Networks and Heterogeneous Media

https://www.aimspress.com/journal/nhm

NHM, 20(4): 1108–1144.
DOI: 10.3934/nhm.2025048
Received: 05 April 2025
Revised: 20 September 2025
Accepted: 09 October 2025
Published: 24 October 2025

Research article

Numerical computation of generalized Wasserstein distances with applications to traffic model analysis

Maya Briani[1], **Emiliano Cristiani**[1,*], **Giovanni Franzina**[1] **and Francesca L. Ignoto**[1,2,*]

[1] Istituto per le Applicazioni del Calcolo, Consiglio Nazionale delle Ricerche, Rome, Italy

[2] Dipartimento di Scienze di Base e Applicate per l'Ingegneria, Sapienza Università di Roma, Rome, Italy

* **Correspondence:** Email: e.cristiani@iac.cnr.it, francescalourdes.ignoto@uniroma1.it.

Abstract: Generalized Wasserstein distances allow us to quantitatively compare two continuous or atomic mass distributions with equal or different total masses. In this paper, we propose four numerical methods for the approximation of three different generalized Wasserstein distances introduced in the past few years, giving some insights into their physical meaning. After that, we explore their usage in the context of a sensitivity analysis of differential models for traffic flow. The quantification of the models' sensitivity is obtained by computing the generalized Wasserstein distances between two (numerical) solutions corresponding to different inputs, including different boundary conditions.

Keywords: Generalized Wasserstein distance; Wasserstein distance; traffic modeling; sensitivity analysis; linear programming; nonlinear programming; computational methods

1. Introduction

Context. In this paper, we are concerned with the numerical computation of Wasserstein distance (WD) and generalized Wasserstein distance (GWD) in the context of traffic flow models. As for all dynamical systems, the study of the sensitivity of the models can be realized measuring the 'distance' between two solutions obtained with different inputs (e.g., any of the model's parameters, initial conditions, boundary conditions, etc.). This allows one to understand their impact on the final solution, and ultimately quantify the degree of chaoticity of the system. The question arises as to which distance is more suitable for this kind of investigation. It is by now well understood that L^p distances do not catch the natural concept of distance among traffic (vehicle) densities, see, for example, the discussion in [1, Section 7.1] and Section 2 in this paper, while the WD appears more natural, at least in the context of traffic flow, see, e.g., [1–3]. The drawback is that the WD is limited to *balanced* mass distribution (equal masses), while real traffic problems often need to consider

scenarios with different numbers of vehicles, especially because of different inflow/outflow at boundaries. This suggests a move toward GWDs, which allows one to deal with *unbalanced* mass distributions, but serious computational problems arise: All existing definitions are quite abstract and cannot immediately derive suitable numerical approximations.

Relevant literature. A common feature of various optimal transport distances between finite positive Borel measures on a separable complete metric space $(X, \mathfrak{D})$ is that their primal definition requires the minimization of a transportation cost including an additive contribution of the form

$$\int_{X\times X} c(x,y)\,d\gamma\,, \tag{1.1}$$

where $c : X \times X \to [0, +\infty]$ is a fixed lower semi-continuous cost function.

In the balanced problem, where one wants to define the classical Wasserstein distance between two probability measures μ^{s} and μ^{D}, minimization occurs upon transport plans γ, i.e., the probability measures on $X \times X$ with marginals μ^{s} and μ^{D}. Namely, one considers the projection π^{s} (respectively (resp.), π^{D}) onto the first (resp., the second) factor of the Cartesian product $X \times X$, and takes, as admissible competitors, the finite positive measures γ for which

$$\pi^{\mathrm{s}}_\sharp\gamma = \mu^{\mathrm{s}}\,, \qquad \pi^{\mathrm{D}}_\sharp\gamma = \mu^{\mathrm{D}}\,, \tag{1.2}$$

where $\pi^{\mathrm{s}}_\sharp\gamma$ (resp., $\pi^{\mathrm{D}}_\sharp\gamma$) denotes the push-forward measure of γ under π^{s} (resp., π^{D}).

In order to define distances between finite positive measures on a bounded open set Ω in a finite-dimensional Euclidean space, there are various approaches to similar but *unbalanced* problems that correspond each to the appropriately constrained minimization of a specific objective functional. In this paper, we focus on four of these approaches.

1. The approach by Figalli & Gigli [4], which consists of minimizing the integral cost in Eq (1.1) among all finite positive measures γ on $\overline{\Omega}\times\overline{\Omega}$ whose marginals induce on Ω the prescribed finite positive Borel measures.
2. The approach by Piccoli & Rossi [5], which, in the definition of the functional, includes the additive contributions
$$|\pi^{\mathrm{s}}_\sharp\gamma - \mu^{\mathrm{s}}|(\Omega) + |\pi^{\mathrm{D}}_\sharp\gamma - \mu^{\mathrm{D}}|(\Omega)\,, \tag{1.3}$$
in addition to Eq (1.1), with the minimization occurring on finite positive Borel measures on $\Omega\times\Omega$.
3. A third approach relies on the logarithmic entropy–transport distance that was independently introduced by Kondratyev et al. in [6], by Chizat et al. in [7, 8], and by Liero et al. in [9], called Hellinger–Kantorovich or Wasserstein–Fisher–Rao distance. This approach also relies on the minimization, among finite positive Borel measures on $\Omega \times \Omega$, of a perturbation of Eq (1.1), except that the penalization in Eq (1.3) is replaced by
$$\int_\Omega \Big(\sigma^{\mathrm{s}}\log(\sigma^{\mathrm{s}}) - \sigma^{\mathrm{s}} + 1\Big)\,d\mu^{\mathrm{s}} + \int_\Omega \Big(\sigma^{\mathrm{D}}\log(\sigma^{\mathrm{D}}) - \sigma^{\mathrm{D}} + 1\Big)\,d\mu^{\mathrm{D}}\,, \tag{1.4}$$
if $d(\pi^{\mathrm{s}}_\sharp\gamma) = \sigma^{\mathrm{s}}\,d\mu^{\mathrm{s}}$ and $d(\pi^{\mathrm{D}}_\sharp\gamma) = \sigma^{\mathrm{D}}\,d\mu^{\mathrm{D}}$ for appropriate densities σ^{s} and σ^{D}, and $+\infty$ otherwise. (Incidentally, a similar objective, in which an entropic regularization is added to Eqs (1.1) and (1.4), in considered in [10], where the appropriate dual formulation of the problem is addressed by Sinkhorn iterations.)

4. A fourth approach relates to the contribution by Savaré and Sodini [11], in which the logarithmic entropy transport distance is characterized, in its various equivalent formulations, by convex relaxation. In particular, their approach can be considered the natural one for numerically approximating the (Gaussian) Hellinger–Kanrotovich distance in its homogeneous formulation.

We point out that the last two approaches only differ numerically, as, in principle, the corresponding unbalanced optimal transport formulations are analytically equivalent. Both extend WD to the cone of non-negative Borel measures of finite mass that are different from zero (see Remark 5 below). On the contrary, the first two approaches do give an extension of the WD, despite defining a distance on the cone (see Remarks 3 and 4 below). We recall that no extension of the WD to the set of signed measures, possibly zero, can define a distance [12] (noncanonical optimal transport distances have, however, been successfully defined, see [13, 14]).

Main contribution. This paper has multiple goals. First of all, the four approaches are recalled and compared at the theoretical level to have a common and exhaustive reference for all recent developments about GWDs.

Second, four new *ad hoc* computational methods are devised in order to practically compute the four GWDs (actually, they are three GWDs, since two of them coincide under suitable assumptions). The methods are individually studied and their main properties are discussed through some explanatory numerical tests.

Third, the four approaches are compared on the same academic test in order to understand which one is the most suitable in the context of traffic flow models.

Finally, the three most promising approaches are further compared on four realistic traffic problems.

Anticipating the conclusions, we can say that Figalli & Gigli approach appears to be preferable among them all, mainly for two reasons: (i) It is the most manageable from the computational point of view (easy to code, fast computation), and (ii) it is the most informative on the modeling side.

Plan of the paper. The paper is organized as follows. In Section 2, we recall the main motivations of the paper and we set up notations for linear programming and discrete-to-discrete optimal mass transport. In Section 3, we recall the definition and the basic properties of some known unbalanced optimal transport-type distances between positive Borel measures of finite mass. In Section 4, we revisit the distances in the discrete-to-discrete case and, more importantly, we propose computational approaches for their approximation. The distances are then compared on an academic numerical test. In addition—in the Appendix—we present some tests specifically designed for each distance in which the exact value is available analytically, in order to study their interplay with the numerical boundary and their ability to detect significant patterns. In Section 5, the proposed methods are applied in the context of traffic analysis, in order to achieve various goals: (i) To assess the impact of boundary conditions, (ii) to evaluate the sensitivity to traffic lights, (iii) to compare forecasts depending on whether or not the model takes inertial effects into account, and (iv) to compare the usage of the distances in calibration and validation phases. The paper ends with a final discussion.

2. The Wasserstein distance in the context of vehicular traffic

2.1. Motivations

Usually, the quantification of the closeness of two time-dependent distributions is performed by means of the L^1, L^2 or L^∞ distance (in space at the final time or in both space and time). Although this can be satisfactory for nearly equal outputs or for convergence results, it appears to be inadequate for measuring the distance of largely different outputs. To see this, let us focus on the example depicted in Figure 1, where three density functions ρ^i, $i = 1, 2, 3$, corresponding to the same total mass, say M, are plotted. It is plain that the L^1 distances between ρ_1 and ρ_2 and between ρ_1 and ρ_3 are both equal to $2M$. Similarly, all L^p distances are blind with respect to variation of the densities once their supports are disjoint. Our perception of distance suggests instead that $\|\rho_3 - \rho_1\| > \|\rho_2 - \rho_1\|$, and this is exactly what the WD guarantees, as we will see in the next section.

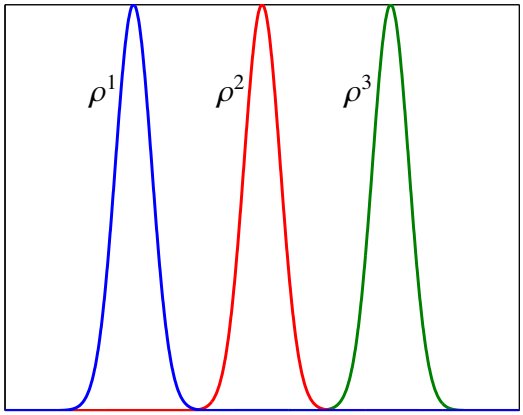

Figure 1. Three density functions with disjoint supports.

2.2. Basic theory

Let us denote by $(X, \mathfrak{D})$ a locally compact complete and separable metric space with a distance $\mathfrak{D}$, and by $\mathcal{B}(X)$ the Borel σ-algebra of $(X, \mathfrak{D})$. Let us also denote by $\mathcal{M}^+(X)$ the set of non-negative finite measures on $(X, \mathcal{B}(X))$. Let μ^{D} (with D standing for the demand or destination) and μ^{S} (with S standing for the supply or source) be two Radon measures in $\mathcal{M}^+(X)$ such that $\mu^{\mathrm{D}}(X) = \mu^{\mathrm{S}}(X)$ (same total mass).

Definition 1 (Wasserstein distance)**.** *For any $p \in [1, +\infty)$, the L^p-Wasserstein distance between μ^{S} and μ^{D} is*

$$W_p(\mu^{\mathrm{S}}, \mu^{\mathrm{D}}) := \left(\inf_{\gamma \in \Gamma(\mu^{\mathrm{S}}, \mu^{\mathrm{D}})} \int_{X \times X} \mathfrak{D}(x, y)^p \, d\gamma(x, y) \right)^{1/p} \tag{2.1}$$

where Γ is the set of transport plans connecting μ^{S} to μ^{D}, i.e.,

$$\Gamma(\mu^{\mathrm{S}}, \mu^{\mathrm{D}}) := \{\gamma \in \mathcal{Q}^+(X \times X) \text{ s.t. } \gamma(A \times X) = \mu^{\mathrm{S}}(A),\ \gamma(X \times B) = \mu^{\mathrm{D}}(B),\ \forall\, A, B \in \mathcal{B}(X)\}\,.$$

Remark 1. *The distance defined by Eq* (2.1) *makes sense for any two given measures μ^{S} and μ^{D} on a finite dimensional normed space, whether they are absolutely continuous or not with respect to the Lebesgue measure.*

It is well known that the notion of Wasserstein distance can be given in relation with the Monge–Kantorovich optimal mass transfer problem [15]. A pile of, say, soil, has to be moved to an excavation with same total volume. Moving a unit quantity of mass has a cost which equals the distance between the source and the destination point. We consequently are looking for a way to rearrange the first mass onto the second which requires a minimum cost.

Remark 2. *In our framework, the mass to be moved corresponds to that of vehicles. We therefore measure the distance between two traffic density distributions by computing the minimal cost to rearrange vehicles from the first configuration to the second one. We also stress that, since we are considering macroscopic models, the vehicles are indistinguishable so we are not able to pinpoint the exact same vehicle in the two scenarios, cf. [3].*

In the particular case of one-dimensional (1D) optimal transport, various characterizations give alternative, more manageable, definitions of the WD. For example, if $(X, \mathfrak{D})$ is $\mathbb{R}$ with the Euclidean distance $|\cdot|$, then, in the case of two Dirac delta functions δ_x, δ_y we simply have $W_p(\delta_x, \delta_y) = \mathfrak{D}(x, y) := |y - x|$. This coincides with the desired distance quantification in the case of only two vehicles on a road, located in x and y respectively. Instead, in the case of two absolutely continuous measures with $d\mu^{\mathrm{s}} = \rho^{\mathrm{s}} dx$ and $d\mu^{\mathrm{D}} = \rho^{\mathrm{D}} dx$, by [16, Remark 2.19], we have

$$W_1(\mu^{\mathrm{s}}, \mu^{\mathrm{D}}) = \int_{\mathbb{R}} |F^{\mathrm{s}}(x) - F^{\mathrm{D}}(x)| dx = \int_0^1 \left| (F^{\mathrm{s}})^{-1}(t) - (F^{\mathrm{D}})^{-1}(t) \right| dt, \tag{2.2}$$

with

$$F^{\mathrm{s}}(x) := \int_{-\infty}^{x} \rho^{\mathrm{s}}(z) dz, \quad F^{\mathrm{D}}(x) := \int_{-\infty}^{x} \rho^{\mathrm{D}}(z) dz \tag{2.3}$$

for $p = 1$, whereas for $p \geq 1$

$$W_p(\mu^{\mathrm{s}}, \mu^{\mathrm{D}}) = \left(\int_{\mathbb{R}} |T^*(x) - x|^p \, \rho^{\mathrm{s}}(x) dx \right)^{1/p}, \tag{2.4}$$

where $T^* : \mathbb{R} \to \mathbb{R}$ satisfies

$$\int_{-\infty}^{x} \rho^{\mathrm{s}}(z) dz = \int_{-\infty}^{T^*(x)} \rho^{\mathrm{D}}(z) dz \quad \forall x \in \mathbb{R}. \tag{2.5}$$

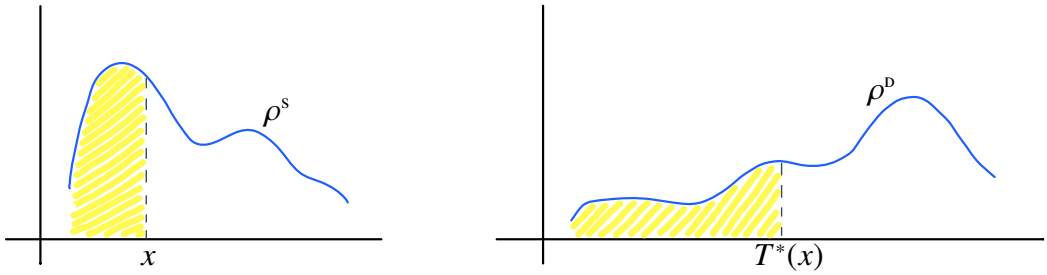

Figure 2. Property of monotone rearrangement.

Equations (2.4) and (2.5) translate the fundamental property of *monotone rearrangement* of the mass [16, Remark 2.19], i.e., the optimal strategy consists of transferring the mass starting from the left; see Figure 2.

2.3. Numerical setting

Let us introduce here the numerical setting that will accompany us throughout the paper; see Figure 3. For simplicity, we will only consider the 1D interval $X = [x_{\text{MIN}}, x_{\text{MAX}}]$, divided into N equispaced cells $C_1, \dots, C_N$ (N is our first parameter of discretization, and we expect the numerical error to vanish for $N \to +\infty$). We also denote by x_i the barycenter of the cell C_i, $i = 1, \dots, N$ and by $\Delta x := x_{i+1} - x_i$ the step of the space discretization. To measure the distance between two cells C_i and C_j, we will always use the Euclidean norm $\mathfrak{D}(x_j, x_k) := |x_k - x_j|$.

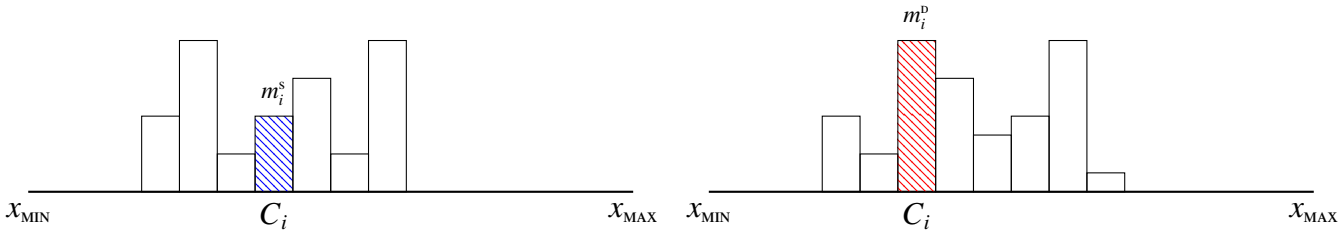

Figure 3. Numerical setting for a generic one-dimensional problem. Two copies of the domain X with the supply mass only (left) and the demand mass only (right).

