\top \mathbf{1} = p, \quad (21)$$

from which we deduce

$$\text{diag} \left(e^{-\frac{\lambda_{l, 1}}{\omega_l}}, \dots, e^{-\frac{\lambda_{l, n^*}}{\omega_l}} \right) = \text{diag} \left(\frac{p}{\bar{\pi}_l^\top \mathbf{1}} \right), \quad (22)$$

where the division is considered element-wise.

We still have to use the result in (19), so, we first rewrite (21) as $(\bar{\pi}_l^\top \mathbf{1})^{\omega_l} = \text{diag} (e^{\lambda_{l, 1}}, \dots, e^{\lambda_{l, n^*}}) p^{\omega_l}$, with element-wise exponentiation. With this relation, we compute $\prod_l (\bar{\pi}_l^\top \mathbf{1})^{\omega_l}$, which is

$$\prod_l (\bar{\pi}_l^\top \mathbf{1})^{\omega_l} = \text{diag} (e^{\sum_l \lambda_{l, 1}}, \dots, e^{\sum_l \lambda_{l, n^*}}) p^{\sum_l \omega_l}.$$

Then, using (19) and the fact that $\sum_l \omega_l = 1$, we can write the measure p in terms of the known quantities $\bar{\pi}_l$ and ω_l as

$$p = \prod_l (\bar{\pi}_l^\top \mathbf{1})^{\omega_l}. \quad (23)$$

Finally, combining (20)–(22)–(23), we obtain

$$\pi_l = \bar{\pi}_l \text{diag} \left(\frac{p}{\bar{\pi}_l^\top \mathbf{1}} \right), \quad \text{where } p = \prod_l (\bar{\pi}_l^\top \mathbf{1})^{\omega_l},$$

which is what we wanted to show. \square

3.3 Description of the proposed approach

Now, we can present the proposed algorithm to solve Problem 2. We use Dykstra's projection algorithm, and together with the support and capacity constraints, we can impose the additional restrictions introduced in the problem statement (Section 2).

For the support constraint (13f), we will take for each matrix only the columns corresponding to the nodes in the support of ρ_t and their neighbors, which we know, since we have the adjacency matrix A . Once we compute ρ_{t+1} , as it might have a smaller dimension $n^* \leq n$, we can redefine ρ_{t+1} as an n -dimensional vector of all zeros except for the nodes that the elements of ρ_{t+1} referred to, which will have the value that we have just computed. While this definition reduces the result to the desired support, nodes in $\text{SUPP}(\rho_t)$ can still send mass to non-neighboring nodes. To fix this issue, we adapt constraint (13d). 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.,

$$[\tilde{C}]_{ij} = \begin{cases} 0 & \text{if } j \in \text{SUPP}(\rho_t) \text{ and } \bar{A}_{ij} = 0, \\ [\tilde{C}]_{ij} & \text{otherwise.} \end{cases} \quad (24)$$

We note that, in this case, constraints (13d) and (13f) could have had a separate matrix for each one and be considered two different projections on the algorithm, but here we merge both into one.

Resizing the matrices to limit the support is unnecessary for the algorithm to converge to the desired solution since it is already taken care of by the capacity matrix (24). Nevertheless, by implementing it, the dimension of the problem can be reduced, so the computations can be executed faster. In the worst-case scenario where all the nodes have mass, there is a direct connection to all the nodes or similar settings, the matrices and vectors would not be modified, and the algorithm would proceed as if this support constraint was not implemented.

Finally, for the storage capacity on each node (13e), we resize ρ to only consider the elements corresponding to the nodes on the new support.

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 in the definition of the Wasserstein distance (6) forces every node to send a small amount of mass to the rest, even if it does not correspond to the distribution described by ν , the solution obtained can be more or less diffused depending on the regularization strength γ . 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 $\gamma(t)$, $\omega(t) \rightarrow 0$ as $t \rightarrow +\infty$ and $\sum_t \omega(t) = +\infty$, we can ensure its convergence (Benamou et al. 2015)(Peyré 2015). We have introduced the second condition on the weight $\omega(t)$ to prevent the parameter from vanishing too quickly. Otherwise, in the computation of the WB, we would obtain the final distribution or one close to it, but the subsequent projections introducing the graph constraints could prevent us from reaching such measure, since we may still have no access to those target nodes. Despite that, in the simulations carried out in Section 4, we consider the weight ω to be both tending to zero (without vanishing too fast) and fixed, since we have observed how, for a constant $\omega < 1/2$, the mass reaches the target distribution as well.

We state the following lemma regarding the convergence of the computation of each intermediate distribution in the discrete flow.

Lemma 9 *For each step t , let \tilde{C} be the capacity matrix defined in (24) such that it verifies $\tilde{C}^\top \mathbf{1} > \rho_t$, 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 3, 5, 6 and 7 converges to the solution of (13a).*

PROOF. The condition $\tilde{C}^\top \mathbf{1} > \rho_t$ ensures that the mass defined by the initial distribution in the t -th step, ρ_t , can be moved or even kept still in some of the nodes in its support. Similarly, if ρ verifies $\rho_t < \rho$, then the same initial distribution ρ_t is a feasible solution. In particular, we have $\text{ri}(\mathcal{C}_f) \cap \text{ri}(\mathcal{C}_e) \cap \text{ri}(\mathcal{C}_1) \cap \text{ri}(\mathcal{C}_2) \cap \text{ri}(\mathcal{C}_3) \neq \emptyset$, where $\text{ri}(\mathcal{C})$ is the relative interior of the set \mathcal{C} . Thus, by Proposition 3.1 in (Peyré 2015), the iterative computation of proximal steps converges to the desired solution. \square

Remark 10 *The conditions on \tilde{C} and ρ are set only to ensure the existence of a feasible solution. Hence, these*

Algorithm 1 Conceptual procedure of the proposed approach

Input: Initial and final distributions ρ_0 and ν , adjacency matrix A , full cost matrix C^* , full vector of storage capacities ρ^* , regularization parameter $\gamma(t)$ and weight $\omega(t)$ depending on t and such that $\gamma(t)$, $\omega(t) \rightarrow 0$ as $t \rightarrow +\infty$, accuracy parameter $\varepsilon > 0$