Moving to the masses, for any $i = 1, \dots, N$ we denote by

$$m_i^{\text{s}} := \rho^{\text{s}}(x_i)\Delta x, \qquad m_i^{\text{D}} := \rho^{\text{D}}(x_i)\Delta x, \tag{2.6}$$

the supply and the demand mass located in C_i, respectively. Roughly speaking, this corresponds to concentrating the masses on specific points, thus getting purely atomic density distributions $\hat{\rho}^{\text{s}} := \sum_i m_i^{\text{s}} \delta_{x_i}$ and $\hat{\rho}^{\text{D}} := \sum_i m_i^{\text{D}} \delta_{x_i}$ where δ denotes the Dirac delta. Note that, in our context, the vectors $m^{\text{s}} := \{m_i^{\text{s}}\}_i$ and $m^{\text{D}} := \{m_i^{\text{D}}\}_i$ are given. Moreover, to avoid any confusion between theoretical and numerical boundary conditions, we will always assume that

$$m_1^{\text{s}} = m_N^{\text{s}} = m_1^{\text{D}} = m_N^{\text{D}} = 0, \tag{2.7}$$

in such a way that the boundaries remain out of play.

For some reasons which will be clear later on, it is also useful to define the *remaining supply mass* at C_i, denoted by $\mathring{m}_i^{\text{s}}$, and the *remaining demand mass* at C_i, denoted by $\mathring{m}_i^{\text{D}}$, $i = 1, \dots, N$.

In order to get a numerical approximation of the WD in $\mathbb{R}^1$, it is convenient to resort to the definition Eq (2.2). Indeed, any quadrature formula is suitable for getting the desired value with arbitrary precision, provided that $\rho^{\text{s}}, \rho^{\text{D}}$ are sufficiently smooth and N is sufficiently large. For example, the naive composite midpoint formula gives

$$W_1(\rho^{\text{s}}, \rho^{\text{D}}) \approx \tilde{W}_1(m^{\text{s}}, m^{\text{D}}) := \Delta x \sum_{j=1}^{N} \left| \sum_{i=1}^{j} m_i^{\text{s}} - \sum_{i=1}^{j} m_i^{\text{D}} \right|. \tag{2.8}$$

The case of more general spaces, including networks embedded in $\mathbb{R}^2$, is more difficult. Typically, the problem is solved by reformulating it in a fully discrete setting as a linear programming (LP) problem. The interested reader can find, in the books by Santambrogio [17, Section 6.4.1] and Sinha

[18, Chapt. 19], the complete procedure, which was also recently used in the context of traffic flow on networks [2]. Even if we will stick with the 1D case, we now recall the crucial steps of this approach because it will be useful in the following.

Following the mass transport interpretation, let c_{jk} be the cost of shipping a unit quantity of mass from the source C_j, $j = 1, \dots, N$, to the destination C_k, $k = 1, \dots, N$. In general, c_{jk} is usually defined as the length of the shortest path joining x_j and x_k. Let γ_{jk} be the (unknown) quantity shipped from the source j to the destination k. The problem is then formulated as

$$\begin{aligned} \text{minimize} \quad & \sum_{j=1}^{N}\sum_{k=1}^{N} c_{jk}\gamma_{jk} \\ \text{subject to} \quad & \sum_{k} \gamma_{jk} = m_j^{\mathrm{s}}, \quad \forall j \\ & \sum_{j} \gamma_{jk} = m_k^{\mathrm{d}}, \quad \forall k \\ & \gamma_{jk} \geq 0. \end{aligned} \tag{2.9}$$

The solution gives an approximation $\tilde{W}_1$ of W_1. Note that the solution satisfies $\gamma_{jk} \leq \min\{m_j^{\mathrm{s}}, m_k^{\mathrm{d}}\}$ since one cannot move more than m_j^{s} from any source j and it is useless to bring more than m_k^{d} to any destination k. From Eq (2.9), it is easy to recover a standard LP problem with equality constraints

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^{\intercal}\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq 0, \end{aligned} \tag{2.10}$$

simply defining

$$\begin{aligned} \mathbf{x} &:= (\gamma_{11}, \gamma_{12}, \dots, \gamma_{1N}, \gamma_{21}, \gamma_{22}, \dots, \gamma_{2N}, \dots, \gamma_{N1}, \dots, \gamma_{NN})^{\intercal} \\ \mathbf{c} &:= (c_{11}, c_{12}, \dots, c_{1N}, c_{21}, c_{22}, \dots, c_{2N}, \dots, c_{N1}, \dots, c_{NN})^{\intercal} \\ \mathbf{b} &:= (m_1^{\mathrm{s}}, \dots, m_N^{\mathrm{s}}, m_1^{\mathrm{d}}, \dots, m_N^{\mathrm{d}})^{\intercal} \end{aligned} \tag{2.11}$$

and $\mathbf{A}$ is the $2N \times N^2$ block matrix

$$\mathbf{A} := \begin{bmatrix} \mathbb{1}_N & \mathbb{0}_N & \mathbb{0}_N & \cdots & \mathbb{0}_N \\ \mathbb{0}_N & \mathbb{1}_N & \mathbb{0}_N & \cdots & \mathbb{0}_N \\ \mathbb{0}_N & \mathbb{0}_N & \mathbb{1}_N & \cdots & \mathbb{0}_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbb{0}_N & \mathbb{0}_N & \mathbb{0}_N & \cdots & \mathbb{1}_N \\ I_N & I_N & I_N & I_N & I_N \end{bmatrix}, \tag{2.12}$$

where I_N is the $N \times N$ identity matrix, $\mathbb{1}_N := (\underbrace{1\ 1\ \cdots\ 1}_{N \text{ times}})$, and $\mathbb{0}_N := (\underbrace{0\ 0\ \cdots\ 0}_{N \text{ times}})$.

3. Basic theory for Generalized Wasserstein distances

In this section, we recall some definitions of metrics on the space of positive finite Borel measures on the locally compact complete and separable metric space $(X, \mathfrak{D})$, inducing the weak/narrow convergence, all based on the distance defined by Eq (2.1) between probability measures.

3.1. Figalli & Gigli approach [4]

In [4], an optimal transport type distance is introduced, with respect to which the heat equation under Dirichlet conditions at the boundary of an open bounded domain Ω can be described as the gradient flow of an entropy, similarly to the case of the heat equation in free space. In order to encode the correct boundary conditions in the minimizing movement scheme, the GWD between two finite non-negative (Borel) measures $\mu^{\mathrm{s}}, \mu^{\mathrm{D}}$ on Ω is defined by

$$W_p^{\mathrm{FG}}(\mu^{\mathrm{s}}, \mu^{\mathrm{D}}) = \left\{ \inf_{\gamma} \int_{\overline{\Omega}\times\overline{\Omega}} \|x-y\|^p d\gamma(x,y) \right\}^{\frac{1}{p}}, \tag{3.1}$$

where the infimum is over

$$\left\{ \gamma \in \mathcal{M}^+(\overline{\Omega}\times\overline{\Omega}) \,\middle|\, (\pi^{\mathrm{s}}_\sharp \gamma) \llcorner \Omega = \mu^{\mathrm{s}}\,,\ (\pi^{\mathrm{D}}_\sharp \gamma) \llcorner \Omega = \mu^{\mathrm{D}} \right\}. \tag{3.2}$$

In the latter, we use the notation $(\bar{\mu} \llcorner \Omega)(A) = \bar{\mu}(A \cap \Omega)$ for the restriction to Ω of a measure $\bar{\mu}$ on $\overline{\Omega}$, and for all given γ, we denote by $\pi^{\mathrm{s}}_\sharp \gamma$ (resp., $\pi^{\mathrm{D}}_\sharp \gamma$) the push-forward of γ under the projection defined on $\overline{\Omega}\times\overline{\Omega}$ by $\pi^{\mathrm{s}}(x,y) = x$ (resp., $\pi^{\mathrm{D}}(x,y) = y$). Equivalently, a non-negative finite Radon measure γ on $\overline{\Omega}\times\overline{\Omega}$ is an admissible competitor for the minimization in Eq (3.1) if and only if

$$\gamma(A\times\overline{\Omega}) = \mu^{\mathrm{s}}(A)\,,\ \gamma(\overline{\Omega}\times B) = \mu^{\mathrm{D}}(B)\,, \qquad \text{for all } A, B \in \mathcal{B}(\Omega). \tag{3.3}$$

Let us note that Eq (3.2) coincides with the set

$$\bigcup_{(\bar{\mu}^{\mathrm{s}}, \bar{\mu}^{\mathrm{D}})} \left\{ \gamma \in \mathcal{M}^+(\overline{\Omega}\times\overline{\Omega}) \,\middle|\, \pi^{\mathrm{s}}_\sharp \gamma = \bar{\mu}^{\mathrm{s}}\,,\ \pi^{\mathrm{D}}_\sharp \gamma = \bar{\mu}^{\mathrm{D}} \right\},$$

where the union is over all pairs $(\bar{\mu}^{\mathrm{s}}, \bar{\mu}^{\mathrm{D}}) \in \mathcal{M}^+(\overline{\Omega}) \times \mathcal{M}^+(\overline{\Omega})$ such that

$$\bar{\mu}^{\mathrm{s}} \llcorner \Omega = \mu^{\mathrm{s}}\,,\ \bar{\mu}^{\mathrm{D}} \llcorner \Omega = \mu^{\mathrm{D}}\,,\ \bar{\mu}^{\mathrm{s}}(\overline{\Omega}) = \bar{\mu}^{\mathrm{D}}(\overline{\Omega})\,.$$

Hence, an equivalent variational definition of the unbalanced optimal transport distance Eq (3.1) is

$$\min \left\{ W_p(\bar{\mu}^{\mathrm{s}}, \bar{\mu}^{\mathrm{D}}) \,\middle|\, \bar{\mu}^{\mathrm{s}} \llcorner \Omega = \mu^{\mathrm{s}}\,,\ \bar{\mu}^{\mathrm{D}} \llcorner \Omega = \mu^{\mathrm{D}}\,,\ \bar{\mu}^{\mathrm{s}}(\partial\Omega) - \bar{\mu}^{\mathrm{D}}(\partial\Omega) = \mu^{\mathrm{s}}(\Omega) - \mu^{\mathrm{D}}(\Omega) \right\}, \tag{3.4}$$

where the classical WD is defined as in Eq (2.1) with $\mathfrak{D}(x,y) = \|x-y\|$ for the probability measures, and is extended by homogeneity to non-negative Radon measures with equal masses. With this definition, if $\bar{\mu}^{\mathrm{s}}, \bar{\mu}^{\mathrm{D}} \in \mathcal{M}^+(\overline{\Omega})$ are fixed, then $W_p(t\bar{\mu}^{\mathrm{s}}, t\bar{\mu}^{\mathrm{D}})$ is arbitrarily large if $t > 0$ is large enough. Hence the minimum in Eq (3.4) is indeed achieved. Notice that the mass unbalance $\mu^{\mathrm{s}}(\Omega) - \mu^{\mathrm{D}}(\Omega)$ is fixed if the marginals μ^{s} and μ^{D} are also fixed.

The authors of [4] deal explicitly with the case of a quadratic unitary cost, i.e., $c(x,y) = \|x-y\|^2$ in Eq (3.1), but their main results also hold for a more general cost, e.g., $c(x,y) = \|x-y\|^p$, with $p \geq 1$. As intermediate steps toward their main results, they prove lower semicontinuity with respect to the weak-$*$ convergence on Ω and provide a gluing lemma implying the triangle inequality, so that Eq (3.1) defines a distance on $\mathcal{M}^+(\Omega)$, and the infimum is achieved by an appropriate optimal plan γ.

The idea of minimizing Eq (3.1) among all γ as in Eq (3.2) is that $\partial\Omega$ models an infinite mass reservoir and the transportation policy described by the convenient competitors γ might include the

option to cut out part of the material from the actual transference plan, by transferring it to $\partial\Omega$, as well as the possibility of integrating the transportation plan by transferring material from $\partial\Omega$ to the final destination, provided that the transportation cost to and from $\partial\Omega$ is accounted for in the overall integral cost, together with that of moving the mass that is authentically transported. A dual formulation of the problem leads to a generalization of Kantorovich potentials and Kantorovich norms, which have been already investigated especially in the case of a linear cost ($p = 1$); see [19, 20].

Remark 3. *It is clear from Eq* (3.4) *that the distance between the positive finite Radon measures introduced in [4] is not an extension of the classical WD: If the two prescribed marginals are concentrated close to the boundary with the supports being far apart from each other, then in solving Eq* (3.4)*, there will be a preference for $\bar{\mu}^{\mathrm{s}}$ and $\bar{\mu}^{\mathrm{D}}$ in charging appropriately two corresponding disjoint boundary regions, after which the WD to be minimized will be achieved by transport plans that only move mass from or to the boundary. In fact, the goal of [4] was to generalize Otto's calculus to partial differential equations (PDEs) under Dirichlet conditions, rather than extending the optimal transport metrics to finite positive Radon measures.*

3.2. Piccoli & Rossi approach [5]

In reference [5], the set of positive finite Borel measures on Ω is endowed with a metric, defined by

$$W_p^{\mathrm{PR}}(\mu^{\mathrm{s}},\mu^{\mathrm{D}};a,b) := \inf_{\gamma}\left\{a\left|\mu^{\mathrm{s}}(\Omega)-\pi^{\mathrm{s}}_{\sharp}\gamma(\Omega)\right| + a\left|\mu^{\mathrm{D}}(\Omega)-\pi^{\mathrm{D}}_{\sharp}\gamma(\Omega)\right| + b\int_{\Omega\times\Omega}\|x-y\|^p\,d\gamma\right\}^{\frac{1}{p}}, \tag{3.5}$$

where $a, b > 0$ are fixed parameters and the infimum is among all finite positive Borel measures γ on $\Omega\times\Omega$.

Note that Eq (3.5) can be equivalently written in the form

$$W_p^{\mathrm{PR}}(\mu^{\mathrm{s}},\mu^{\mathrm{D}};a,b) = \inf\left\{a\left(\left|\mu^{\mathrm{s}}(\Omega)-\mathring{\mu}^{\mathrm{s}}(\Omega)\right| + \left|\mu^{\mathrm{D}}(\Omega)-\mathring{\mu}^{\mathrm{D}}(\Omega)\right|\right) + bW_p(\mathring{\mu}^{\mathrm{s}},\mathring{\mu}^{\mathrm{D}}) \;\middle|\; \mathring{\mu}^{\mathrm{s}}(\Omega)=\mathring{\mu}^{\mathrm{D}}(\Omega)\right\}, \tag{3.6}$$

where μ^{s} and μ^{D} form an arbitrary pair of positive finite Borel measures on Ω with equal mass. In the right-hand side of Eq (3.6), the term $W_p(\mathring{\mu}^{\mathrm{s}},\mathring{\mu}^{\mathrm{D}})$ is understood to be the value of the WD between probability measures taken by the extension by homogeneity to pairs of finite positive measures of equal mass.*

The infimum in Eq (3.6), and hence in Eq (3.5), is indeed achieved and it defines a distance [5, Prop. 1].

The mass transfer may include either physical transportation, or creation and destruction. Adding mass to and removing it from μ^{s}, as well as from μ^{D}, are both allowed actions, with the implied costs being weighted by the parameter a. After these actions have been taken, the supplies eligible for transport, as well as the actual demands, are established and the corresponding optimal transfer policy contributes to the overall cost, via a classical WD weighted by b.

The constraints $\mathring{\mu}^{\mathrm{s}}(\Omega) \le \mu^{\mathrm{s}}(\Omega)$ and $\mathring{\mu}^{\mathrm{D}}(\Omega) \le \mu^{\mathrm{D}}(\Omega)$ may be added in the minimization without changing the minimum value; see [21, Prop. 2].

*Namely, we set $W_p(\mathring{\mu}^{\mathrm{s}},\mathring{\mu}^{\mathrm{D}}) = mW_p(\mathring{\mu}^{\mathrm{s}}/m,\mathring{\mu}^{\mathrm{D}}/m)$, where $m = \mathring{\mu}^{\mathrm{D}}(\Omega) = \mathring{\mu}^{\mathrm{s}}(\Omega)$ and $\mathring{\mu}^{\mathrm{s}},\mathring{\mu}^{\mathrm{D}}$, are thought of as measures on $\overline{\Omega}$ that vanish on all measurable subsets of the boundary $\partial\Omega = \overline{\Omega}\setminus\Omega$.

Remark 4. *Heuristically, all combinations of the actions above are admissible for the minimization. An arbitrary fraction of mass may be removed from the marginals, along with transporting the remaining mass. Accordingly, Eq* (3.5) *does not extend the WD because in the case of two marginals with equal masses, the permissible actions include the possibility of removing mass from both. On the other hand, in the case of* $p = 1$ *with* $a = b = 1$*, it still offers a natural generalization of the* 1*-WD, being equivalent to the flat metric.*

The Piccoli & Rossi distance Eq (3.5) differs from that of Figalli & Gigli Eq (3.1). Recalling Eq (3.4), it is clear that, in the latter, the mass added to the marginals is located at the boundary and that no other pay-off is required, except for the transportation cost between all locations, including those at the boundary. By contrast, the boundary's geometry plays no role in Eq (3.5).

On the other hand, at least in the case of a quadratic cost ($p = 2$), it is known (see [7, Prop. 1.1]) that Eq (3.5) is equivalent to the partial optimal transport problem [22, 23], i.e.,

$$\inf_{\gamma} \int_{\Omega\times\Omega} \Big(\|x-y\|^2 - \lambda \Big)\, d\gamma\,,$$

where the infimum is among all non-negative Radon measures on $\Omega\times\Omega$, whose left and right marginals are dominated by μ^{s} and μ^{D}, which are subject to have a fixed overall mass m, and the Lagrange multiplier for this constraint satisfies $\lambda = 2a^2/b^2$.

Furthermore, the distance Eq (3.5) is inspired by the dynamic formulation of the Monge–Kantorovich problem: By generalizing the Benamou and Brenier's approach [24], the supply and the demand distributions are thought of as the initial and the final prescribed configurations for a fluid motion, and the minimal value of Eq (3.5) is obtained by minimizing the total kinetic energy necessary to connect the two configurations so as to balance mass conservation with the creation and destruction of mass, thanks to the presence of an additional source term in a differential constraint for the macroscopic fluid variables.

3.3. Gaussian Hellinger–Kantorovich approach [9]

Another distance between positive finite Borel measures on Ω is defined by

$$W_2^{\mathrm{GHK}}(\mu^{\mathrm{s}},\mu^{\mathrm{D}};a,b) := \inf\left\{ a\,\mathcal{U}(\pi^{\mathrm{s}}_\sharp\gamma \,|\, \mu^{\mathrm{s}}) + a\,\mathcal{U}(\pi^{\mathrm{D}}_\sharp\gamma \,|\, \mu^{\mathrm{D}}) + b\int_{\Omega\times\Omega} \|x-y\|^2\, d\gamma \right\}^{\frac{1}{2}}, \tag{3.7a}$$

with $a, b > 0$ being fixed parameters, and the infimum is among all positive finite measures γ on $\Omega\times\Omega$.

Here

$$\mathcal{U}(\tilde{\mu}\,|\,\mu) = \begin{cases} \displaystyle\int U(\sigma)\, d\mu\,, & \text{if } d\tilde{\mu} = \sigma d\mu\,, \\ +\infty\,, & \text{otherwise,} \end{cases}, \qquad \text{where } U(\sigma) = \sigma\log(\sigma) - \sigma + 1\,. \tag{3.7b}$$

In reference [9], such a distance is called the *Gaussian Hellinger–Kantorovich distance* in view of the occurrence of a Gaussian function in its homogeneous formulation; see Section 3.4 below. We note that Eq (3.7) can be equivalently written in the form

$$W_2^{\mathrm{GHK}}(\mu^{\mathrm{s}},\mu^{\mathrm{D}};a,b) = \inf\left\{ a\,\mathcal{U}(\hat{\mu}^{\mathrm{s}} \,|\, \mu^{\mathrm{s}}) + a\,\mathcal{U}(\hat{\mu}^{\mathrm{D}} \,|\, \mu^{\mathrm{D}}) + bW_2^2(\hat{\mu}^{\mathrm{s}},\hat{\mu}^{\mathrm{D}}) \right\}^{\frac{1}{2}}, \tag{3.8}$$

where the infimum is among all positive finite measures $\hat{\mu}^{\mathrm{s}}$ and $\hat{\mu}^{\mathrm{D}}$ of equal mass on Ω. In order to define the last term on the right-hand side of Eq (3.8), we have implicitly understood both $\hat{\mu}^{\mathrm{s}}$ and $\hat{\mu}^{\mathrm{D}}$ as multiples of probability measures on $\overline{\Omega}$ vanishing on $\partial\Omega$ and we have defined their WD accordingly, with the obvious extension by homogeneity to pairs of finite positive measures of equal mass on Ω of the WD between probability measures.

When the given measures μ^{s} and μ^{D} themselves happen to have equal mass, they form an admissible pair for the minimization: The right-hand side of Eq (3.8) is then the classical (quadratic) WD, whose the value is a computed by solving a balanced optimal transport problem, because in view of Eq (3.7b), we have $U(1) = 0$. In the unbalanced case, the two additional non-negative contributions in the objective penalize the deviation of the competitors from the prescribed measures.

The variational principle Eq (3.7) belongs to the wide class of *optimal entropy-transport problems* considered in [9], in which the penalization is proportional to the *Csiszár–Morimoto divergence*, defined in terms of proper convex functions $U : [0, +\infty) \to [0, +\infty]$, called the *entropy functions*, possibly different from Eq (3.7b). That large class covers the Piccoli & Rossi distance Eq (3.6) for $p = 1$, with the entropy function $U(\sigma) = |\sigma - 1|$, and it comprises several alternative models to [5] but equally inspired by dynamical formulations with source terms in the differential constraints, which have appeared in the subsequent literature.

In particular, the model case of the *logarithmic optimal entropy-transport problem* in Eq (3.7) has been investigated under various perspectives: Its value coincides with that of the metrics studied in [6, 7] (also called the Wasserstein–Fisher–Rao distance, as it interpolates between the quadratic Wasserstein and the Fisher–Rao distances), for which a static definition had been provided in [8].

Remark 5. *It is known, e.g., by [9, Part II], that the logarithmic entropy-transport minimization Eq* (3.7) *defines a distance that metrizes the weak/narrow convergence of the non-negative finite measure, extending the quadratic WD (otherwise defined only for probability measures).*

3.4. Savaré & Sodini approach [11]

In reference [9, Part II] the logarithmic entropy-transport minimization Eq (3.7) is proved to be equivalent to an optimal transport type problem in the *metric cone* $\mathfrak{C}[X]$ over X. More precisely, one defines $\mathfrak{C}[X]$ as the quotient of $X \times [0, +\infty)$ obtained by collapsing the whole fiber $X \times \{0\}$ to a single element, endowed with a distance that is defined, once an appropriate function g is chosen, by setting

$$d_{\mathfrak{C},g}([x^{\mathrm{s}}, r^{\mathrm{s}}], [x^{\mathrm{D}}, r^{\mathrm{D}}]) = \sqrt{H_g(x^{\mathrm{s}}, r^{\mathrm{s}}; x^{\mathrm{D}}, r^{\mathrm{D}})}\,,$$

where

$$H_g(x, r; y, s) = r + s - 2\sqrt{rs}\, g(\|x - y\|) \quad \text{and} \quad g(t) = e^{-\frac{t^2}{2}}\,. \tag{3.9}$$

One then considers the minimization of

$$\int_{\mathfrak{C}[X]\times\mathfrak{C}[X]} d^2_{\mathfrak{C},g}([x^{\mathrm{s}}, r^{\mathrm{s}}], [x^{\mathrm{D}}, r^{\mathrm{D}}])\, d\alpha$$

among positive measures α on $\mathfrak{C}[X] \times \mathfrak{C}[X]$ with *homogeneous marginals* μ^{s} and μ^{D}, i.e., such that

$$\mu^{\mathrm{s}}(A) = \int_{\mathfrak{C}[A]\times\mathfrak{C}[X]} r^{\mathrm{s}}\, d\alpha([x^{\mathrm{s}}, r^{\mathrm{s}}], [x^{\mathrm{D}}, r^{\mathrm{D}}])\,, \quad \mu^{\mathrm{D}}(B) = \int_{\mathfrak{C}[X]\times\mathfrak{C}[B]} r^{\mathrm{D}}\, d\alpha([x^{\mathrm{s}}, r^{\mathrm{s}}], [x^{\mathrm{D}}, r^{\mathrm{D}}])$$

for all Borel sets $A, B \subset X$.

The constrained minimum is achieved, and it coincides with the Gaussian Hellinger–Kantorovich distance defined by Eq (3.7) in the case of $a = b = 1$.[†]

4. Discrete and computational approaches for GWDs

In this section, we first revise the GWDs previously introduced in the special case of atomic measures (sum of Dirac delta's). This allows us to naturally introduce a discretization in the space domain, which is the first step toward a numerical computation. Second, we propose original computational approaches to get algorithms that are ready to be implemented on a computer. In the final Section 4.5, we present a comparative numerical test of all the proposed approaches. Specific tests for each GWD are given later in Appendix A.

4.1. Figalli & Gigli approach

4.1.1. Discrete approach

In the special case of two purely atomic marginals with finite supports in a 1D domain, i.e.,

$$\mu^{\mathrm{s}} = \sum_{i=2}^{N-1} m_i^{\mathrm{s}} \delta_{x_i}\,, \qquad \mu^{\mathrm{D}} = \sum_{i=2}^{N-1} m_i^{\mathrm{D}} \delta_{x_i}\,,$$

we get an interpretation of Eq (3.4) as a linear programming problem (the indices 1 and N are discarded because of Eq (2.7)). Namely, we consider the problem

$$\min_{(m_1^{\mathrm{s}}, m_N^{\mathrm{s}}, m_1^{\mathrm{D}}, m_N^{\mathrm{D}}) \in \mathbb{R}^4} \left\{ W_p^p\Big(\sum_{i=1}^{N} m_i^{\mathrm{s}} \delta_{x_i}\,,\, \sum_{i=1}^{N} m_i^{\mathrm{D}} \delta_{x_i} \Big) \;\Big|\; m_1^{\mathrm{s}} + m_N^{\mathrm{s}} - (m_1^{\mathrm{D}} + m_N^{\mathrm{D}}) = \Delta M \right\},$$

where

$$\Delta M := \sum_{i=2}^{N-1} m_i^{\mathrm{s}} - \sum_{i=2}^{N-1} m_i^{\mathrm{D}}.$$

After setting $c_{jk} = \|x_j - x_k\|^p$, for all $j, k \in \{1, \dots, N\}$, we write it in the equivalent form

$$\begin{aligned} \tilde{W}_p^{\mathrm{FG}}(m^{\mathrm{s}}, m^{\mathrm{D}}) := \min \Bigg\{ \sum_{j,k=1}^{N} c_{jk}\gamma_{jk} \;\Big|\; \gamma \in [0,+\infty)^{N\times N}\,,\; & \sum_{k=1}^{N} \gamma_{jk} = m_j^{\mathrm{s}}\,,\ \text{for } j \in \{2,\dots,N-1\}\,, \\ & \sum_{j=1}^{N} \gamma_{jk} = m_k^{\mathrm{D}}\,,\ \text{for } k \in \{2,\dots,N-1\}, \\ & \sum_{k=1}^{N} (\gamma_{1k} + \gamma_{Nk}) - \sum_{j=1}^{N} (\gamma_{j1} + \gamma_{jN}) = \Delta M \Bigg\}^{\frac{1}{p}}. \end{aligned}$$

That amounts to the minimization of a linear objective in a convex polytope obtained by intersecting $2N - 3$ affine half-spaces in $\mathbb{R}^{N\times N}$ with $[0, +\infty)^{N\times N}$. Alternatively, the last constraint may be converted into an additional contribution to the total cost involving suitable Lagrange multipliers.

[†]A match is also possible for different choices of the parameters a, b, provided that the definition of H_g changes accordingly in Eq (3.9). Moreover, the choice $g(t) = \cos\left(t^2 \wedge \frac{\pi}{2}\right)$ is also possible in Eq (3.9), leading to the *Hellinger–Kantorovich distance*, which equals Eq (3.7a) if Eq (3.7b) is replaced by the entropy function $U(\sigma) = \sigma - 1 - \log\sigma$.

4.1.2. Computational approach

The minimization problem introduced in the previous section can be formulated as an LP problem in the standard form Eq (2.10) defining

$$\begin{aligned}\mathbf{x} &:= (\gamma_{11}, \gamma_{12}, \dots, \gamma_{1N}, \gamma_{21}, \gamma_{22}, \dots, \gamma_{2N}, \dots, \gamma_{N1}, \dots, \gamma_{NN})^{\intercal}\\ \mathbf{c} &:= (c_{11}, c_{12}, \dots, c_{1N}, c_{21}, c_{22}, \dots, c_{2N}, \dots, c_{N1}, \dots, c_{NN})^{\intercal}\\ \mathbf{b} &:= (m_2^{\mathrm{s}}, \dots, m_{N-1}^{\mathrm{s}}, m_2^{\mathrm{d}}, \dots, m_{N-1}^{\mathrm{d}}, \Delta M)^{\intercal}\end{aligned}$$

and $\mathbf{A}$ is the $(2N-3)\times N^2$ block matrix

$$\mathbf{A} := \begin{bmatrix} \mathbb{0}_N & \mathbb{1}_N & \mathbb{0}_N & \cdots & \mathbb{0}_N & \mathbb{0}_N \\ \mathbb{0}_N & \mathbb{0}_N & \mathbb{1}_N & \cdots & \mathbb{0}_N & \mathbb{0}_N \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbb{0}_N & \mathbb{0}_N & \mathbb{0}_N & \cdots & \mathbb{1}_N & \mathbb{0}_N \\ J_N & J_N & J_N & \cdots & J_N & J_N \\ U_N & Z_N & Z_N & \cdots & Z_N & U_N \end{bmatrix}, \tag{4.1}$$

where I_N is the $N\times N$ identity matrix, $\mathbb{1}_N := (\underbrace{1\ 1\ \cdots\ 1}_{N \text{ times}})$, $\mathbb{0}_N := (\underbrace{0\ 0\ \cdots\ 0}_{N \text{ times}})$, $Z_N := (-1\ \underbrace{0\ 0\ \cdots\ 0}_{N-2 \text{ times}} -1)$, $U_N := (0\ \underbrace{1\ 1\ \cdots\ 1}_{N-2 \text{ times}}\ 0)$, and J_N is the $(N-2)\times N$ block matrix $[\mathbb{0}_{N-2}^{\intercal}\ I_{N-2}\ \mathbb{0}_{N-2}^{\intercal}]$.

The LP formulation described above can be slightly generalized to account for the costs of creating/destroying mass at the boundaries. Still keeping the simplifying assumption Eq (2.7), the vector $\mathbf{x}$ of the unknowns is modified by adding four new unknowns M_1^{s}, M_N^{s}, M_1^{d}, and M_N^{d}, corresponding to the masses taken from the left/right boundaries, and the masses brought to the left/right boundaries, respectively. Note that these masses are potentially unlimited. Moreover, four input parameters c_1^{s}, c_N^{s}, c_1^{d}, and c_N^{d} related to the four new unknowns are added. For example, the quantity $c_1^{\mathrm{s}}M_1^{\mathrm{s}}$ will give the cost to create the mass M_1^{s} at the left boundary (just creating it, not transporting it inside the domain), while the quantity $c_N^{\mathrm{d}}M_N^{\mathrm{d}}$ will give the cost of destroying the mass M_N^{d} at the right boundary.

The LP formulation becomes

$$\begin{aligned}\tilde{W}_p^{\mathrm{FG}'}(m^{\mathrm{s}}, m^{\mathrm{d}}) := \min\Bigg\{ &\sum_{j,k=1}^{N} c_{jk}\gamma_{jk} + c_1^{\mathrm{s}}M_1^{\mathrm{s}} + c_N^{\mathrm{s}}M_N^{\mathrm{s}} + c_1^{\mathrm{d}}M_1^{\mathrm{d}} + c_N^{\mathrm{d}}M_N^{\mathrm{d}} \;\Big|\; \gamma \in [0,+\infty)^{N\times N}\,,\\ &\sum_{k=1}^{N}\gamma_{jk} = m_j^{\mathrm{s}}\,,\ \text{for } j\in\{2,\dots,N-1\}\,, \quad \sum_{j=1}^{N}\gamma_{jk} = m_k^{\mathrm{d}}\,,\ \text{for } k\in\{2,\dots,N-1\},\\ &\sum_{k=1}^{N} x_{jk} = M_j^{\mathrm{s}}, \quad \text{for } j\in\{1,N\}, \quad \sum_{j=1}^{N} x_{jk} = M_k^{\mathrm{d}}, \quad \text{for } k\in\{1,N\},\\ &M_1^{\mathrm{s}}, M_N^{\mathrm{s}}, M_1^{\mathrm{d}}, M_N^{\mathrm{d}} \geq 0, \quad c_{11} = c_{1N} = c_{N1} = c_{NN} = +\infty\Bigg\}^{\frac{1}{p}},\end{aligned} \tag{4.2}$$

which corresponds to Eq (2.10) with

$$\begin{aligned}
\mathbf{x} &:= (x_{11}, x_{12}, \dots, x_{1N}, x_{21}, x_{22}, \dots, x_{2N}, \dots, x_{N1}, \dots, x_{NN}, M_1^{\mathrm{s}}, M_N^{\mathrm{s}}, M_1^{\mathrm{D}}, M_N^{\mathrm{D}})^\intercal \\
\mathbf{c} &:= (c_{11}, c_{12}, \dots, c_{1N}, c_{21}, c_{22}, \dots, c_{2N}, \dots, c_{N1}, \dots, c_{NN}, c_1^{\mathrm{s}}, c_N^{\mathrm{s}}, c_1^{\mathrm{D}}, c_N^{\mathrm{D}})^\intercal \\
\mathbf{b} &:= (0, m_2^{\mathrm{s}}, \dots, m_{N-1}^{\mathrm{s}}, 0, 0, m_2^{\mathrm{D}}, \dots, m_{N-1}^{\mathrm{D}}, 0)^\intercal
\end{aligned}$$

and $\mathbf{A}$ is a $2N \times (N^2 + 4)$ block matrix, obtained from the matrix Eq (2.12) by extending it with four columns. The extra columns have a single nonzero element each, namely

$$\mathbf{A}_{1,N^2+1} = -1, \quad \mathbf{A}_{N,N^2+2} = -1, \quad \mathbf{A}_{N+1,N^2+3} = -1, \quad \mathbf{A}_{2N,N^2+4} = -1, \tag{4.3}$$

due to the four extra unknowns and constraints of the problem Eq (4.2). The last constraints of Eq (4.2) are needed to assure that it is never convenient creating mass at the boundaries just to keep it there or to transport it to the other boundary.

In Appendix A.1, we present two academic examples which serve both to illustrate the features of the Figalli & Gigli approach and to facilitate a deeper understanding of the proposed methodology.

4.2. Piccoli & Rossi approach

4.2.1. Discrete approach

In the case of two purely atomic marginals with finite supports

$$\mu^{\mathrm{s}} = \sum_{i=1}^{N} m_i^{\mathrm{s}} \delta_{x_i}, \qquad \mu^{\mathrm{D}} = \sum_{i=1}^{N} m_i^{\mathrm{D}} \delta_{x_i},$$

the distance Eq (3.5) is computed by minimizing

$$a \sum_{i=1}^{N} (m_i^{\mathrm{s}} - \mathring{m}_i^{\mathrm{s}}) + a \sum_{i=1}^{N} (m_i^{\mathrm{D}} - \mathring{m}_i^{\mathrm{D}}) + b \sum_{j,k=1}^{N} c_{jk} \gamma_{jk},$$

where $c_{jk} = \|x_j - x_k\|^p$, among all $\gamma \in [0, +\infty)^{N \times N}$ such that

$$\begin{aligned}
\sum_{k=1}^{N} \gamma_{jk} &= \mathring{m}_j^{\mathrm{s}}, \quad \text{for } j \in \{1, \dots, N\}, \\
\sum_{j=1}^{N} \gamma_{jk} &= \mathring{m}_k^{\mathrm{D}}, \quad \text{for } k \in \{1, \dots, N\},
\end{aligned}$$

and among all $\mathring{m}_1^{\mathrm{s}}, \dots, \mathring{m}_N^{\mathrm{s}}, \mathring{m}_1^{\mathrm{D}}, \dots, \mathring{m}_N^{\mathrm{D}}$ with $\mathring{m}_i^{\mathrm{s}} \le m_i^{\mathrm{s}}$ and $\mathring{m}_i^{\mathrm{D}} \le m_i^{\mathrm{D}}$, $i \in \{1, \dots, N\}$. By taking such conditions into account, the objective is to minimize

$$a \sum_{i=1}^{N} (m_i^{\mathrm{s}} + m_i^{\mathrm{D}}) + \sum_{j,k=1}^{N} (b\, c_{jk} - 2a) \gamma_{jk},$$

among all $\gamma \in [0, +\infty)^{N\times N}$ such that

$$\sum_{k=1}^{N} \gamma_{jk} \le m_j^{\mathrm{S}}, \qquad \text{for all } j = 1, \dots, N,$$
$$\sum_{j=1}^{N} \gamma_{jk} \le m_k^{\mathrm{D}}, \qquad \text{for all } k = 1, \dots, N.$$

4.2.2. Computational approach

This time, the problem can be solved through a linear programming problem with inequality constraints. First, one solves the problem

$$\begin{aligned} &\text{minimize} && \hat{\mathbf{c}}^{\intercal}\mathbf{x} \\ &\text{subject to} && \mathbf{A}\mathbf{x} \le \mathbf{b} \\ & && \mathbf{x} \ge 0, \end{aligned} \tag{4.4}$$

where $\hat{\mathbf{c}} := b\mathbf{c} - 2a\mathbb{1}_{N^2}$, $\mathbf{x}$, $\mathbf{c}$, $\mathbf{b}$ are defined as in Eq (2.11), and $\mathbf{A}$ is defined as in Eq (2.12). Once the solution $\mathbf{x}^* = \mathbf{x}^*(a, b)$ of Eq (4.4) is found, the final GWD is given by

$$\tilde{W}_1^{\mathrm{PR}}(m^{\mathrm{S}}, m^{\mathrm{D}}; a, b) := \mathbf{x}^*(a, b) + a \sum_{i=1}^{N} (m_i^{\mathrm{S}} + m_i^{\mathrm{D}}).$$

In Appendix A.2, we present two academic examples which serve both to illustrate the features of the Piccoli & Rossi approach and to facilitate a deeper understanding of the proposed methodology.