- 1: $t = 0$
- 2: **while** $\frac{1}{2} \|\nu - \rho_t\|_1 > \varepsilon$ **do**
- 3: Find the support of the new measure ρ_{t+1}
- 4: Define C as the cost matrix C^* but taking only the columns corresponding to the new support
- 5: Define ρ as the vector of storage capacity ρ^* but taking only the elements corresponding to the new support
- 6: Define the capacity matrix \tilde{C} as seen in (24)
- 7: 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 (15), (14) and (17) (with ρ) for both transport plans, and (16) only for transport plan π_1 to enforce the capacity constraint (13d) with capacity matrix \tilde{C}
- 8: $t \leftarrow t + 1$
- 9: **end while**

Output: $\{\rho_t\}_t$

hypotheses could be exchanged for other expressions as long as they are not so restrictive that a solution cannot satisfy all the constraints. The ones proposed in the statement of Lemma 9 are reasonably lax and could be expected in more practical applications.

4 Numerical simulations

In this section, we show numerical simulations that provide evidence for the effectiveness of the proposed approach and analyze its performance. We further show how to implement it to solve a supply and demand problem related to a drinking water network (DWN).

4.1 Implementation details

Before discussing the results obtained in the simulations, some remarks about the implementation of Algorithm 1 are in order. Algorithm 2 shows the detailed steps of Algorithm 1 to solve Problem 2. The computations are carried out using logarithms, as some of the values are of the order of $e^{-1/\gamma}$, so when the regularization is really small we might obtain machine precision issues if we did it outside the logarithmic domain. Moreover, to compute $\ln((\pi_i q_j) \mathbf{1})$ (the product of matrices is element-wise) using $L_{\pi_i} + L_{q_j} = \ln(\pi_i) + \ln(q_j)$, one

Algorithm 2 Detailed implementation of Algorithm 1

Input: Initial and final distributions ρ_0 and ν , adjacency matrix A , full cost matrix C^* , full vector of storage capacities ρ^* , regularization parameter $\gamma(t)$ and weight $\omega(t)$ depending on t and such that $\gamma(t), \omega(t) \rightarrow 0$ as $t \rightarrow +\infty$, accuracy parameter ε

```
1:  $t = 0$ 
2: while  $\frac{1}{2} \|\nu - \rho_t\|_1 > \varepsilon$  do
3:    $\text{supp}_{\text{new}} = \text{SUPP}(\bar{A}\rho_t)$ 
4:   Define  $C$  as the cost matrix  $C^*$  but taking only
     the columns corresponding to  $\text{supp}_{\text{new}}$ 
5:   Define the capacity matrix  $\tilde{C}$  as seen in (24)
6:    $L_{\tilde{C}} = \ln \tilde{C}$ 
7:    $L_{\rho} = \ln \rho$ 
8:    $L_{\pi_1} = L_{\pi_2} = -\frac{1}{\gamma} C$ 
9:    $L_{q_1} = L_{q_2} = L_{q_3} = L_{q_4} = L_{q_5} = L_{q_6} = L_{q_7} = \ln(\mathbf{1}\mathbf{1}^\top)$ 
10:   $L_p = \mathbf{1}$ 
11:   $k = 0$ 
12:  while  $\|L_p - \ln(\pi_1^\top \mathbf{1})\|_1 > \varepsilon$  or  $|\ln(\mathbf{1}^\top(\pi_1 \mathbf{1}))| > \varepsilon$  do
13:     $L'_{\pi_1} = L_{\pi_1}$ 
14:     $L'_{\pi_2} = L_{\pi_2}$ 
15:    if  $k \bmod 4 = 0$  then
16:       $L_{\pi_1} \leftarrow L_{\pi_1} + L_{q_1} + (\ln \rho_t - \ln((\pi_1 q_1)\mathbf{1})) \mathbf{1}^\top$ 
17:       $L_{q_1} \leftarrow L_{q_1} + (L'_{\pi_1} - L_{\pi_1})$ 
18:       $L_{\pi_2} \leftarrow L_{\pi_2} + L_{q_2} + (\ln \nu - \ln((\pi_2 q_2)\mathbf{1})) \mathbf{1}^\top$ 
19:       $L_{q_2} \leftarrow L_{q_2} + (L'_{\pi_2} - L_{\pi_2})$ 
20:    else if  $k \bmod 4 = 1$  then
21:       $L_p \leftarrow \omega \ln((\pi_1 q_3)^\top \mathbf{1}) + (1 - \omega) \ln((\pi_2 q_4)^\top \mathbf{1})$ 
22:       $L_{\pi_1} \leftarrow L_{\pi_1} + L_{q_3} + \mathbf{1} (L_p - \ln((\pi_1 q_3)^\top \mathbf{1}))^\top$ 
23:       $L_{q_3} \leftarrow L_{q_3} + (L'_{\pi_1} - L_{\pi_1})$ 
24:       $L_{\pi_2} \leftarrow L_{\pi_2} + L_{q_4} + \mathbf{1} (L_p - \ln((\pi_2 q_4)^\top \mathbf{1}))^\top$ 
25:       $L_{q_4} \leftarrow L_{q_4} + (L'_{\pi_2} - L_{\pi_2})$ 
26:    else if  $k \bmod 4 = 2$  then
27:       $L_{\pi_1} \leftarrow \min(L_{\pi_1} + L_{q_5}, L_{\tilde{C}})$ 
28:       $L_{q_5} \leftarrow L_{q_5} + (L'_{\pi_1} - L_{\pi_1})$ 
29:    else
30:       $L_{\pi_1} \leftarrow L_{\pi_1} + L_{q_6} + \mathbf{1} (\min(L_{\rho} - \ln((\pi_1 q_6)^\top \mathbf{1})))^\top$ 
31:       $L_{q_6} \leftarrow L_{q_6} + (L'_{\pi_1} - L_{\pi_1})$ 
32:       $L_{\pi_2} \leftarrow L_{\pi_2} + L_{q_7} + \mathbf{1} (\min(L_{\rho} - \ln((\pi_2 q_7)^\top \mathbf{1})))^\top$ 
33:       $L_{q_7} \leftarrow L_{q_7} + (L'_{\pi_2} - L_{\pi_2})$ 
34:    end if
35:     $k \leftarrow k + 1$ 
36:  end while
37:   $\rho_{t+1} = \exp(L_p)$ 
38:  Rewrite  $\rho_{t+1}$  so that it is an  $n$ -dimensional vector
    of all zeros except on  $\text{supp}_{\text{new}}$ 
39:   $t \leftarrow t + 1$ 
40: end while
Output:  $\{\rho_t\}_t$ 
```