4.3. Gaussian Hellinger–Kantorovich approach

4.3.1. Discrete approach

When computing the distance Eq (3.7) of two purely atomic measures with finite supports

$$\mu^{\mathrm{S}} = \sum_{i=1}^{N} m_i^{\mathrm{S}} \delta_{x_i}, \qquad \mu^{\mathrm{D}} = \sum_{i=1}^{N} m_i^{\mathrm{D}} \delta_{x_i},$$

we may resort to a dual formulation. Namely, by adapting the proof of [9, Theorem 6.3] to the case in which a, b are not necessarily both equal to 1, one sees that the distance equals the maximum of

$$a \sum_{i=1}^{N} m_i^{\mathrm{S}} \delta_{x_i} \left(1 - e^{-b\varphi_i/a}\right) + a \sum_{i=1}^{N} m_i^{\mathrm{D}} \delta_{x_i} \left(1 - e^{-b\psi_i/a}\right)$$

among all $(\varphi_1, \dots, \varphi_N, \psi_1, \dots, \psi_N) \in \mathbb{R}^N \times \mathbb{R}^N$ such that

$$\varphi_j + \psi_k \le c_{jk} := |x_j - x_k|^2, \qquad \text{for all } j, k \in \{1, \dots, N\}. \tag{4.5}$$

4.3.2. Computational approach

In this case, the optimization problem can be written as a nonlinear maximization problem with linear constraints. To this end, we define the function $f : \mathbb{R}^{2N} \to \mathbb{R}$

$$f(\mathbf{z}) := a \sum_{i=1}^{N} m_i^{\mathrm{s}} \left(1 - e^{-bz_i/a}\right) + a \sum_{i=N+1}^{2N} m_{i-N}^{\mathrm{D}} \left(1 - e^{-bz_i/a}\right), \qquad \mathbf{z} \in \mathbb{R}^{2N}.$$

We then aim to find

$$\tilde{W}_2^{\mathrm{GHK}}(m^{\mathrm{s}}, m^{\mathrm{D}}) := \max_{\mathbf{z}} f(\mathbf{z})$$

under the constraints

$$\mathbf{A}^{\intercal}\mathbf{z} \leq \mathbf{c},$$

where $\mathbf{c}$ is defined as in Eq (2.11) (with c_{jk} as in Eq (4.5)) and $\mathbf{A}$ is the $N^2 \times 2N$ matrix defined as

$$\mathbf{A} := \begin{bmatrix} \mathbf{e}_1\, \mathbf{e}_1 \cdots \mathbf{e}_1 & \mathbf{e}_2\, \mathbf{e}_2 \cdots \mathbf{e}_2 & \cdots & \mathbf{e}_N\, \mathbf{e}_N \cdots \mathbf{e}_N \\ \mathbf{e}_1\, \mathbf{e}_2 \cdots \mathbf{e}_N & \mathbf{e}_1\, \mathbf{e}_2 \cdots \mathbf{e}_N & \cdots & \mathbf{e}_1\ \mathbf{e}_2\ \cdots \mathbf{e}_N \end{bmatrix},$$

with $\mathbf{e}_i = (0\,0\ \cdots \underbrace{1}_{i^{th}} \cdots\ 0\,0)^{\intercal} \in \mathbb{R}^N$ being, as usual, the ith element of the canonical base of $\mathbb{R}^N$.

Numerical tests for the GWD discussed here are in common with those of the following GWD; see Appendix A.3.

4.4. Savaré & Sodini approach

4.4.1. Discrete approach

For the purpose of empiric approximation, it may be convenient to resort to the convex relaxation approach introduced in [11]. In particular, in view of [11, Theorem 3.5], in the case of two atomic homogeneous marginals

$$\mu^{\mathrm{s}} = \sum_{i=1}^{N} m_i^{\mathrm{s}} \delta_{x_i}, \qquad \mu^{\mathrm{D}} = \sum_{i=1}^{N} m_i^{\mathrm{D}} \delta_{x_i},$$

the Gaussian Hellinger–Kantorovich distance can be obtained by minimizing

$$\mathcal{S} := \sum_{j,h,k,\ell} \gamma_{(jh,k\ell)} H_g(x_j, r_{jh}^{\mathrm{s}}; x_k, r_{k\ell}^{\mathrm{D}}) \tag{4.6}$$

under the constraints $\gamma_{(jh,k\ell)} \geq 0$, $\sum_{jhk\ell} \gamma_{(jh,k\ell)} = 1$, and for the homogeneous marginals

$$m_j^{\mathrm{s}} = \sum_{h} r_{jh}^{\mathrm{s}} \sum_{k\ell} \gamma_{(jh,k\ell)}, \qquad m_k^{\mathrm{D}} = \sum_{\ell} r_{k\ell}^{\mathrm{D}} \sum_{jh} \gamma_{(jh,k\ell)},$$

and

$$\sum_{h} r_{jh}^{\mathrm{s}} = m_j^{\mathrm{s}}, \qquad \sum_{\ell} r_{k\ell}^{\mathrm{D}} = m_k^{\mathrm{D}},$$

for all $j, k \in \{1, \ldots, N\}$ (see also Figure 4). We point out that this latter approach differs from that of the previous one only numerically because the distance defined in this way in principle coincides with the one discussed in Section 4.3.1.

4.4.2. Computational approach

Starting again from the basic numerical setting introduced in Section 2.3, we introduce the additional, 'vertical', discretization at any fixed grid point x_i and for each mass distribution; see Figure 4.

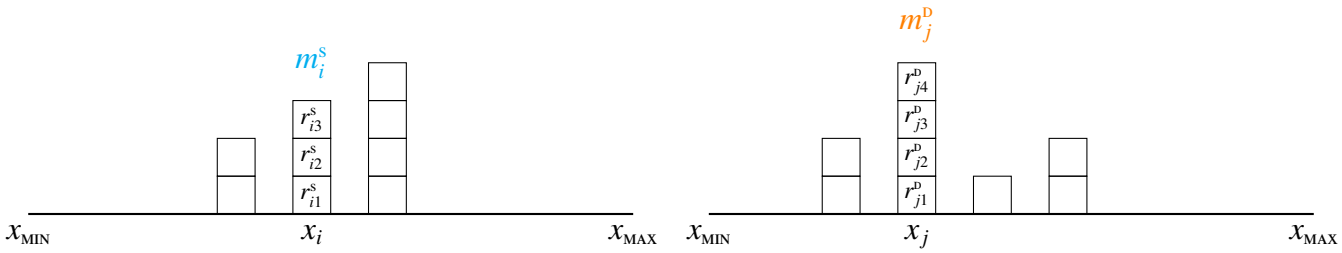

Figure 4. Additional discretization of the concentrated masses.

At each grid point x_i, the mass m_i^{s} is divided into $n_i^{\mathrm{s}} \geq 0$ intervals (reasonably, $n_i^{\mathrm{s}} = 0$ if and only if $m_i^{\mathrm{s}} = 0$). Analogously, the mass m_i^{D} is divided into $n_i^{\mathrm{D}} \geq 0$ intervals (again, $m_i^{\mathrm{D}} = 0$ if and only if $m_i^{\mathrm{D}} = 0$). Note that $n_i^{\{\mathrm{s,D}\}}$ acts as additional discretization parameter, and we expect the numerical error to vanish only for both $n_i^{\{\mathrm{s,D}\}} \to +\infty$ and $N \to +\infty$.

Now, each subdivision of the supply mass is associated with the additional variable r_{jh}^{s}, $j = 1, \dots, N$, $h = 1, \dots, n_j^{\mathrm{s}}$, and each subdivision of the demand mass is associated with the additional variable $r_{k\ell}^{\mathrm{D}}$, $k = 1, \dots, N$, $\ell = 1, \dots, n_k^{\mathrm{D}}$.

We also define

$$\Gamma \in \mathbb{R}^d, \quad \text{with} \quad d = \sum_{i=1}^{N} n_i^{\mathrm{s}} \cdot \sum_{i=1}^{N} n_i^{\mathrm{D}},$$

as the vector of all the weights $\gamma_{(jh,k\ell)}$ associated with the pairs $(r_{jh}^{\mathrm{s}}, r_{k\ell}^{\mathrm{D}})$, $j = 1, \dots, N$, $h = 1, \dots, n_j^{\mathrm{s}}$, $k = 1, \dots, N$, $\ell = 1, \dots, n_k^{\mathrm{D}}$.

The optimization problem to be solved reads as

$$\begin{aligned}
&\text{minimize} \quad \mathcal{S}(\Gamma, r^{\mathrm{s}}, r^{\mathrm{D}}) \\
&\text{subject to} \quad \gamma_{(jh,k\ell)} \geq 0, \quad \forall\, j,h,k,\ell, \qquad \sum_{jhk\ell} \gamma_{(jh,k\ell)} = 1, \\
&\qquad\qquad m_j^{\mathrm{s}} = \sum_{h=1}^{n_j^{\mathrm{s}}} \Big(r_{jh}^{\mathrm{s}} \sum_{k\ell} \gamma_{(jh,k\ell)} \Big), \quad \forall\, j, \\
&\qquad\qquad m_k^{\mathrm{D}} = \sum_{\ell=1}^{n_k^{\mathrm{D}}} \Big(r_{k\ell}^{\mathrm{D}} \sum_{jh} \gamma_{(jh,k\ell)} \Big), \quad \forall k,
\end{aligned} \tag{4.7}$$

where $\mathcal{S}$ is defined in Eq (4.6), and the sums $\sum_{jhk\ell}$, $\sum_{k\ell}$, $\sum_{jh}$ run on all over the index combinations. It is important to note that the linear systems for r^{s}, r^{D} appearing in Eq (4.7) can be actually underdetermined for a given Γ. Analogously to the previous cases, we will denote the solution of the optimization problem by $\tilde{W}_2^{\mathrm{ss}}$.

In order to solve the optimization problem Eq (4.7), we have devised two strategies: The first one is based on a slow but exhaustive search, which guarantees finding the global minimum. The second one is based on a nondifferential descend method and allows for faster computations, although a priori limited to the search of a local minimum.

Algorithm SS-A: Exhaustive search

1. Introduce a grid with Q nodes in the unit interval $[0,1]$, namely $0 = q_1 < q_2 < \ldots, q_Q = 1$. Note that Q is yet another approximation parameter (the associated error is expected to vanish for $Q \to +\infty$).
2. List all the possible vectors $\{\tilde{\Gamma} \in \mathbb{R}^d\}$ such that each element belongs to the set $\{q_1, \ldots, q_Q\}$ and the sum of the elements equals 1. A good algorithm to do this is the following (with $d = 5$), although some generated $\tilde{\Gamma}$ lead to linear systems with no solution (we have discarded them).

```
for i_1 = 0, ..., Q do
    for i_2 = 0, ..., Q − i_1 do
        for i_3 = 0, ..., Q − i_1 − i_2 do
            for i_4 = 0, ..., Q − i_1 − i_2 − i_3 do
                i_5 = Q − i_1 − i_2 − i_3 − i_4
                Γ̃ = (i_1, i_2, i_3, i_4, i_5)/Q
            end for
        end for
    end for
end for
```

3. For each $\tilde{\Gamma}$, find $\{r^{\mathrm{s}}_{jh}\}_{jh}$ and $\{r^{\mathrm{D}}_{k\ell}\}_{k\ell}$ (depending on that $\tilde{\Gamma}$) solving the underdetermined linear systems, selecting the solution with the minimal norm.
4. For each $\tilde{\Gamma}$ (and the corresponding $\{r^{\mathrm{s}}_{jh}\}_{jh}$, $\{r^{\mathrm{D}}_{k\ell}\}_{k\ell}$), evaluate $\mathcal{S}(\tilde{\Gamma})$.
5. Find $\min_{\tilde{\Gamma}} \mathcal{S}(\tilde{\Gamma})$ by exhaustive comparison.

Algorithm SS-B: Random guess and random descend

1. List the largest possible number of vectors $\tilde{\Gamma} \in \mathbb{R}^d$ such that each element belongs to the interval $[0,1]$ and the sum of the elements equals 1. A simple algorithm to do this is the following.

```
for z = 1, ..., d do
    Extract a uniform random variable ω_z ∈ {1, ..., Q}
end for
Γ̃ = (ω_1, ..., ω_d)/ Σ_z ω_z
```

2. For each $\tilde{\Gamma}$, find $\{r^{\mathrm{s}}_{jh}\}_{jh}$ and $\{r^{\mathrm{D}}_{k\ell}\}_{k\ell}$ (depending on that $\tilde{\Gamma}$) solving the underdetermined linear systems, selecting the solution with the minimal norm.
3. For each $\tilde{\Gamma}$ (and the corresponding $\{r^{\mathrm{s}}_{jh}\}_{jh}$, $\{r^{\mathrm{D}}_{k\ell}\}_{k\ell}$), evaluate $\mathcal{S}(\tilde{\Gamma})$.
4. Find $\tilde{\Gamma}^* := \arg\min_{\tilde{\Gamma}} \mathcal{S}(\tilde{\Gamma})$ by exhaustive comparison.

5. Starting from $\tilde{\Gamma}^*$ as an initial guess, explore the control space perturbing the best Γ obtained so far with a small perturbation $\pm\varepsilon$ on random pairs of its elements (thus keeping the sum equal to 1). If the perturbation leads to a lower value, the best Γ^* is updated with the perturbed one; otherwise, the attempt is discarded. As usual, duly diminishing ε while the algorithm runs can improve the accuracy (simulated annealing).

In Appendix A.3, we present some academic examples to illustrate and better understand the two approaches described in Sects. 4.3.2 and 4.4.2.

4.5. Comparison of the three GWDs

In this numerical test, we compare the four computational approaches for the three GWDs individually studied above. This is done by devising a numerical test which can be run by all algorithms. From now on, the four algorithms will be referred to as FG, PR, SS (variant B only), and GHK, with obvious meanings.

Denoting by χ the indicator function, we consider the case of two time-dependent distributions with parameters $\alpha, \beta, \eta > 0$

$$\rho^{\mathrm{s}}(x,t) = \alpha\chi_{[-t-\eta,-t+\eta]}(x), \qquad \rho^{\mathrm{D}}(x,t) = \beta\chi_{[t-\eta,t+\eta]}(x), \qquad 0 \le t \le T \tag{4.8}$$

in the domain $X = [-T-\eta, T+\eta]$. The two step functions start with perfectly overlapping support, then the supply mass moves leftward while the demand mass moves rightward until they both reach the boundary of the domain; see Figure 5.

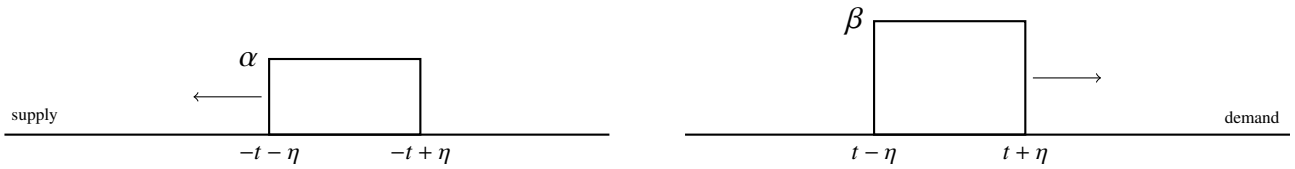

Figure 5. Moving distributions ρ^{s} and ρ^{D} as defined in Eq (4.8).

In particular, for this comparison test, we set $\alpha \in \{1, 2\}$, $\beta = 2$, $\eta = 1$, $T = 4$. Moreover, we have the following algorithm-specific parameters:

- PR: $a = 2.5$, $b = 1$.
- GHK: $a = b = 1$.
- SS: $a = b = 1$, 4 atomic masses per distribution, 4 additional divisions per mass, $Q = 1000$, $\epsilon = 0.001$, iterations 5000 (random searches) + 40000 (random descend), average on 10 runs.

Figure 6 shows the behavior of the three GWDs between ρ^{s} and ρ^{D} as defined in Eq (4.8), as a function of time. The WD is also shown in the balanced case $\alpha = \beta$.

This comparison test allows us to sketch some preliminary conclusions. First of all, SS seems to be computationally unfeasible. It is impossible to deal with a reasonable number of concentrated masses and the additional divisions; therefore, it will be discarded in the following analysis.

Among the others, PR and GHK share an important drawback with L^p: the distances saturate when the distributions do not overlap and they are far enough from each other. This does not seem to be a nice feature in the context of traffic flow models.