can take advantage of the identity $\ln \sum_{i=0}^N a_i = \ln a_0 + \ln \left(1 + \sum_{i=1}^N e^{\ln a_i - \ln a_0}\right)$, where $a_0 \geq a_1 \geq \dots \geq a_N$.

Additionally, for the loop condition at line 12 of Algorithm 2, we have added the second condition $|\ln(\mathbf{1}^\top(\pi_1 \mathbf{1}))| > \varepsilon$, to check if a capacity constraint has been enforced on any position on the transport plan π_1 . This is done to avoid numerical issues where, depending on the precision parameter ε , the first *while* condition might not be verified but the solution has not yet converged to \mathcal{C}_1 in (13d).

4.2 Synthetic examples

To illustrate the steps described in Algorithm 1, in Figure 2, we show a simple example, where we start with a Dirac measure at the center of the graph, whose mass has to be distributed among the outermost nodes. Each subsequent plot shows the intermediate measure obtained after one iteration until the final distribution is reached.

In Figure 3, on the left we plot the total variation distance between the intermediate distribution ρ_t and the target measure ν , for $\omega(t)$ tending to zero at different rates and also fixed at $\omega(t) = 0.1$. In any case, we see how we eventually converge to the final distribution. Due to the symmetrical nature of the network and the probability measures, we observe how for $\omega = 0.1$, since it gives more weight to minimizing the distance to ν rather than the previous distribution, the mass advances until it eventually covers the target in a single step. Similarly, for $\omega(t) = 1/\ln t$, the weight decreases at a slow rate, and so the mass is transported gradually until ω is small enough to cover ν in a single step. For $\omega(t) = 1/t$, the decrease rate is faster, but when it finally starts covering ν , it does so fractionally in a couple of steps, since it is still large enough to give some significant weight to the previous distribution. On the right of Figure 3, we have the cost of transportation (in other words, the Wasserstein distance) of each step, and we observe how the cost adds up to be similar for each case, and we can reach the same conclusions we had with the study of the total variation distance. In particular, we notice how for $\omega(t) = 1/\ln t$, the mass does not move until ω is small enough at the sixth iteration. From there, the transport is similar to what we have for the other cases. As a side note, this is a suitable illustration of the information that the Wasserstein metric can provide with regard to the difference between two measures in the particular domain they are in, which can be lost when using other metrics.

In this and the following examples, we take $\gamma = 10^{-3}$ (except stated otherwise) to have low diffusion. We could take $\gamma(t) \rightarrow 0$ as $t \rightarrow +\infty$, as we have commented earlier. However, the computational speed significantly decreases as the regularization tends to zero, and with this small fixed value, the results obtained have been satisfactory in terms of precision and convergence speed.

Moreover, as we have just commented, apart from taking $\omega(t) \rightarrow 0$ as $t \rightarrow +\infty$, we have also considered a constant

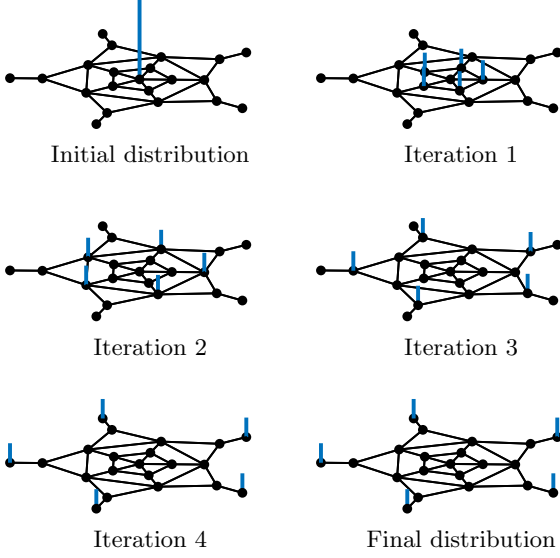


Fig. 2. Simple example showing the steps obtained with the algorithm to go from the (top left) initial distribution to the (bottom right) final one.

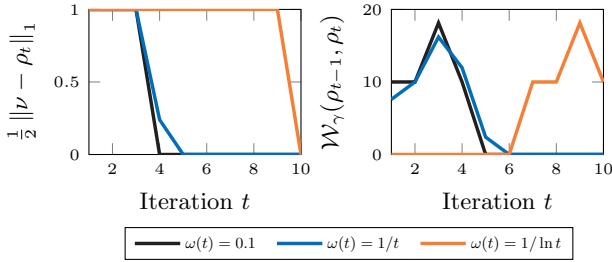


Fig. 3. (Left) Total variation distance between the distribution obtained at iteration t (ρ_t) and the final distribution (ν), and (right) cost of transportation for each iteration, for the example depicted in Figure 2 and taking different weight functions $\omega(t)$.

weight $\omega = 0.1$, in favor of the final distribution. This constant weight parameter plays an important role in how the mass is transported along with the graph, since being closer to the final distribution rather than the previous measure in terms of the Wasserstein distance does not mean that once the mass has moved from ρ_0 to ν , the followed path is the cheapest. To illustrate this case, Figure 4 shows an instance where there are two paths to reach the same node from a certain position, and the whole mass must be sent from one place to the other. Taking $\omega = 0.1$, the left plot shows one iteration without adding extra capacity constraints, which results in the mass being transported through the straight line, as one might surmise. The right plot shows the same setting but with a capacity bound of 0.5 at each link. This value prevents the mass from moving directly to the closest node, and instead, it is forced to be divided and sent through more than one path. With $\omega = 0.1$, the mass tends much more towards the final distribution than the initial one