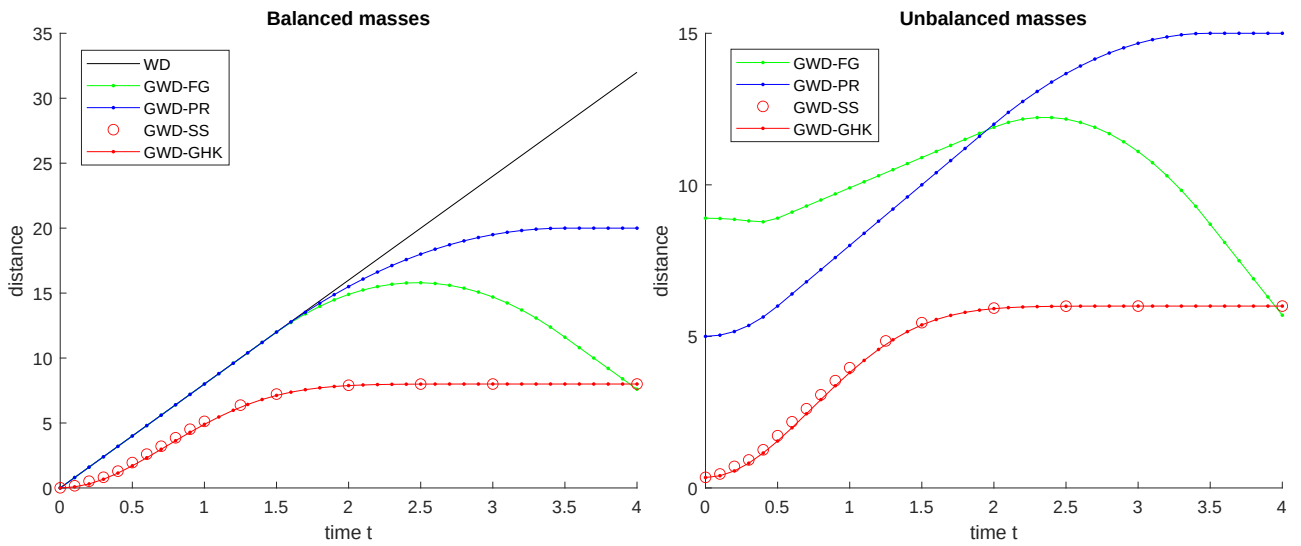

Figure 6. Comparison of the three GWDs between ρ^{s} and ρ^{p} as defined in Eq (4.8), as a function of time t ($\beta = 2$, $\eta = 1$, $T = 4$). $\alpha = 2$ (balanced), including also the classical WD (left) and $\alpha = 1$ (unbalanced, right).

Lastly, FG has an important feature which deserves discussion: The distance decreases whenever the distributions are close to the boundaries, even if they are very far from each other. On first glance, this seems to be an issue, but it can be also be seen as a good point. Basically, it implies that what happens near the boundaries is less important than what happens on the rest of the road. Considering the fact that the boundary conditions are typically unknown in traffic flow modeling, it could make sense to give priority to that part of the road which is less or not at all influenced by the boundary conditions.

5. Applications to traffic model analysis

In this section, we will try to understand to what extent the GWDs previously studied are suitable in the context of traffic flow models. It is useful to recall here that in [2], the authors have already extensively studied the sensitivity of the LWR model [25, 26] using the WD. More precisely, many simulations were performed by varying some key parameters like the initial vehicle density, the road capacity (maximal flux), the critical density (vehicle density associated with the maximal flux), the size of the road network, and the distribution coefficients at junctions (how traffic is distributed across a junction with more than one outgoing road). Interestingly, the authors found that the parameter the model is most sensitive to is the last one, namely the distribution coefficients at junctions. That said, the usage of the WD has limited the authors to the comparison of *balanced* vehicle distributions only. This means, in practice, that the number of vehicles should be the same in the two scenarios considered.

In the following, we try to complete the sensitivity analysis, including the cases in which the mass distributions are unbalanced.

5.1. Mathematical modeling

In a macroscopic setting, the traffic flow is described by means of average quantities only, like the *density* ρ (vehicles per unit of space, at a given time), the *flux* f (vehicles per unit of time crossing

a given point, at a given time), and the *velocity* v. Models can be *first-order* (i.e., velocity-based), if the velocity is given as a function of the density and acceleration is neglected, or *second-order* (i.e., acceleration-based) if, instead, velocity is an independent variable and acceleration is given as a function of both density and velocity.

In the following, we will assume that the road is a finite segment $X = [x_{\text{MIN}}, x_{\text{MAX}}]$ with a single lane and a single class of vehicles.

The LWR model

The LWR model, introduced independently by Lighthill and Whitham [25] and by Richards [26], describes the evolution of the vehicle density $\rho(x,t)$ by means of the following initial-boundary value problem involving an hyberbolic PDE

$$\begin{cases} \partial_t \rho + \partial_x f(\rho) = 0, & t \in [t_0, T], \quad x \in [x_{\text{MIN}}, x_{\text{MAX}}], \\ \rho(x,0) = \rho_0(x), & x \in [x_{\text{MIN}}, x_{\text{MAX}}], \\ \rho(x_{\text{MIN}}, t) = \rho_{\text{IN}}(t), & t \in [t_0, T], \\ \rho(x_{\text{MAX}}, t) = \rho_{\text{OUT}}(t), & t \in [t_0, T], \end{cases} \tag{5.1}$$

for some initial time t_0, a final time $T > t_0$, and a concave flux function $f : [0, \rho_{\text{MAX}}] \to \mathbb{R}^+$ (called the *fundamental diagram* in the context of traffic flow) such that $f(0) = f(\rho_{\text{MAX}}) = 0$, for some maximal density ρ_{MAX}. Note that we have naturally guaranteed the constraint $0 \le \rho \le \rho_{\text{MAX}}$ at any time, provided that $0 \le \rho_0, \rho_{\text{IN}}, \rho_{\text{OUT}} \le \rho_{\text{MAX}}$.

Recalling the well-known physical law $f = \rho v$, which is always valid for any x and t, we find that the LWR model implicitly assumes that the vehicle velocity v only depends on the density ρ (since $f = f(\rho)$, therefore $v = v(\rho)$). From the modeling point of view, this means that the velocity adapts instantaneously to the changes in the density, i.e., acceleration is infinite. For this reason, the LWR model is said to be *first-order*.

As usual in the mathematical literature, in the following Tests 1–3 we will define the following:

$$f(\rho) = \rho(1 - \rho), \tag{5.2}$$

implicitly assuming that the maximal density and the maximal velocity of vehicles are normalized and equal to 1. This is done for simplicity, since a qualitative analysis is sufficient to highlight the differences among the GWDs. In Test 4, instead, we consider realistic parameters.

Equation (5.1) can be numerically approximated by means of, e.g., the Godunov scheme. Using again the numerical setting introduced in Section 2.3, this scheme reads as

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta x}{\Delta t}\Big(\mathcal{G}(\rho_i^n, \rho_{i+1}^n) - \mathcal{G}(\rho_{i-1}^n, \rho_i^n)\Big), \qquad i = 2, \dots, N-1, \quad n = 0, \dots, \left[\frac{T - t_0}{\Delta t}\right], \tag{5.3}$$

where ρ_i^n represents the approximated value of ρ at the center of the cell C_i at time $t_0 + n\Delta t$, and $\mathcal{G}$ is Godunov's numerical flux defined by

$$\mathcal{G}(\rho_-, \rho_+) := \begin{cases} \min\limits_{z \in [\rho_-, \rho_+]} f(z), & \text{if } \rho_- \le \rho_+, \\ \max\limits_{z \in [\rho_+, \rho_-]} f(z), & \text{if } \rho_- \ge \rho_+. \end{cases} \tag{5.4}$$

Remark 6. *Boundary conditions can be also given in form of incoming and outgoing flux, rather than as left and right densities ρ_{IN}, ρ_{OUT}. This happens, for example, if the boundary data are given by fixed sensors counting the vehicles passing through $x = x_{\text{MIN}}$ and $x = x_{\text{MAX}}$ in a time interval Δt. If this is the case, the values $\rho_1^n = \rho_{\text{IN}}^n$ and $\rho_N^n = \rho_{\text{OUT}}^n$ are dropped, and the incoming/outgoing fluxes F_{IN}^n, F_{OUT}^n are directly injected into the scheme Eq* (5.3) *taking the place of $\mathcal{G}(\rho_{i-1}^n, \rho_i^n)$ if $i = 1$, and of $\mathcal{G}(\rho_i^n, \rho_{i+1}^n)$ if $i = N$, respectively.*

The ARZ model

The ARZ model, introduced independently by Aw and Rascle [27] and by Zhang [28], describes the evolution of the vehicle density $\rho(x,t)$ and the velocity $v(x,t)$ by means of the following initial-boundary value problem involving two hyberbolic PDEs:

$$\begin{cases} \partial_t \rho + \partial_x(\rho v) = 0, & t \in [t_0, T], \quad x \in [x_{\text{MIN}}, x_{\text{MAX}}], \\ \partial_t(\rho w) + \partial_x(v\rho w) = \rho \frac{v_{eq}(\rho) - v}{\tau}, & t \in [t_0, T], \quad x \in [x_{\text{MIN}}, x_{\text{MAX}}], \\ \rho(x, 0) = \rho_0(x), & x \in [x_{\text{MIN}}, x_{\text{MAX}}], \\ v(x, 0) = v_0(x), & x \in [x_{\text{MIN}}, x_{\text{MAX}}], \\ \rho(x_{\text{MIN}}, t) = \rho_{\text{IN}}(t), & t \in [t_0, T], \\ \rho(x_{\text{MAX}}, t) = \rho_{\text{OUT}}(t), & t \in [t_0, T], \\ v(x_{\text{MIN}}, t) = v_{\text{IN}}(t), & t \in [t_0, T], \\ v(x_{\text{MAX}}, t) = v_{\text{OUT}}(t), & t \in [t_0, T], \end{cases} \tag{5.5}$$

where $w(x,t) := v(x,t) + P(\rho(x,t))$, $P(\rho) := \frac{v_{ref}}{\gamma}\left(\frac{\rho}{\rho_{\text{MAX}}}\right)^{\gamma}$, and v_{ref}, γ, and τ are strictly positive parameters.

This time, the velocity v is an independent unknown and the acceleration $\frac{v_{eq}(\rho)-v}{\tau}$ appears explicitly in the model, depending on ρ and v. For this reason, the ARZ model is said to be *second-order*.

In order to fairly compare the two models, we define $v_{eq}(\rho) := \frac{f(\rho)}{\rho}$, such that the desired equilibrium velocity of the ARZ model coincides with the instantaneous velocity of the LWR model. In the boundary conditions, we also set $v_{\text{IN}} = v_{eq}(\rho_{\text{IN}})$ and $v_{\text{OUT}} = v_{eq}(\rho_{\text{OUT}})$.

Equation (5.5) is numerically approximated by means of the Lax–Friedrichs scheme.

5.2. Numerical tests

5.2.1. Test 1: Sensitivity to boundary conditions

In this test, we investigate the sensitivity to the boundary conditions. We consider the case described in Remark 6, which is a scenario often met in real life [29]. In the same spirit as [29], we assume that the time interval is divided into two subintervals: $[t_0, 0]$ and $(0, T]$, where $t = 0$ corresponds to the current time (the 'now'), i.e., the time the simulation is actually performed. Until time 0, we can rely on the sensors' measurements F_{IN}, F_{OUT}, and we use it as boundary conditions (see Remark 6). After that time, we have no such data because they are not yet available, but we assume that we are be able to (exactly) forecast their average value,

$$\hat{F}_{\text{IN}} = \frac{1}{T}\int_0^T F_{\text{IN}}(t)dt, \qquad \hat{F}_{\text{OUT}} = \frac{1}{T}\int_0^T F_{\text{OUT}}(t)dt.$$

In short, we denote by ρ^{S} the 'exact' solution obtained by using the boundary fluxes $F_{\text{IN}}(t)$, $F_{\text{OUT}}(t)$ in the whole interval $[t_0, T]$, and by ρ^{D} the 'predicted' solution obtained by using the flux boundary data

$F_{\text{IN}}(t)$, $F_{\text{OUT}}(t)$ until time $t = 0$ and then by using the predicted average values $\hat{F}_{\text{IN}}$, $\hat{F}_{\text{OUT}}$ until time T. At any time t, the vehicle distributions $\rho^{\text{s}}(\cdot, t)$ and $\rho^{\text{D}}(\cdot, t)$ are compared computing the GWDs.

The initial conditions are $\rho^{\text{s}}(x, t_0) = \rho^{\text{D}}(x, t_0) = 0.2$ for any x. Boundary fluxes are randomly generated: F_{IN} has a uniform distribution in $[0, \frac{1}{4}]$, and F_{OUT} has a uniform distribution in $[0, \mathcal{G}(0.2, 0.2)]$. The other parameters are: $x_{\text{MIN}} = 0$, $x_{\text{MAX}} = 40$, $t_0 = -10$, $T = 30$, $\Delta x = 0.4$ (corresponding to $N = 100$), and $\Delta t = 0.25$.

Figure 7 shows two screenshots of the simulations at times $t = 0$ and $t = 12$, while Figure 8 shows the total mass on the road and the GWDs between $\rho^{\text{s}}(\cdot, t)$ and $\rho^{\text{D}}(\cdot, t)$ at any time.

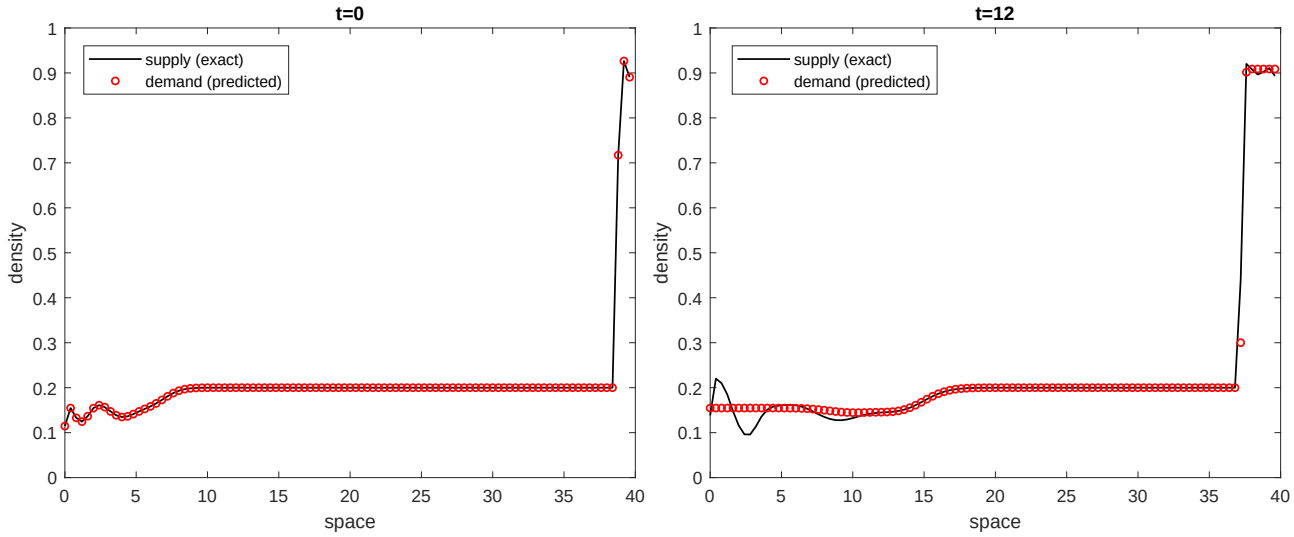

Figure 7. Test 1 on sensitivity to the boundary conditions. Two screenshots at $t = 0$ ('now'), i.e., when the solutions cease to be equal (left), and $t = 12$ (right).

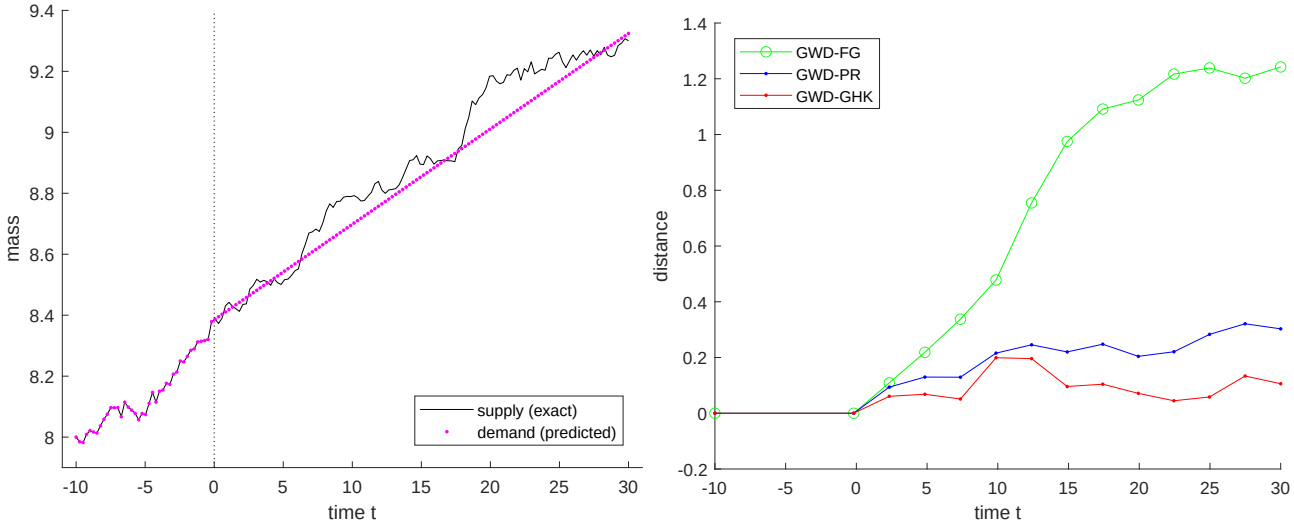

Figure 8. Test 1 on sensitivity to the boundary conditions. Difference in mass between the two simulations (left) and comparison of the GWDs (right).

Observing the numerical results, we tend to confirm what we anticipated in Section 4.5, namely that Figalli & Gigli approach seems to be the most suitable GWD to be used in the context of traffic

flow modeling, mainly because of how boundaries are regarded. In this test, the discrepancy between the exact and predicted solution slowly propagates from the boundaries to the interior of the road, impacting more and more on the reliability of traffic prediction. This behavior of the error is clearly better caught by FG (see Figure 8 (right)), which is the last distance that is saturated.