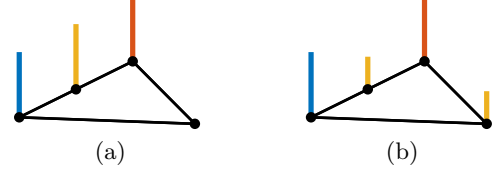


Fig. 4. (a) First step computed without adding physical capacities to the links, and (b) with an additional capacity of 0.5 at each link. We have the initial, final and intermediate distributions in blue, orange, and yellow, respectively.

at each step. This amount of ω forces the mass to be sent through the available paths, which sets the obtained distribution as close as possible to ν while verifying the constraints of Problem (13). However, if ω is increased, giving preference to the initial measure, what we would observe in the first iteration is a certain amount of mass being sent through the straight path and the rest staying in place in the initial position since it is closer than the secondary path. Thus, we move everything through the straight path, taking more steps to reach the final destination, rather than using all the paths at our disposal to finish in fewer steps, which is cheaper. This scenario highlights the potential use of this weight parameter to model the relationship between distributions at play and determine how we move through the graph.

4.3 Real case application: drinking water network

4.3.1 Small case study

Now we proceed to study the case of a DWN. Figure 5 depicts a basic topology of a generic drinking water transport network. The interaction along the most relevant constitutive elements is described by the water supply from the sources towards the network through pumps or valves, depending on the nature of the particular source (either superficial or underground). Therefore, drinking water is moved using manipulated actuators to fill retention tanks and supply water to demand sectors (city neighborhoods). The reader is referred to (Ocampo-Martinez et al. 2013) for further details about this system. Here, this case study is used to discuss and analyze how the proposed approach works and how different parameters can be modified, showing the consequent effects over the whole performance of the considered system.

We note that transporting water through a pipe requiring a pump adds a cost of operation to that edge. This added value can be modeled by including the extra expense into the cost matrix.

Figure 6 shows a simulation on the small network in Figure 5, ignoring the pumps (so, no additional costs on the edges). Here, we again take advantage of the parameter ω to regulate how the water is transported. In particular, in the first step, we use a fairly high weight $\omega = 0.75$ in

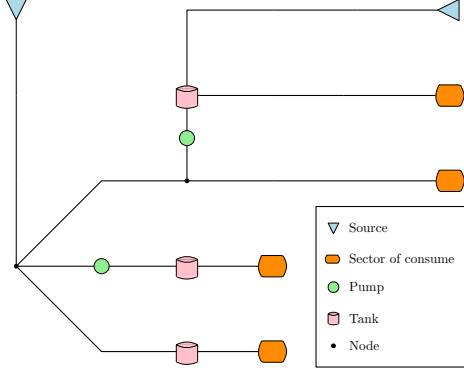


Fig. 5. Topology of the small DWN case study.

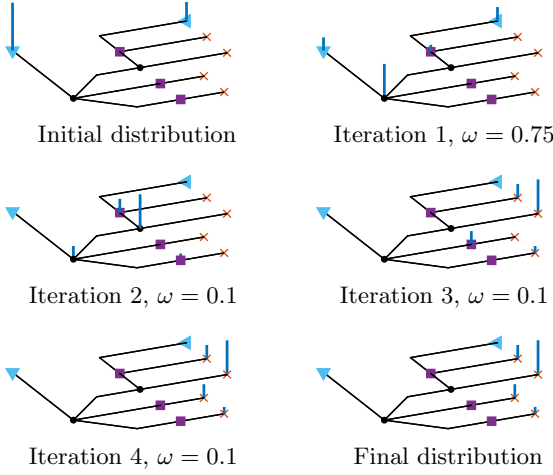


Fig. 6. Steps obtained for the small DWN case study.

favor of the initial distribution so that the transportation is done more gradually. In the following steps, as each one is independent of the preceding iteration, the weight is reduced to $\omega = 0.1$ so that the demand is covered much faster. Similarly to Figure 3, Figure 7 shows the total variation and Wasserstein distance between ρ_t and ν at each iteration t , and we see how we eventually converge to the final distribution with the different weight functions $\omega(t)$ considered. In this case with the chosen network topology and distributions, for $\omega(t) = 0.1$ and $\omega(t) = 1/t$, the transportation is identical.

We have seen that with Algorithm 1 presented as it is, we can account for some additional constraints regarding physical limitations, such as capacities on the pipes or additional costs to operate pumps to be able to send mass between certain locations. Constraint (13e) in particular has been added with DWN-modelling in mind. According to (Ocampo-Martinez et al. 2013), the nodes that are neither tanks nor sources cannot hold as much water, but they do have a certain retention capacity.

Another issue one can find is having to update specific parameters due to external factors, for instance, the initial or final distributions if, for example, there is a sud-

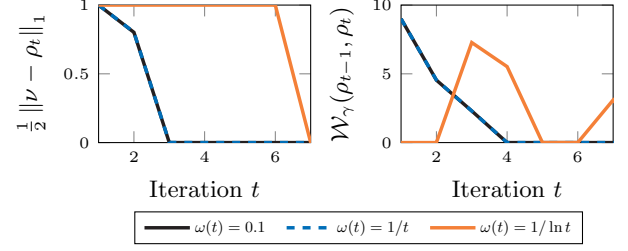


Fig. 7. (Left) Total variation distance between the distribution obtained at iteration t (ρ_t) and the final distribution (ν), and (right) cost of transportation for each iteration, for the example depicted in Figure 6 (taking different weight functions).

den peak in demand, or even the graph topology if a pipe breaks or needs to be cut for maintenance. In the former case, since the scheme is memoryless, the initial and final distributions at any step can be changed, and the algorithm will proceed from there without having to make any modifications to it. For the latter, a change in the topology means that the adjacency and cost matrices are updated, so, as long as these updated values are provided at that step, just as with the change of distributions, the algorithm automatically adapts and proceeds with the computations since the support and capacity constraints are computed at each iteration.

This last case highlights this feature in our approach that we have mentioned several times: each step does not depend on the previous one, which allows the algorithm to adapt to different changes as it advances. If, for example, we wanted to find our sequence of distributions $\{\rho_t\}_{t \geq 0}$ by solving a minimum-cost flow problem, since the flow is computed all at once, each change in the middle of the transportation would mean having to recompute the whole solution (or at least restart taking as the initial measure the distribution obtained at that stage). Simultaneously, with our approach, we only need to update the affected parameters, and the algorithm proceeds from there. Here lies the main difference between our computation of a discrete flow and the continuous flow one would obtain by solving a minimum-cost flow problem.

4.3.2 Performance assessment with the Barcelona drinking water network

To show the effectiveness of the proposed approach, a bigger version of a DWN, particularly the one corresponding to Barcelona (Spain) and its metropolitan area, is considered. In this DWN, the water sources are the Ter and Llobregat rivers regulated at their head by some dams with an overall capacity of 600 cubic hectometres. With four drinking water treatment plants, water from rivers and underground sources (wells) is turned into potable water and served to Barcelona and surrounding towns. Those different water sources currently provide a raw flow of around $7 \text{ m}^3/\text{s}$. Water flow

from each source is limited, implying different water prices depending on water treatments and legal extraction canons.

The Barcelona DWN is structurally organized in two functional layers: an upper layer named *transport network* links the water treatment plants with the reservoirs distributed all over the city, while a lower layer, named *distribution network*, links a reservoir with each consumer sector (water demand). Notice that the upper layer can be managed using control approaches, while the distribution system follows a pre-established behavior given by the water pressure determined. Figure 9 depicts the whole scheme of the transport network.

Our objective is to implement our algorithm for the management of the upper layer. The setting is analogous to what we have seen in Section 4.3.1 for the small case study: we want to find the (discrete) flow that moves the mass from an initial distribution (water provided by the treatment plants and reservoirs) to a target distribution (expected water in the reservoirs to cover the consumers' water demand) such that it follows the sparsity pattern and constraints induced by the network, and each step is the most cost-efficient (depending on the weight parameter ω). By computing the discrete flow, we can also adapt the solution's next step to changes on the network or the other agents.

To perform the simulations, for the initial distribution ρ_0 we have taken the set of source nodes together with close to half of the total amount of tanks (selected at random), assigned them a value following a uniform distribution, and normalized the obtained vector so that $\rho_0 \in \text{Prob}(V)$. The final distribution ν is computed following the same steps with the remaining tanks. For the nodes that are neither tanks nor sources, we have considered that those on the periphery have a retention capacity of 0.05. For the weight parameter, we have tested it first with a small value $\omega = 0.1$ so that the final distribution is reached in fewer iterations, and then with a larger value $\omega = 0.45$, so that the transport is slightly more gradual. Further below we also comment on the convergence when taking $\omega(t) = 1/t$ and $\omega(t) = 1/\ln t$.

For comparison, the sequence $\{\rho_t\}_{t \geq 0}$ is found by solving Problem 2, on one side with Algorithm 1, using different values of the regularization parameter γ , and on the other, using the CPLEX solver, which uses the dual simplex algorithm with the default parameters (**MaxIter** = 9.2234×10^{18} , **TolFun** = 10^{-6}).

Figure 8 shows on the top plot the total variation distance between the final distribution ν and the distribution obtained at every iteration with each method. We notice how with low regularization, the solution obtained is really close (in terms of the total variation distance) to the non-regularized solution obtained with CPLEX,

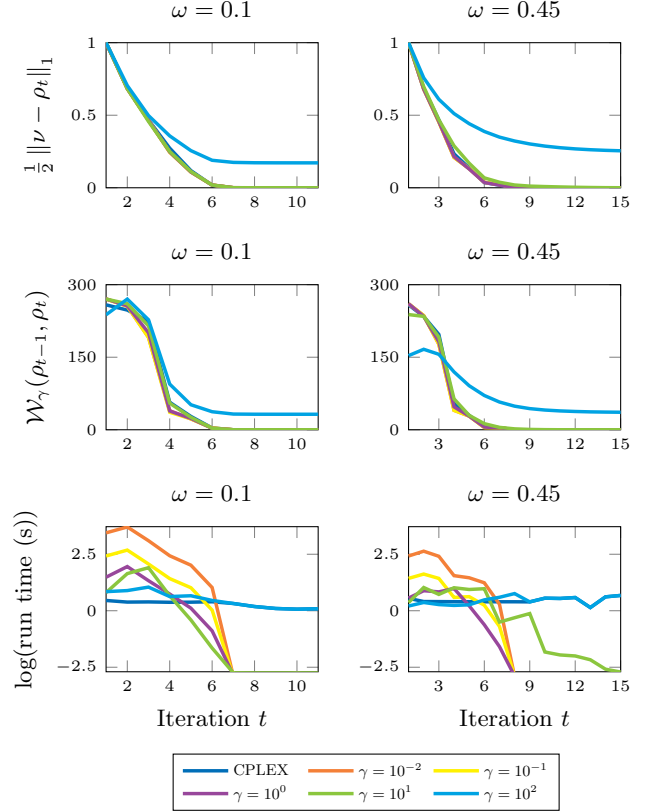


Fig. 8. Performance comparison between Algorithm 1 (using increasing values of the regularization parameter) and CPLEX, using $\omega = 0.1$ (left column) and $\omega = 0.45$ (right column). (Top) Total variation distance between the final distribution ν and the distribution obtained at iteration t (ρ_t), (middle) cost of transportation for each iteration, and (bottom) time elapsed (in seconds) for each iteration (the plot is in logarithmic scale for visualization purposes).

as expected, but even with higher values of the regularization parameter ($\gamma = 1, 10$), there are no noticeable differences, especially in the case with $\omega = 0.1$. However, with higher values ($\gamma = 100$), even though the first iterations are close to the other results, the solution eventually becomes too diffused and is not valid in the setting of DWN. The bottom plot shows the running time of each iteration, i.e., the time elapsed to solve Problem (13) with the new distribution found in the previous step. As expected, the speed of convergence rapidly decreases as $\gamma \rightarrow 0$, which is a known issue with this kind of algorithms (Essid & Solomon 2018). Nonetheless, having seen how with higher regularization, the results obtained are really close even to the CPLEX output, it would be safe to consider a small enough constant γ instead of taking $\gamma(t) \rightarrow 0$ as we do in Algorithm 1, in exchange of higher performance speed and without losing too much accuracy.