5.2.2. Test 2: Sensitivity to traffic light cycle

In this test, we consider a shorter road with respect to the previous test, and we assume that there is a traffic light in the middle of the road, at the interface between two numerical cells. The duration of the red phase equals that of the green phase. Technically, the red traffic light is modeled by imposing null flux on the right (resp., left) boundary of the cell before (resp., after) the traffic light. The aim is to investigate the sensitivity to the duration of the traffic light cycle.

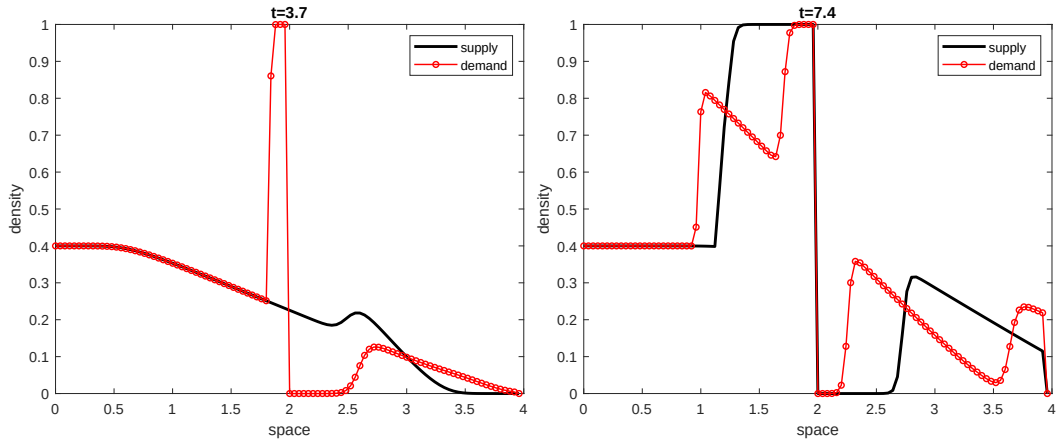

Figure 9. Test 2 on sensitivity to traffic light cycle. Two screenshots at $t = 3.7$ with red traffic light for demand and green traffic light for supply (left), and $t = 7.4$ with red traffic light for both distributions (right).

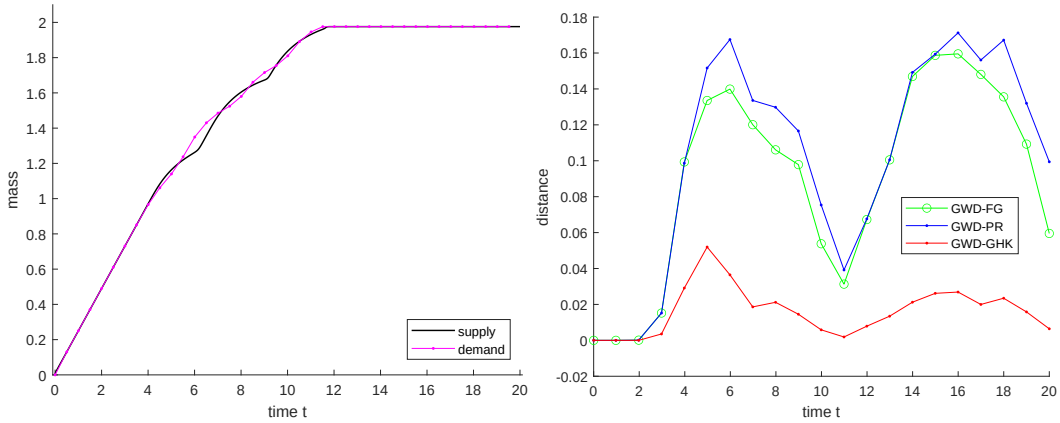

Figure 10. Test 2 on sensitivity to traffic light cycle. Difference in mass between the two simulations (correctly masses differ only in a certain time interval, left), and comparison of the GWDs (right).

The initial conditions are $\rho^s(x, t_0) = \rho^p(x, t_0) = 0$ for any x. The left Dirichlet boundary conditions

are $\rho^{\mathrm{s}}_{\mathrm{IN}} = \rho^{\mathrm{D}}_{\mathrm{IN}} = 0.4$. The right Dirichlet boundary conditions are $\rho^{\mathrm{s}}_{\mathrm{OUT}} = \rho^{\mathrm{D}}_{\mathrm{OUT}} = 0$. The other parameters are: $x_{\mathrm{MIN}} = 0$, $x_{\mathrm{MAX}} = 4$, $t_0 = 0$, $T = 20$, $\Delta x = 0.04$ (corresponding to $N = 100$), and $\Delta t = 0.025$. Both red and green phases last 50 time steps for supply distribution and 40 time steps for demand distribution.

Figure 9 shows two screenshots of the simulations at times $t = 3.7$ and $t = 7.4$.

Figure 10 shows the total mass on the road in the two scenarios and the GWDs between $\rho^{\mathrm{s}}(\cdot, t)$ and $\rho^{\mathrm{D}}(\cdot, t)$ at any time.

This test does not show much difference among the three GWDs, as they have basically the same cycling behavior. That behavior is correct, since the two solutions cyclically move away from each other and then come closer again.

5.2.3. Test 3: Sensitivity to model order

In this test, we quantify the difference between two simulations obtained by means of the LWR model (supply, first-order) and the ARZ model (demand, second-order).

The initial conditions are $\rho^{\mathrm{s}}(x, t_0) = \rho^{\mathrm{D}}(x, t_0) = 0.7$ for $x \in [1.6, 2.4]$ and 0.1 elsewhere. The left and right Dirichlet boundary conditions are null. The other parameters are: $\tau = 0.05$, $\gamma = 2$, $v_{ref} = 1$, $x_{\mathrm{MIN}} = 0$, $x_{\mathrm{MAX}} = 4$, $t_0 = 0$, $T = 3$, $\Delta x = 0.02$ (corresponding to $N = 200$), and $\Delta t = 0.01$.

Figure 11 shows two screenshots of the simulations at times $t = 0.3$ and $t = 1$.

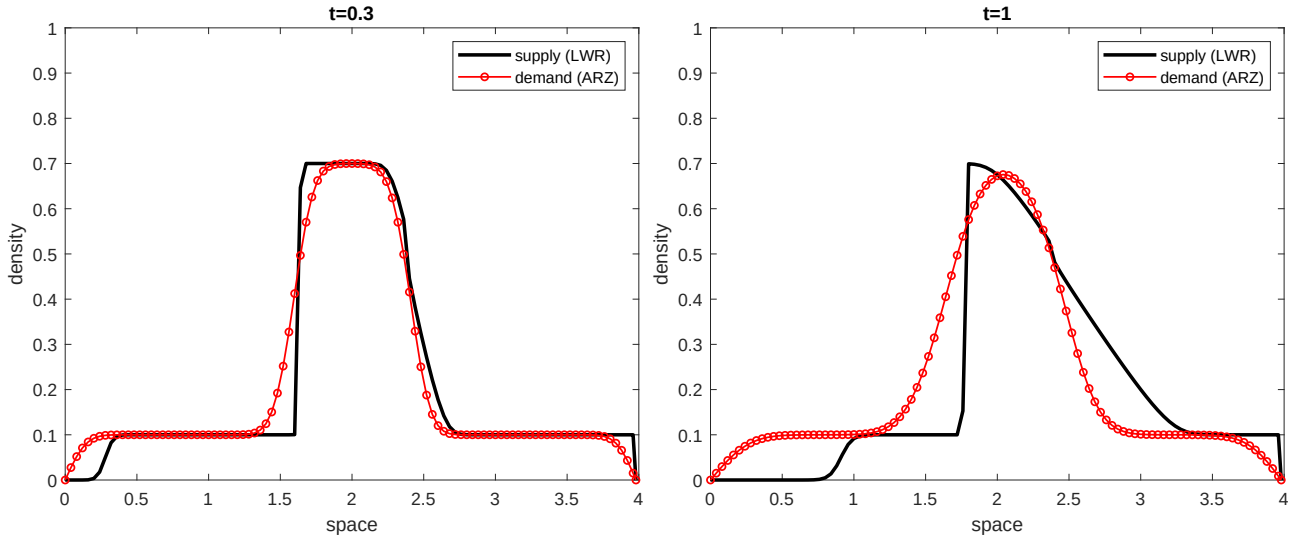

Figure 11. Test 3 on sensitivity to model order. Two screenshots at $t = 0.3$ (left), and $t = 1$ (right).

Figure 12 shows the total mass on the road in the two scenarios and the GWDs between $\rho^{\mathrm{s}}(\cdot, t)$ and $\rho^{\mathrm{D}}(\cdot, t)$ at any time. Here, we have chosen $a = 0.5$ and $b = 1$ as parameters for computing $\tilde{W}_1^{\mathrm{PR}}$.

This test is similar to Test 1, even if the conclusions are less evident: Here the two solutions start differing immediately, both at the boundary and inside the road. Moreover, intuition tells us that the difference increases as the time goes by. Among the three GWDs, FG is the one which increases more rapidly at the initial time, and it is the one with the largest relative increment at later times.

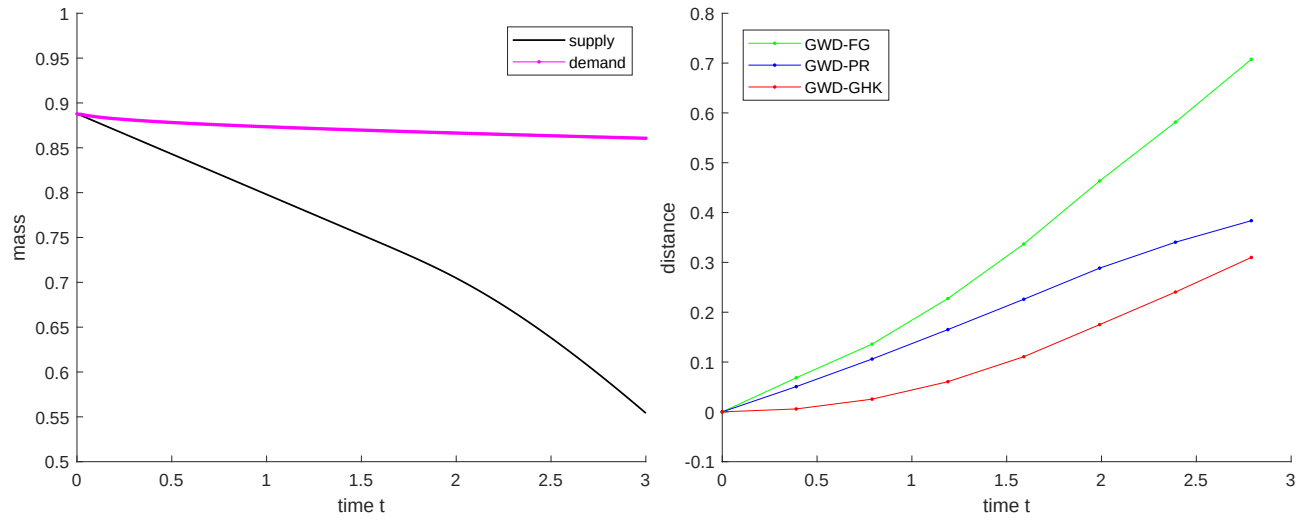

Figure 12. Test 3 on sensitivity to model order. Difference in mass between the two simulations (left) and comparison of the GWDs (right).

5.2.4. Test 4: Calibration and validation

In this section, we consider two important problems in traffic modeling, namely *calibration* and *validation*. In fact, once a new mathematical model is devised, the calibration is needed for finding the best values of its parameters to be used in order to fit some reference solution as much as possible, typically coming from real observations (real traffic data). Once the calibration is done, one has to *validate* the model, checking if it is also able to compute accurate solutions in scenarios which are different from those used for calibration.

In both steps, the choice of a distance function is crucial, since it is used to measure the difference between the exact and the simulated traffic conditions. Different distances can highlight different features, thus affecting the values of the optimal parameters.

To study these aspects of traffic modeling, we resort again to the LWR model (Section 5.1). Here, we do not normalize the maximal density ρ_{MAX} and maximal velocity v_{MAX} to 1, so we have

$$f(\rho) = \rho\left(-\frac{v_{\text{MAX}}}{\rho_{\text{MAX}}}\rho + v_{\text{MAX}}\right)$$

instead of Eq (5.2). Summarizing, to run the model, we need the two parameters ρ_{MAX} and v_{MAX}, the initial condition ρ_0, and the boundary conditions $\rho_{\text{IN}}, \rho_{\text{OUT}}$. The proposed numerical tests are performed in two steps:

- *Direct problem*: We fix the parameters/conditions and we solve Eq (5.1), saving the final solution $\rho_T^{\text{REF}}(\cdot) := \rho(\cdot, T)$ as reference solution;
- *Inverse problem*: We assume that some parameters/conditions are unknown, and we use the other parameters/conditions and the reference solution ρ_T^{REF} to find the unknown parameters/conditions by means of an optimization problem.

Note that the second step requires us to measure the distance between the reference solution and the solution obtained with tentative parameters/conditions several times, until the minimum distance (best match) is found.

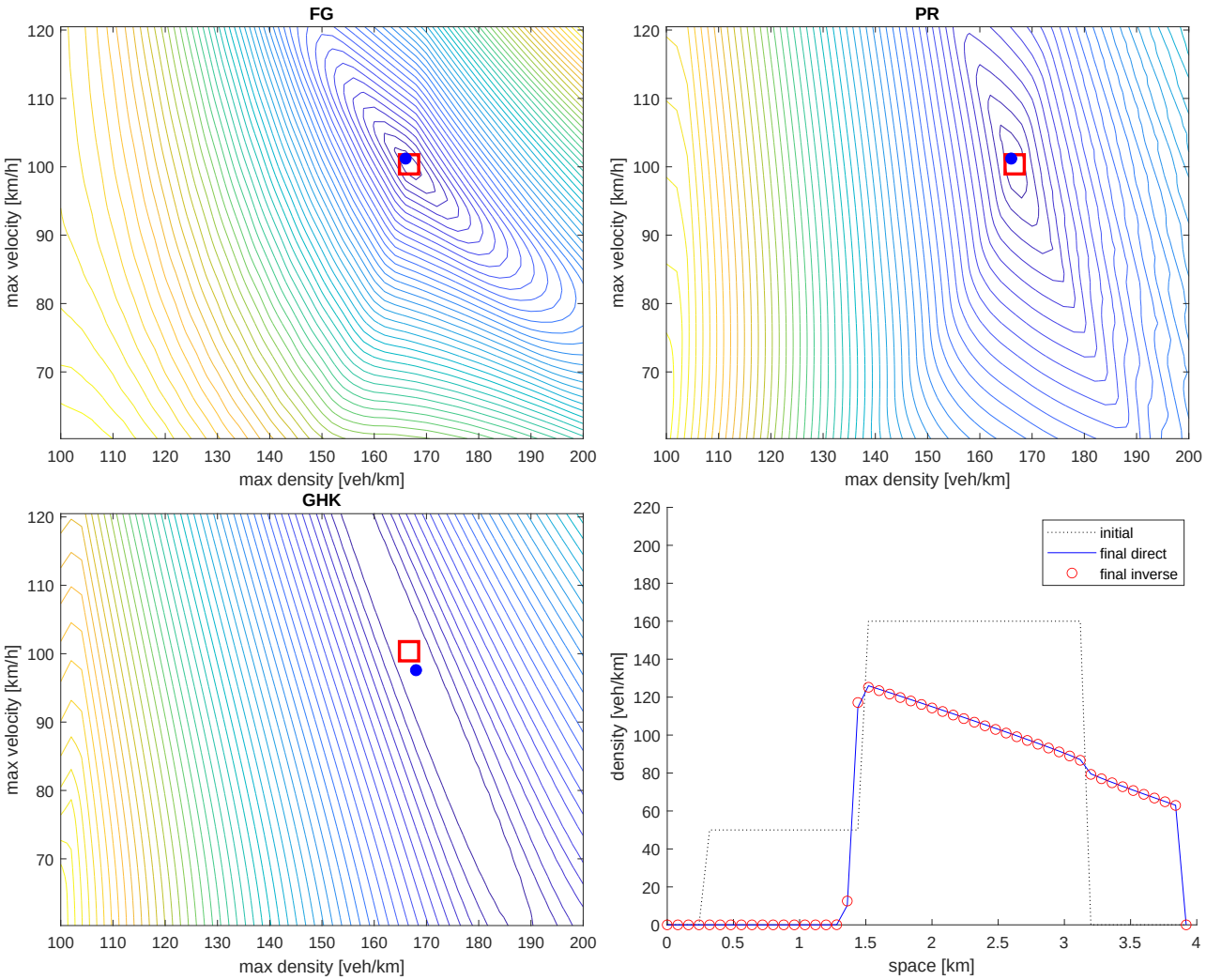

Figure 13. Test 4a on calibration. Level sets of the GWD functions $(\rho_{\text{MAX}}, v_{\text{MAX}}) \to \tilde{W}(\rho(\cdot, T; \rho_{\text{MAX}}, v_{\text{MAX}}), \rho_T^{\text{REF}})$ for FG (top left), PR (top right), and GHK (bottom left). The red square locates the exact pair $(\rho^*_{\text{MAX}}, v^*_{\text{MAX}})$, while the blue dot locates the argmin of the function. In the bottom right figure, we show the initial condition ρ_0, the reference density function ρ_T^{REF} computed in the direct problem, and the optimal density function ρ^* at the final time corresponding to the solution to the optimization problem.

Test 4a. In this test, for the direct problem, we set $\rho^*_{\text{MAX}} = 167$ veh/km, $v^*_{\text{MAX}} = 100$ km/h, $x_{\text{MIN}} = 0$ km, $x_{\text{MAX}} = 4$ km, $T = 2$ min, $\rho_{\text{IN}} = 0$ veh/km, and $\rho_{\text{OUT}} = 0$ veh/km. ρ_0 is plotted in Figure 13 (bottom right).

For the inverse problem, instead, we assume that ρ_{MAX} and v_{MAX} are unknown, and we run an exhaustive search for the best match. The intervals for the search are $[100, 200]$ vehicles/km for the density and $[60, 120]$ km/h for the velocity.

In Figure 13, we show the level sets of the function $(\rho_{\text{MAX}}, v_{\text{MAX}}) \to \tilde{W}(\rho(\cdot, T; \rho_{\text{MAX}}, v_{\text{MAX}}), \rho_T^{\text{REF}})$ for the FG, PR, and GHK approaches.