Moreover, we have noticed how by removing the capacity constraint to enforce both (13d) and (13f), the algorithm performance vastly improves in terms of conver-

gence speed, which makes sense, considering that it can force sharp changes on the transport plan. In this regard, it would be interesting to find a different approach to improve the computation of the projection with the capacity matrix \tilde{C} in (24), or directly bypass it by rethinking the constraint in terms of the other parameters and variables at play.

Figure 10 shows some selected iterations illustrating how the water is transported towards the target distribution (the bigger the point, the higher the amount of resource is held in that node). Figure 11 shows the total variation distance between ρ_t and ν at each iteration t , taking $\omega(t) = 1/t$ (left) and $\omega(t) = 1/\ln t$ (right). As one might expect, since for $\omega(t) = 1/t$ the weight tends to zero at a higher rate, we reach the solution in fewer iterations than taking $\omega(t) = 1/\ln t$. Since the Barcelona DWN is highly connected to cover the whole city and metropolitan area and accounts for any incidents on the network, we have also carried out simulations in different graphs of similar dimensions (around 10^2 nodes), shown in Figure 11 for comparison. In any case, we observe how the total variation distance eventually converges to zero, taking more steps for the case where the weight decreases slower ($\omega(t) = 1/\ln t$).

From the point of view related to the management of a DWN, in particular, the considered case of Barcelona, the proposed approach opens new ways of improving existent management criteria in the sense of scalability and modularity of the control approaches (Tedesco et al. 2018), apart from adding robustness capabilities to the system. This latter aspect has been previously reported for the particular case given the importance of rejecting the system disturbances and their nature (water costumers demands) (Grosso et al. 2017). In any case, a straightforward comparison with existing methods for management and control of DWNs is nowadays not fair since our approach is presented as a proof of concept for the proposed objectives related to the case study, and then some additional design criteria should be considered.

5 Concluding remarks 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 use Dijkstra’s projection algorithm, which 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.

We have observed how modifying the capacity matrix to avoid sending mass between non-neighboring nodes

forces sharp changes on the transport plan entries, drastically decreasing the execution speed of the algorithm. Future work should be finding an efficient approach to ensure that this condition is verified. However, this paper focuses on the application of these optimal transport concepts in the context of more *real-life* scenarios and how they can automatically adapt to sudden changes in the topology of the networks or the parameters and distributions.

We have illustrated how the value of the weight ω alters how the mass is transported from node to node, even mimicking the behavior we can observe if we implement additional physical capacities on the links. It would be interesting to gain more insight into the weight parameter’s role in shaping the resulting distribution, not only in our setting but also in the multi-marginal case, with several weights. Moreover, the fact that the methodology proposed can be extended to consider more than two distributions and can adapt to different changes could be used to tackle problems involving decentralized or distributed models, where not all the information is available for every agent.

Acknowledgements

This work was partially funded by ARPA-H Strategic Initiative Seed Fund #916012, Sustainable Futures Fund #919027, and the Spanish project PID2020-115905RB-C21 (L-BEST) funded by MCIN/ AEI /10.13039/501100011033.

References

- Ahuja, R., Magnanti, T. & Orlin, J. (1993), *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall.
- Ambrosio, L. & Gigli, N. (2013), *A User’s Guide to Optimal Transport*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–155.
- Ambrosio, L., Gigli, N. & Savare, G. (2008), *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics. ETH Zürich, Birkhäuser Basel.
- Arjovsky, M., Chintala, S. & Bottou, L. (2017), ‘Wasserstein GAN’, *arXiv* **1701.07875**.
- Barré, M., Giron, C., Mazzolini, M. & d’Aspremont, A. (2020), ‘Averaging atmospheric gas concentration data using Wasserstein barycenters’, *arXiv* **2010.02762**.
- Bassetti, F., Gualandi, S. & Veneroni, M. (2020), ‘On the computation of kantorovich–wasserstein distances between two-dimensional histograms by uncapacitated minimum cost flows’, *SIAM Journal on Optimization* **30**(3), 2441–2469.
- Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L. & Peyré, G. (2015), ‘Iterative Bregman projections for