It is interesting to note that in all the three cases, the optimization problem is able to find the optimal pair $(\rho^*_{\text{MAX}}, v^*_{\text{MAX}})$ quite precisely, which is indeed the argmin of the GWDs. Nevertheless, the qualitative behavior of the level sets around the minimum point is different. FG and PR show the best

behavior, with the minimum point well contoured by level sets, meaning that the minimum is well localized. Conversely, GHK shows a plateau around the argmin, which clearly makes the minimum less identifiable.

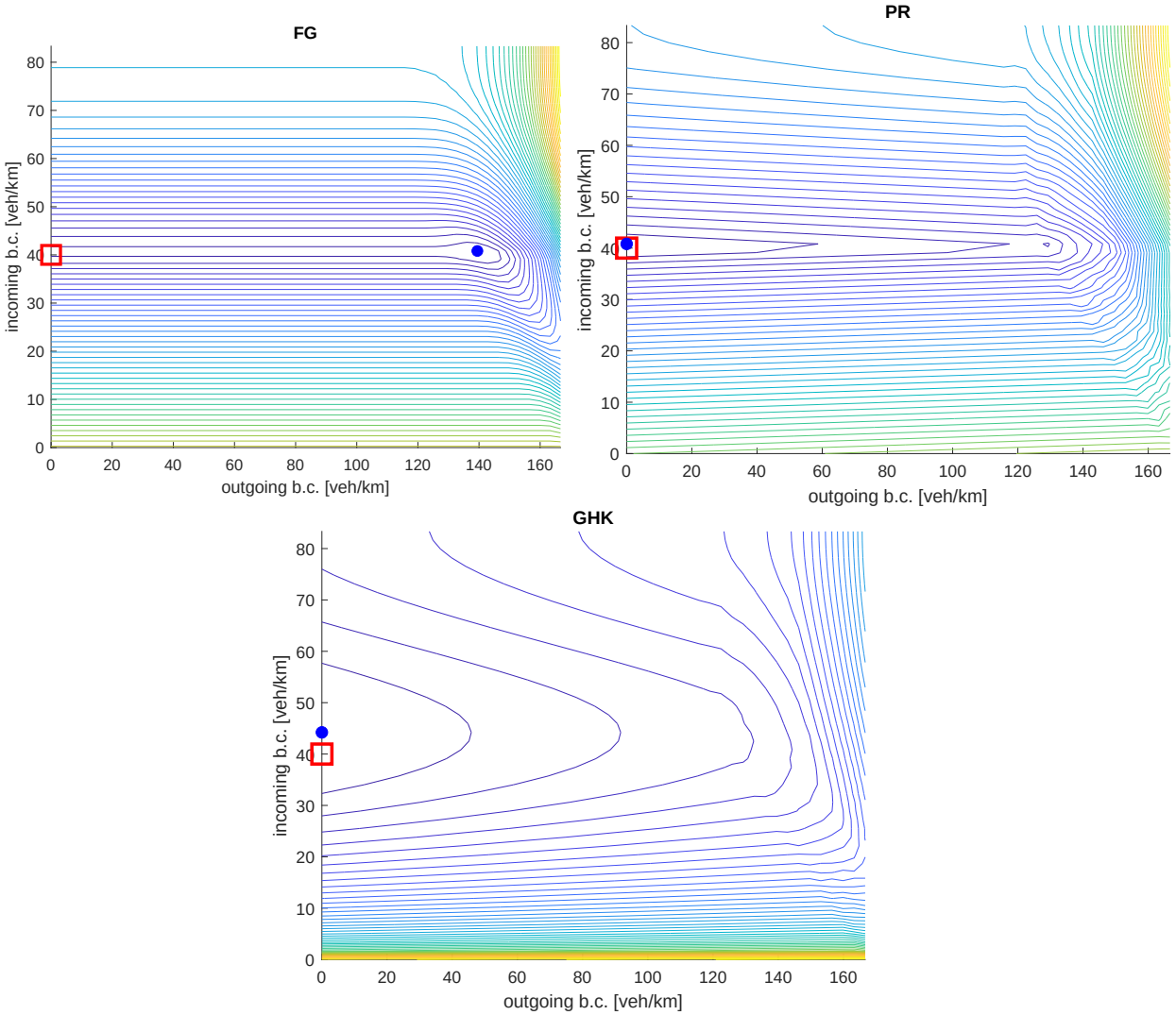

Figure 14. Test 4b on validation. Level sets of the GWD functions $(\rho_{\text{OUT}}, \rho_{\text{IN}}) \to \tilde{W}(\rho(\cdot, T; \rho_{\text{IN}}, \rho_{\text{OUT}}), \rho_T^{\text{REF}})$ for FG (top left), PR (top right), and GHK (bottom). The red square locates the exact pair $(\rho^*_{\text{OUT}}, \rho^*_{\text{IN}})$, while the blue dot locates the argmin of the function.

Test 4b In this test, for the direct problem, we set $\rho_{\text{MAX}} = 167$ vehicles/km, $v_{\text{MAX}} = 100$ km/h, $x_{\text{MIN}} = 0$ km, $x_{\text{MAX}} = 2$ km, $T = 3$ min, $\rho^*_{\text{IN}} = 40$ vehicles/km, $\rho^*_{\text{OUT}} = 0$ vehicles/km, and $\rho_0 \equiv 0$. Moreover, we assume that between $t = 1$ min and $t = 1.7$ min, an accident at 1.6 km blocks the traffic.

For the inverse problem, instead, we assume that ρ_{IN} and ρ_{OUT} are unknown, and the accident is also unknown (this is a typical situation which can happen during model validation). The intervals for the search are $[0, \rho_{\text{MAX}}/2]$ vehicles/km for the inflow boundary conditions and $[0, \rho_{\text{MAX}}]$ km/h for the outflow boundary conditions.

Conversely to Test 4a, here, it is very possible that the optimization problem is not able to find the perfect match with the reference solution, since the accident is not considered at all in the inverse problem. Indeed, the optimization problem could tend to compensate for the internal discrepancy due to the accident using different boundary conditions.

In Figure 14, we show the level sets of the function $(\rho_{\text{OUT}}, \rho_{\text{IN}}) \to \tilde{W}(\rho(\cdot, T; \rho_{\text{IN}}, \rho_{\text{OUT}}), \rho_T^{\text{REF}})$ for the FG, PR, and GHK approaches.

Since the accident happens close to the boundary, and the control parameters $\rho_{\text{IN}}, \rho_{\text{OUT}}$ act at the boundaries as well, FG struggles to find the optimal parameters, being basically insensitive to any possible variation. This is evident by observing the plateau around the minimum, which also lead to a completely wrong estimation of the boundary conditions. The other two methods, instead, are more sensitive to variations near the boundaries and are able to locate the exact solution even if the accident is ignored.

6. Conclusions

In this paper, we have proposed four algorithms for computing, in a discrete context, four GWDs (two of them theoretically equivalent to each other). One of them (SS) proves to be numerically unfeasible, while the other three (FG, PR, GHK) are all relatively simple to implement once the computational procedure is arranged.

Regarding traffic modeling, the different ways the road boundaries are treated makes the choice of the GWD to lean toward PR in the phase of calibration and validation, i.e., when one has to estimate, at best, the model parameters, while it makes the choice to lean toward FG when the model is used for traffic forecast, both for computational and modeling reasons: Indeed, FG is the fastest in terms of CPU time and naturally focuses on the internal part of the road rather than on the boundaries. This feature is useful, since traffic forecast at the boundaries is necessarily inaccurate due to the impossibility of estimating inflow and outflow traffic with good precision.

In the future, one could explore the possibility of computing the considered GWDs in more complicated domain like $\mathbb{R}^2$, $\mathbb{R}^3$, and (road) networks.

Funding

M.B. and E.C. would like to thank the Italian Ministry of University and Research (MUR) to support this research with funds coming from PRIN Project 2022 PNRR entitled “Heterogeneity on the road-Modeling, analysis, control”, No. 2022XJ9SX.

E.C. would like to thank the Italian Ministry of University and Research (MUR) to support this research with funds coming from PRIN Project 2022 entitled “Optimal control problems: Analysis, approximation and applications”, No. 2022238YY5.

G.F. would like to thank the Italian Ministry of University and Research (MUR) to support this research with funds coming from PRIN Project 2022 entitled “Variational analysis of complex systems in materials science, physics and biology”, No. 2022HKBF5C, CUP B53D23009290006.

This study was carried out within the Spoke 7 of the MOST – Sustainable Mobility National Research Center and received funding from the European Union Next-Generation EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2,

INVESTIMENTO 1.4 – D.D. 1033 17/06/2022, CN00000023).

This manuscript reflects only the authors' views and opinions. Neither the European Union nor the European Commission can be considered responsible for them.

E.C. and F.L.I. are funded by INdAM–GNCS Project, CUP E53C23001670001, entitled "Numerical modeling and high-performance computing approaches for multiscale models of complex systems".

F.L.I. is funded by Sapienza – University of Rome, project "Partial differential equations towards control theory and climate modeling" (project code: RM124190DEC62D0A).

M.B., E.C., and F.L.I. are members of the e INdAM research group GNCS, while G.F. is member of the e INdAM research group GNAMPA.

Use of AI tools declaration

The authors declare they have not used artificial intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors want to thank Benedetto Piccoli, Francesco Rossi, Giuseppe Savaré, and Giacomo Enrico Sodini for the useful discussions, and Anna Thünen for her contribution to Section 4.1.2.

Conflict of interest

Emiliano Cristiani is an editorial board member of *Networks and Heterogeneous Media* and was not involved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

Author contributions

All authors contributed equally to this work.

References

1. E. Cristiani, B. Piccoli, A. Tosin, *Multiscale Modeling of Pedestrian Dynamics*, Springer Cham, 978-3-319-06620-2, 2014. https://doi.org/10.1007/978-3-319-06620-2
2. M. Briani, E. Cristiani, E. Iacomini, Sensitivity analysis of the LWR model for traffic forecast on large networks using Wasserstein distance, *Commun. Math. Sci.*, **16** (2018), 123–144. https://doi.org/10.4310/CMS.2018.v16.n1.a6
3. E. Cristiani, M. C. Saladino, Comparing comparisons between vehicular traffic states in microscopic and macroscopic first-order models, *Math. Methods Appl. Sci.*, **42** (2019), 918–934. https://doi.org/10.1002/mma.5395
4. A. Figalli, N. Gigli, A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions, *J. Math. Pures Appl.*, **94** (2010), 107–130. https://doi.org/10.1016/j.matpur.2009.11.005

5. B. Piccoli, F. Rossi, Generalized Wasserstein distance and its application to transport equations with source, *Arch. Ration. Mech. Anal.*, **211** (2014), 335–358. https://doi.org/10.1007/s00205-013-0669-x
6. S. Kondratyev, L. Monsaingeon, D. Vortnikov, A new optimal transport distance on the space of finite Radon measures, *Adv. Differ. Equ.*, **21** (2016), 1117–1164.
7. L. Chizat, G. Peyré, B. Schmitzer, F. X. Vialard, An interpolating distance between optimal transport and Fisher–Rao metrics, *Found. Comput. Math.*, **18** (2018), 1–44. https://doi.org/10.1007/s10208-016-9331-y
8. L. Chizat, G. Peyré, B. Schmitzer, F. X. Vialard, Unbalanced optimal transport: Dynamic and Kantorovich formulations, *J. Funct. Anal.*, **274** (2018), 3090–3123. https://doi.org/10.1016/j.jfa.2018.03.008
9. M. Liero, A. Mielke, G. Savaré, Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures, *Invent. Math.*, **211** (2018), 969–1117. https://doi.org/10.1007/s00222-017-0759-8
10. Z. Ma, X. Wei, X. Hong, H. Lin, Y. Qiu, Y. Gong, Learning to count via unbalanced optimal transport, *Proc. AAAI Conf. Artif. Intell.*, **35** (2021), 2319–2327. https://doi.org/10.1609/aaai.v35i3.16332
11. G. Savaré, G. E. Sodini, A relaxation viewpoint to Unbalanced Optimal Transport: Duality, optimality and Monge formulation, *J. Math. Pures Appl.*, **188** (2024), 114–178. https://doi.org/10.1016/j.matpur.2024.05.009
12. L. Lombardini, F. Rossi, Obstructions to extension of Wasserstein distances for variable masses, *Proc. Am. Math. Soc.*, **150** (2022), 4879–4890. https://doi.org/10.1090/proc/16030
13. S. Serfaty, L. Ambrosio, E. Mainini, Gradient flow of the Chapman–Rubinstein–Schatzman model for signed vortices, *Ann. Inst. H. Poincaré Anal. Non Linéaire*, **28** (2011), 217–246. https://doi.org/10.1016/j.anihpc.2010.11.006
14. E. Mainini, A description of transport cost for signed measures, *J. Math. Sci.*, **181** (2012), 837–855. https://doi.org/10.1007/s10958-012-0718-2
15. C. Villani, *Optimal Transport*, Springer Berlin, Heidelberg, 2009. https://doi.org/10.1007/978-3-540-71050-9
16. C. Villani, *Topics in Optimal Transportation*, American Mathematical Society, Lyon, France, 2003. https://doi.org/10.1090/gsm/058
17. F. Santambrogio, *Optimal Transport for Applied Mathematicians,* Birkhäuser Cham, 2015. https://doi.org/10.1007/978-3-319-20828-2
18. S. M. Sinha, *Mathematical Programming*, Elsevier Cham, 2006. https://doi.org/10.1016/B978-81-312-0376-7.X5000-3
19. L. G. Hanin, Kantorovich–Rubinstein norm and its application in the theory of Lipschitz spaces, *Proc. Am. Math. Soc.*, **115** (1992), 345–352.
20. L. Kantorovich, G. S. Rubinstein, On a space of totally additive functions, *Vestn. Leningr. Univ.*, **13** (1958), 52–59.

21. B. Piccoli, F. Rossi, On properties of the generalized Wasserstein distance, *Arch. Ration. Mech. Anal.*, **222** (2016), 1339–1365. https://doi.org/10.1007/s00205-016-1026-7
22. L. Caffarelli, R. J. McCann, Free boundaries in optimal transport and Monge–Ampère obstacle problems, *Ann. Math.*, **171** (2010), 673–730. https://doi.org/10.4007/annals.2010.171.673
23. A. Figalli, The optimal partial transport problem, *Arch. Ration. Mech. Anal.*, **195** (2010), 533–560. https://doi.org/10.1007/s00205-008-0212-7
24. J. D. Benamou, Y. Brenier, A computational fluid mechanics solutions to the Monge–Kantoriovich mass transfer problem, *Numer. Math.*, **84** (2000), 375–393. https://doi.org/10.1007/s002110050002
25. M. J. Lighthill, G. B. Whitham, On kinematic waves II. A theory of traffic flow on long crowded roads, *Proc. R. Soc. Lond. Ser. A*, **229** (1955), 317–345. https://doi.org/10.1098/rspa.1955.0089
26. P. I. Richards, Shock waves on the highway, *Oper. Res.*, **4** (1956), 42–51. https://doi.org/10.1287/opre.4.1.42
27. A. Aw, M. Rascle, Resurrection of "second order" models of traffic flow, *SIAM J. Appl. Math.*, **60** (2000), 916–938. https://doi.org/10.1137/S0036139997332099
28. H. M. Zhang, A non-equilibrium traffic model devoid of gas-like behavior, *Transportation Res. Part B*, **36** (2002), 275–290. https://doi.org/10.1016/S0191-2615(00)00050-3
29. M. Briani, E. Cristiani, E. Onofri, Inverting the fundamental diagram and forecasting boundary conditions: how machine learning can improve macroscopic models for traffic flow, *Adv. Comput. Math.*, **50** (2024), 115. https://doi.org/10.1007/s10444-024-10206-8

A. Appendix

A.1. Numerical tests for the Figalli & Gigli approach

Test 1

In this test, we consider the case of atomic balanced masses in the domain $X = [0, 5]$. We set $N = 50$ so that $\Delta x = 0.1$. The total supply mass is 0.2, divided into two equal parts and concentrated in two points. The total demand mass is also 0.2, concentrated in one point only.

1.1. Here, the supply mass is located at $x = 1$ and $x = 2.5$, while the demand mass is located at $x = 3$; see Figure A1 (left). The optimal mass transport map prescribes to move the supply mass onto the demand mass, without resorting to mass at the boundaries. We have $\tilde{W}_1^{\text{FG}} = 0.25$.

1.2. Here, the supply mass is located at $x = 0.5$ and $x = 2.5$, while the demand mass is located at $x = 4$; see Figure A1 (right). The optimal mass transport map prescribes to move the leftmost supply mass to the left boundary, the rightmost supply mass onto the demand mass, and take mass from the right boundaries to bring them onto the demand mass. We have $\tilde{W}_1^{\text{FG}} = 0.3$.

Test 2

In this test, we consider the case of continuous unbalanced distributions in the domain $X = [-5, 5]$,

$$\rho^{\text{S}}(x) = e^{-x^2}, \qquad \rho^{\text{D}}(x) = \frac{e^{-(x-3)^2}}{4}.$$

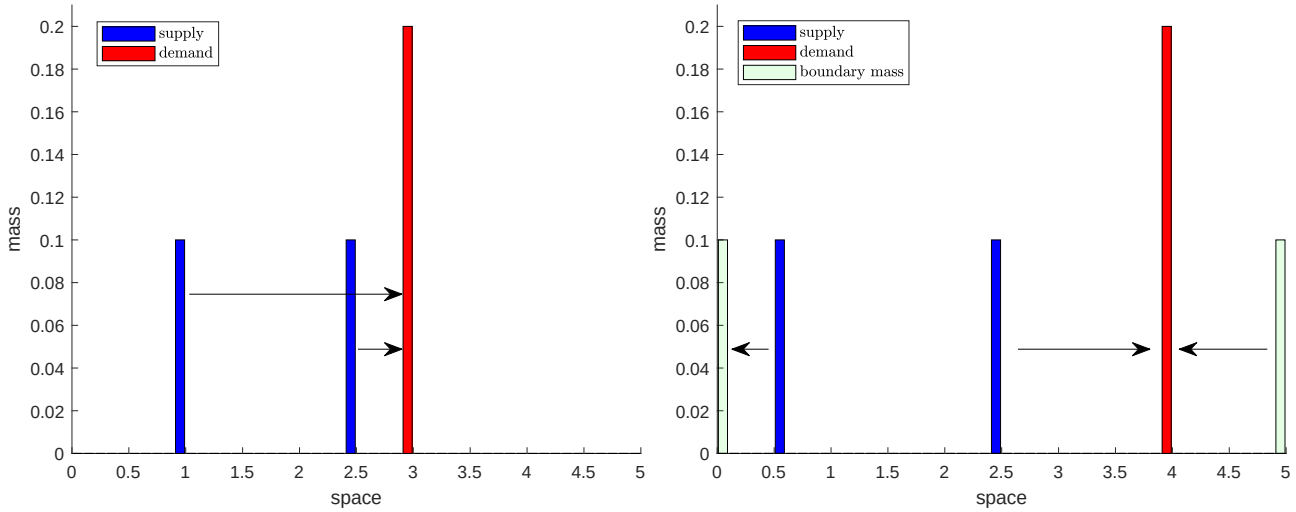

Figure A1. Test 1. Case 1.1. where boundaries are not used (left), and Case 1.2. where the mass is both taken from and brought to the boundaries (right).

See Figure A2. We set $N = 50$ such that $\Delta x = 0.2$.

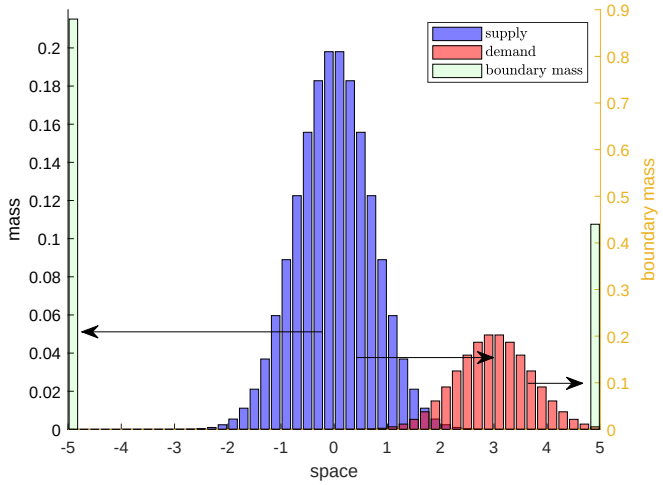

Figure A2. Test 2. Some supply (resp., demand) mass is brought to the left (resp., right) boundary, and the rest is transported.

In this case, the optimal mass transport map prescribes to move part of the supply mass to the left boundary and part of the demand mass to the right boundary, then transport the remaining mass. We have $\tilde{W}_1^{\mathrm{FG}} = 6.84$.

A.2. Numerical tests for the Piccoli & Rossi approach

Test 1

Consider the simple example proposed in [5, Section 2.2], where the authors consider the two constant distributions

$$\rho^{\mathrm{S}}(x) = \chi_{[-1,0]}(x), \qquad \rho^{\mathrm{D}}(x) = \chi_{[\xi,1+\xi]}(x), \qquad x \in \mathbb{R}$$

with the parameter $\xi \geq 0$. The exact value of the GWD (with $a = b = p = 1$) is easily computed and gives

$$W_1^{\mathrm{PR}}(\rho^{\mathrm{s}}, \rho^{\mathrm{D}}; 1, 1) = \begin{cases} 1 + \xi - \xi^2/4, & \text{if } \xi \in [0, 2], \\ 2, & \text{if } \xi \geq 2. \end{cases}$$

Figure A3 shows the result of the numerical computation in the domain $X = [-2, 5]$ divided into $N = 200$ cells for $\xi \in [0, 3]$.

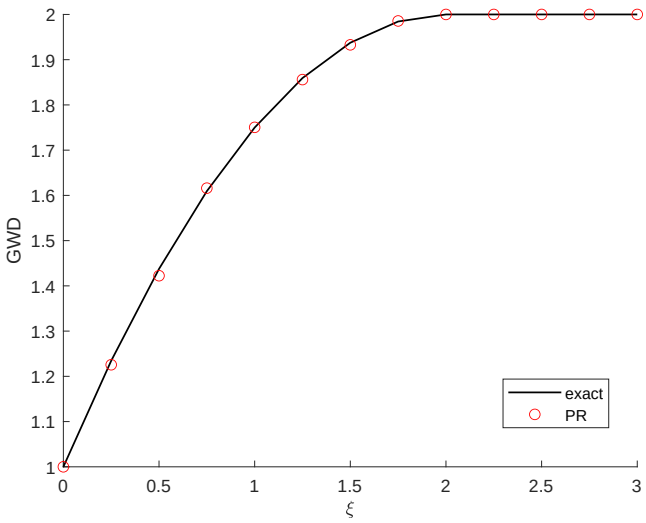

Figure A3. Test 1. Exact solution compared with numerical solution ($N = 200$).

Test 2

In this test, we run the algorithm in a nontrivial case where the exact solution is not known. In more detail, we consider $X = [-4, 4]$, $N = 100$, $a = b = p = 1$, and

$$\rho^{\mathrm{s}}(x) = \frac{e^{-x+1}}{5}\chi_{[-2,0]}(x), \qquad \rho^{\mathrm{D}}(x) = e^{-(x-1)^2}.$$

Figure A4 shows the result in terms of the remaining mass after minimization. One can see that here, the optimal solution ($\tilde{W}_1^{\mathrm{PR}} = 4.36$) requires to destroy only part of the mass.

A.3. Numerical tests for the Gaussian Hellinger–Kantorovich and the Savaré & Sodini approaches

Here, we present numerical tests for the approaches described in Sects. 4.3.2 and 4.4.2. We recall that the two approaches, although quite different from the computational point of view, are theoretically equivalent (with $a = b = 1$, which will be always the case from now on). This is also confirmed by our numerical experiments.

Test 1

This test is mainly intended to check the correctness of the code on simple examples with known exact solution. Let us consider the simplest possible scenario, with only one concentrated supply mass δ^{s} and only one concentrated demand mass δ^{D}. To simplify the notations, we denote by m^{s}, x^{s} and by

m^{D}, x^{D} the mass and the position of the supply mass and the demand mass, respectively. In this case the exact solution is $\tilde{W}_2^{\mathrm{GHK}} = \tilde{W}_2^{\mathrm{SS}} = H_g(x^{\mathrm{S}}, m^{\mathrm{S}}; x^{\mathrm{D}}, m^{\mathrm{D}})$, where H_g is defined in Eq (3.9).

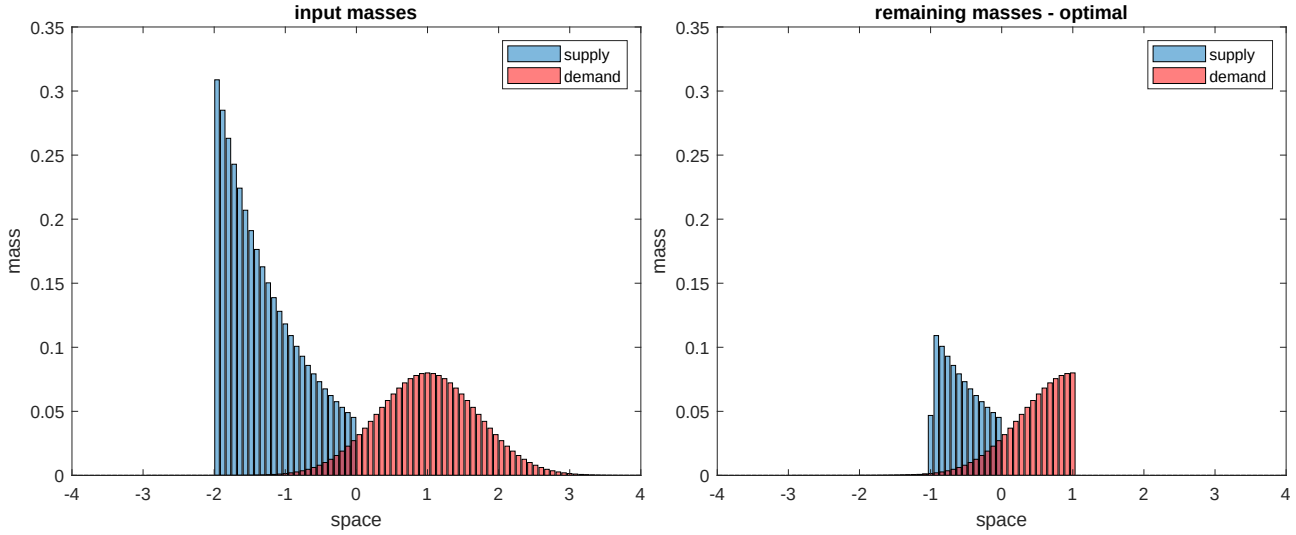

Figure A4. Test 2. The initial distributions m^{S} and m^{D} (left) and the remaining masses $\mathring{m}^{\mathrm{S}}$ and $\mathring{m}^{\mathrm{D}}$ corresponding to the optimal solution (right).

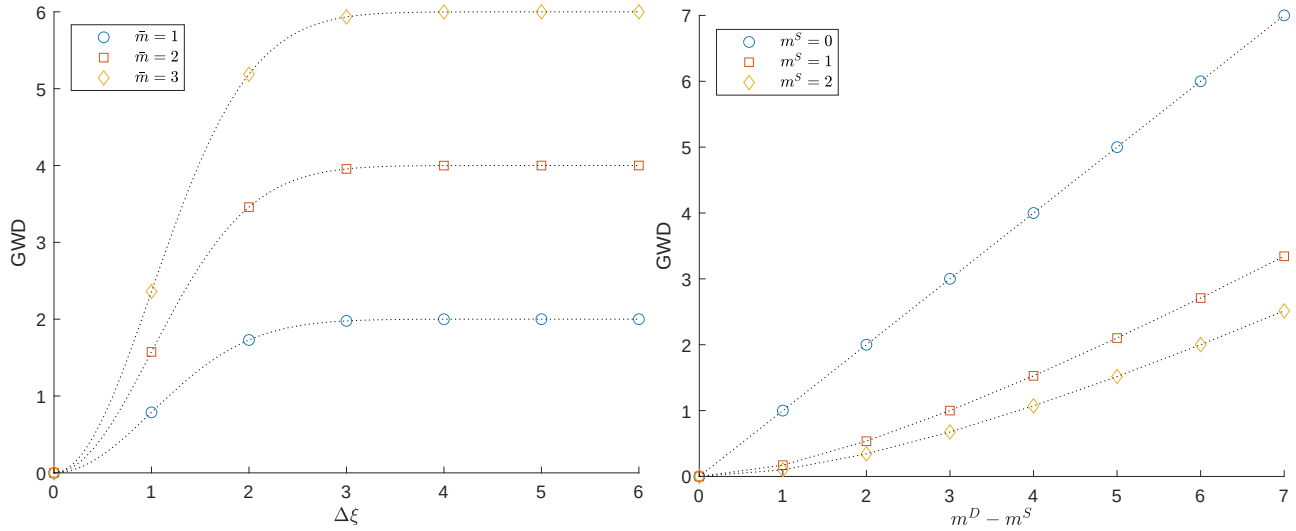

Figure A5. Test 1. The GWD between two equal atomic mass distributions with a mass $\bar{m}$ as a function of their distance $\Delta\xi$ (left), and the GWD between two overlapping atomic mass distributions as a function of the difference of mass (right). The dotted black line is the exact solution.

Figure A5 shows the results of the algorithm for $\tilde{W}_2^{\mathrm{GHK}}$ as well as both algorithms for $\tilde{W}_2^{\mathrm{SS}}$ (SS-A and SS-B), alongside the exact solution.

In more detail, Figure A5 (left) shows the GWD in the case when the two distributions are balanced ($m^{\mathrm{S}} = m^{\mathrm{D}} = \bar{m}$) and their distance progressively increases from 0 to 6. One can see that the GWD rapidly saturates at $2\bar{m}$. As for $\tilde{W}_2^{\mathrm{GHK}}$, this is consistent with the fact that increasing the distance results in relaxing the constraint (4.5) for the maximization of the dual functional, which involves

exponential functions: From this, we obtain the maximum value $2\bar{m}$ in the limit case of an infinite distance and the exponential decay of the difference between this value and the actual maximum for finite distances that grow large. As for $\tilde{W}_2^{\mathrm{ss}}$, recalling Eq (3.9), the objective in Eq (4.6) decays exponentially to $r^{\mathrm{s}}_{jh} + r^{\mathrm{D}}_{k\ell}$ which, in turn, equals at most $2\bar{m}$ due to the constraints on the homogeneous marginals.

Figure A5 (right) instead shows the GWD in the case when the two distributions are located at the same point and the difference of their masses progressively increases from 0 to 7.

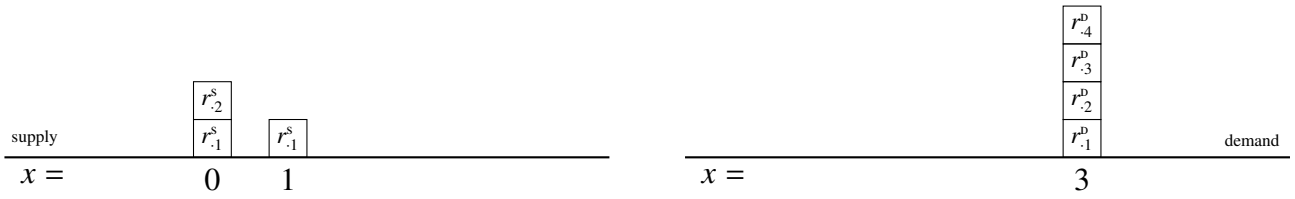

Figure A6. Test 2: Two concentrated masses for the supply mass and one for the demand mass.

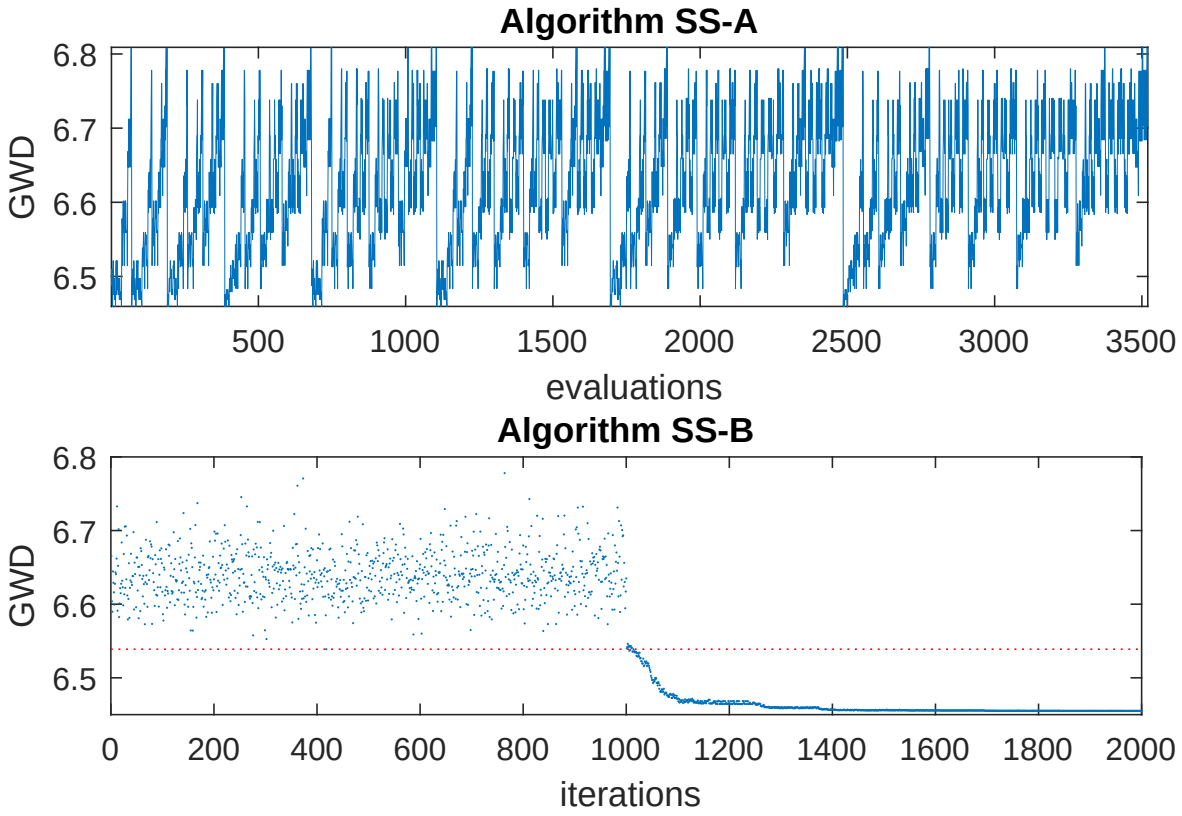

Figure A7. Test 2. Algorithm SS-A, complete view of all the evaluations of $\tilde{W}_2^{\mathrm{ss}}$ (top), and Algorithm SS-B, with 1000 random searches followed by about 1000 descend iterations (bottom). The dotted red line indicates the minimum value found in the first phase of SS-B.

Test 2

The goal of this test is mainly to compare the algorithms SS-A and SS-B. We consider another simple scenario with only two concentrated masses for supply, and only one concentrated mass for demand; see Figure A6. We assume that the supply masses are located at $x = 0$ and $x = 1$, have the masses $m = 2$ and $m = 1$, and have $n = 2$, $n = 1$ additional divisions, respectively. The demand mass is instead located at $x = 3$, has a mass $m = 4$, and has $n = 4$ additional divisions. This scenario results in $d = 12$.

Figure A7 (top) shows the history of all the evaluations of $\tilde{W}_2^{\text{ss}}$ done by the exhaustive algorithm SS-A, in the same order as they are computed, with $Q = 6$. One can observe a recursive pattern due to the way Γ's are generated (Step 2). The minimum is already found among the first evaluations and it is $\tilde{W}_2^{\text{ss}} = 6.45$.

Figure A7 (bottom) shows the tentative values of $\tilde{W}_2^{\text{ss}}$ found in the first exploratory phase of the algorithm SS-B (Steps 1–3), followed by the minimum values found during the random descent, with $Q = 1000$.

The algorithm again returns $\tilde{W}_2^{\text{ss}} = 6.45$. Convergence seems to be well guaranteed, and the two algorithms give the same value for $\tilde{W}_2^{\text{ss}}$ ($\tilde{W}_2^{\text{GHK}}$ is also the same).

For these reasons, we use algorithm SS-B only, which is obviously faster.