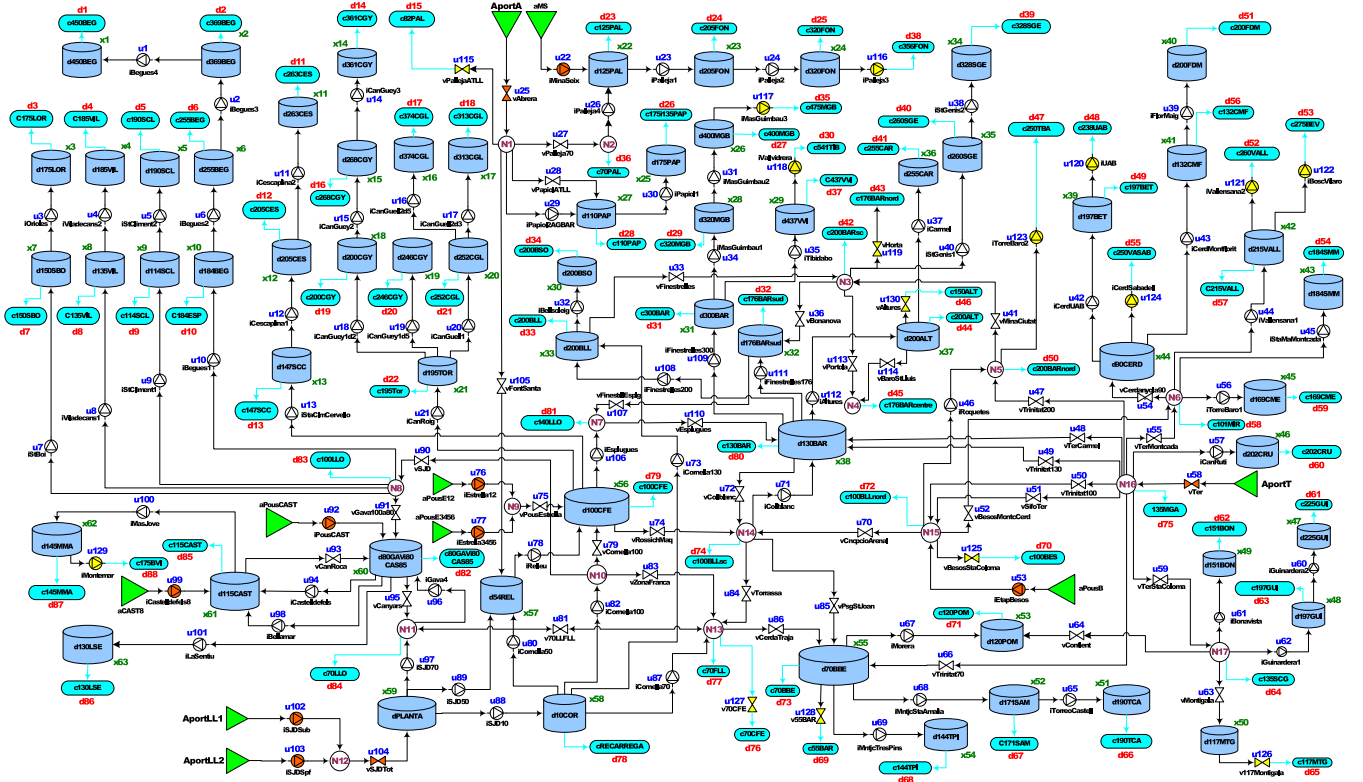


Fig. 9. Scheme of the Barcelona drinking water transport network. Taken from (Ocampo-Martinez et al. 2013).

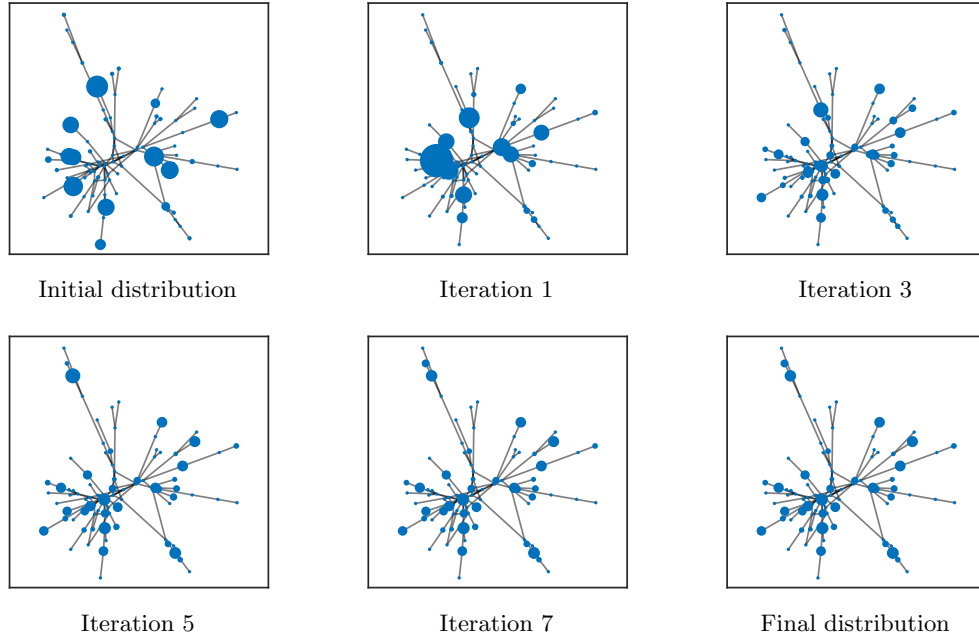


Fig. 10. Selected steps showing the transportation of mass through the Barcelona drinking water network from the (top left) initial distribution to the (bottom right) final distribution. Parameters used: $\omega = 0.1$ and $\gamma = 10^{-1}$.

regularized transportation problems', *SIAM Journal on Scientific Computing* **37**(2), A1111–A1138.
 Brenier, Y. (1991), 'Polar factorization and monotone

rearrangement of vector-valued functions', *Communications on Pure and Applied Mathematics* **44**(4), 375–417.

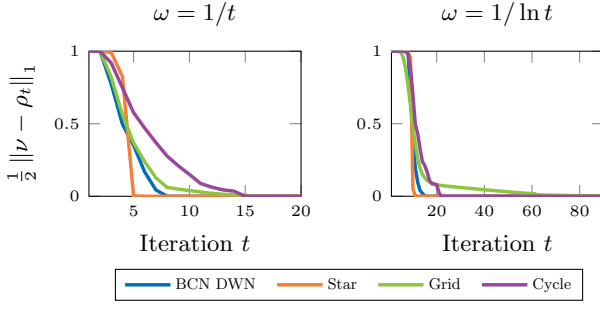


Fig. 11. Total variation distance between the final distribution (ν) and the one computed at iteration t (ρ_t), with $\omega = 1/t$ (left) and $\omega = 1/\ln t$ (right), computed on different graphs, all of size 10^2 . $\gamma = 10^{-1}$ for every case.

Bunne, C., Meng-Papaxanthos, L., Krause, . & Cuturi, M. (2021), ‘Jkonet: Proximal optimal transport modeling of population dynamics’, *arXiv* **2106.06345**.

Chen, Y., Georgiou, T. T., Pavon, M. & Tannenbaum, A. (2018), ‘Efficient robust routing for single commodity network flows’, *IEEE Transactions on Automatic Control* **63**(7), 2287–2294.

Chow, S.-N., Li, W. & Zhou, H.-M. (2017), ‘A discrete Schrodinger equation via optimal transport on graphs’, *arXiv: Dynamical Systems*.

Chzhen, E., Denis, C., Hebiri, M., Oneto, L. & Pontil, M. (2020), ‘Fair regression with wasserstein barycenters’, *arXiv* **2006.07286**.

Cuturi, M. (2013), Sinkhorn distances: Lightspeed computation of optimal transport, in ‘Advances in neural information processing systems’, pp. 2292–2300.

Cuturi, M. & Doucet, A. (2014), ‘Fast computation of Wasserstein barycenters’.

Cuturi, M. & Peyré, G. (2016), ‘A smoothed dual approach for variational Wasserstein problems’, *SIAM J. Imaging Sci.* **9**, 320–343.

Cuturi, M., Teboul, O., Niles-Weed, J. & Vert, J.-P. (2020), ‘Supervised quantile normalization for low-rank matrix approximation’, *arXiv* **2002.03229**.

Dvurechenskii, P., Dvinskikh, D., Gasnikov, A., Uribe, C. & Nedich, A. (2018), Decentralize and randomize: Faster algorithm for wasserstein barycenters, in ‘Advances in Neural Information Processing Systems 31’, Curran Associates, Inc., pp. 10760–10770.

Dykstra, R. L. (1983), ‘An algorithm for restricted least squares regression’, *Journal of the American Statistical Association* **78**(384), 837–842.

Erbar, M., Rumpf, M., Schmitzer, B. & Simon, S. (2020), ‘Computation of optimal transport on discrete metric measure spaces’, *Numerische Mathematik* **144**, 157–200.

Essid, M. & Solomon, J. (2018), ‘Quadratically regularized optimal transport on graphs’, *SIAM Journal on Scientific Computing* **40**(4), A1961–A1986.

Grosso, J., Velarde, P., Ocampo-Martinez, C., Maestre, J. & Puig, V. (2017), ‘Stochastic model predictive control approaches applied to drinking water networks’, *Optimal Control, Applications and Methods*

38(4), 541–558.

Guex, G., Kivimäki, I. & Saeens, M. (2019), ‘Randomized optimal transport on a graph: framework and new distance measures’, *Netw. Sci.* **7**(1), 88–122.

Haasler, I., Chen, Y. & Karlsson, J. (2021), ‘Optimal steering of ensembles with origin-destination constraints’, *IEEE Control Systems Letters* **5**(3), 881–886.

Hu, Y., Zhao, X., Liu, J., Liang, B. & Ma, C. (2020), ‘An efficient algorithm for solving minimum cost flow problem with complementarity slack conditions’, *Mathematical Problems in Engineering* **2020**, 1–5.

Kantorovitch, L. V. (1958), ‘On the translocation of masses’, *Management Science* **5**(1), 1–4.

Kovács, P. (2015), ‘Minimum-cost flow algorithms: an experimental evaluation’, *Optimization Methods and Software* **30**(1), 94–127.

Krawtschenko, R., Uribe, C., Gasnikov, A. & Dvurechensky, P. (2020), ‘Distributed optimization with quantization for computing Wasserstein barycenters’, *arXiv* **2010.14325**.

Mielke, A. (2013), ‘Geodesic convexity of the relative entropy in reversible Markov chains’, *Calculus of Variations and Partial Differential Equations* **48**, 1–31.

Monge, G. (1781), *Mémoire sur la théorie des déblais et des remblais*, De l’Imprimerie Royale.

Morton, K. (1967), ‘A primal method for minimal cost flows with applications to the assignment and transportation problems’, *Management Science* **14**(3), 205–220.

Nejatbakhsh, A., Varol, E., Yemini, E., Hobert, O. & Paninski, L. (2020), Probabilistic joint segmentation and labeling of c. elegans neurons, in ‘International Conference on Medical Image Computing and Computer-Assisted Intervention’, Springer, pp. 130–140.

Nielsen, F., Marti, G., Ray, S. & Pyne, S. (2020), ‘Clustering patterns connecting covid-19 dynamics and human mobility using optimal transport’, *arXiv* **2007.10677**.

Ocampo-Martinez, C., Puig, V., Cembrano, G. & Quevedo, J. (2013), ‘Application of predictive control strategies to the management of complex networks in the urban water cycle’, *IEEE Control Systems Magazine* **33**(1), 15–41.

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

Peyré, G. & Cuturi, M. (2019), ‘Computational optimal transport: With applications to data science’, *Foundations and Trends in Machine Learning* **11**(5-6), 355–607.

Richard, J., David, K. & Felix, O. (1998), ‘The variational formulation of the Fokker-Planck equation’, *SIAM Journal on Mathematical Analysis* **29**(1), 1–17.

Richemond, P. H. & Maginnis, B. (2017), ‘On Wasserstein reinforcement learning and the Fokker-Planck equation’, *ArXiv abs/1712.07185*.

Rubner, Y., Tomasi, C. & Guibas, L. J. (2000), ‘The earth mover’s distance as a metric for image retrieval’,

- International journal of computer vision* **40**(2), 99–121.
- Schrödinger, E. (1931), *Über die Umkehrung der Naturgesetze*, Vol. 9, Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter u. Co., Berlin.
- Solomon, J., de Goes, F., Peyré, G., Cuturi, M., Butscher, A., Nguyen, A., Du, T. & Guibas, L. J. (2015), ‘Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains’, *ACM Trans. Graph.* **34**(4).
- Tedesco, F., Ocampo-Martinez, C., Cassavola, A. & Puig, V. (2018), ‘Centralised and distributed command governor approaches for the operational control of drinking water networks’, *IEEE Transactions on Systems, Man & Cybernetics: Systems* **48**(4), 586–595.
- Tupitsa, N., Dvurechensky, P., Gasnikov, A. & Uribe, C. (2020), ‘Multimarginal optimal transport by accelerated gradient descent’, *arXiv* **2004.02294**.
- Vaserstein, L. N. (1969), ‘Markov processes over denumerable products of spaces describing large systems of automata’, *Problems of Information Transmission* **5**(3), 47–52.
- Villani, C. (2008), *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg.