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:** e.cristiani@iac.cnr.it; francescalourdes.ignoto@uniroma1.it

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

Contents

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 (like, e.g., any of the model 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 which distance is more suitable to 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, e.g., the discussion in [9, 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., [4, 9, 10]. The drawback is that WD is limited to *balanced* mass distribution (equal masses), while real traffic problems often need to consider scenarios with a different amount of

vehicles, especially because of different inflow/outflow at boundaries. This suggests to move towards GWDs, which allows to deal with *unbalanced* mass distributions, but serious computational problems arise: all existing definitions are quite abstract and it is not immediate to 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, that one considers to define the classical Wasserstein distance between two probability measures μ^s and μ^D, the minimization occurs upon transport plans γ, i.e., probability measures on $X \times X$ with marginals μ^s and μ^D. Namely, one considers the projection π^s (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

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

where $\pi^s_\sharp \gamma$ (resp., $\pi^D_\sharp \gamma$) 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 exist 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 such approaches.

1. The approach by Figalli & Gigli [12], which consists in minimizing the integral cost (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 [21], which includes in the definition of the functional the additive contributions

$$|\pi^s_\sharp \gamma - \mu^s|(\Omega) + |\pi^D_\sharp \gamma - \mu^D|(\Omega) \,, \tag{1.3}$$

   in addition to (1.1), with the minimization occurring upon 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, Monsaingeon & Vorotnikov in [15], by Chizat, Peyré, Schmitzer & Vialard in [7, 8], and by Liero, Mielke & Savaré in [16], 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 (1.1), except that the penalization (1.3) is replaced by

$$\int_\Omega \left(\sigma^s \log(\sigma^s) - \sigma^s + 1 \right) d\mu^s + \int_\Omega \left(\sigma^D \log(\sigma^D) - \sigma^D + 1 \right) d\mu^D \,, \tag{1.4}$$

   if $d(\pi^s_\sharp \gamma) = \sigma^s \, d\mu^s$ and $d(\pi^D_\sharp \gamma) = \sigma^D \, d\mu^D$ for appropriate densities σ^s and σ^D, and $+\infty$ otherwise. (Incidentally, a similar objective, in which an entropic regularization is added to (1.1) and (1.4), in considered in [19], where the appropriate dual formulation of the problem is addressed by Sinkhorn iterations.)

4. A fourth approach relates to the contribution by Savaré & Sodini [25], 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 to numerically approximate 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 WD, despite defining a distance on the cone (see Remarks 3 and 4 below). We recall that no extension of WD to the set of signed measures, possibly zero, can define a distance [18] (non-canonical optimal transport distance have however been successfully defined, see [1, 20]).

Main contribution. This paper has multiple goals: first of all, the four approaches are recalled and compared at the theoretical level, so 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 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 important, we propose computational approaches for their approximation. The distances are then compared on an academic numerical test. In addition – postponed 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 capability 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 final time or in both space and time). Although this can be satisfactory for nearly equal outputs or for convergence results, it appears 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
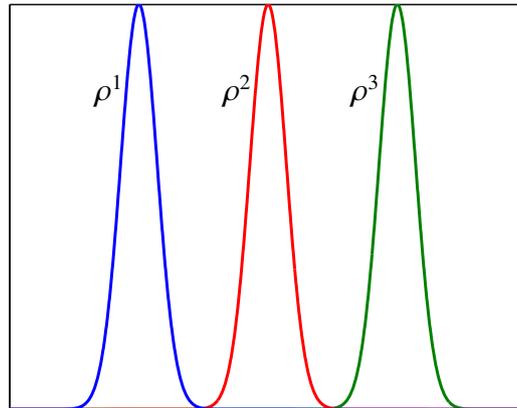

Figure 1. Three density functions with disjoint supports.

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 the supports of them are disjoint. Our perception of distance suggests instead that $\|\rho_3 - \rho_1\| > \|\rho_2 - \rho_1\|$, and this is exactly what Wasserstein distance guarantees, as we will see in the next section.

2.2. Basic theory

Let us denote by $(X, \mathfrak{D})$ a locally compact complete and separable metric space with 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} (D standing for demand or destination) and μ^{s} (s standing for supply or source) be two Radon measures in $\mathcal{M}^+(X)$ such that $\mu^{\text{D}}(X) = \mu^{\text{s}}(X)$ (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*

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

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

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

Remark 1. *The distance defined by* (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 put in relation with the Monge–Kantorovich optimal mass transfer problem [28]: 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 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 onto the second one. We also stress that, since we are considering macroscopic models, vehicles are indistinguishable so we are not able to pinpoint the exact same vehicle in the two scenarios, cf. [10].*

In the particular case of 1D optimal transport various characterizations give alternative, more manageable, definitions of 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 $W_p(\delta_x, \delta_y) = \mathfrak{D}(x, y) := |y - x|$. That 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 $d\mu^s = \rho^s dx$ and $d\mu^D = \rho^D dx$, by [27, Remark 2.19] we have

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

with

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

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

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

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

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

Equations (2.4)-(2.5) translate the fundamental property of *monotone rearrangement* of the mass [27, Remark 2.19], i.e. the optimal strategy consists in 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 one-dimensional interval $X = [x_{\text{MIN}}, x_{\text{MAX}}]$, divided in N equispaced cells $C_1, \ldots, C_N$ (N is our first parameter of discretization, and we expect to vanish the numerical error for $N \to +\infty$). We also denote by x_i the barycenter of the cell C_i, $i = 1, \ldots, N$ 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 $\mathfrak{D}(x_j, x_k) := |x_k - x_j|$.
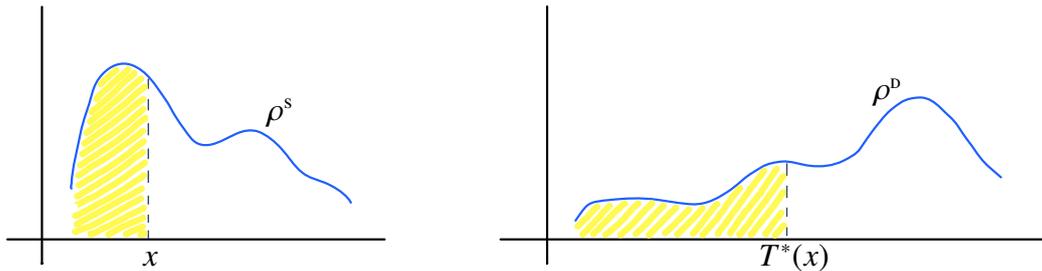
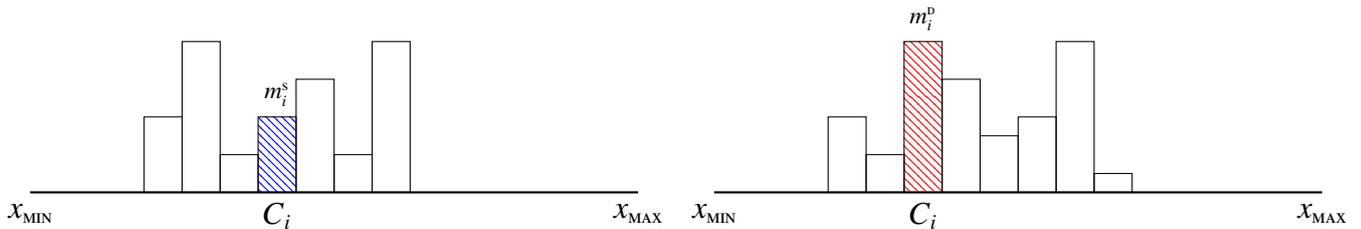**Figure 2.** Property of monotone rearrangement.

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

Moving to the masses, for any $i = 1, \ldots, N$ 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 concentrate 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 boundaries remain out of play.

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

In order to get a numerical approximation of the WD in $\mathbb{R}^1$, it is convenient to resort to the definition (2.2). Indeed, any quadrature formula is suitable to get the desired value with arbitrary precision, provided $\rho^\text{s}, \rho^\text{D}$ are sufficiently smooth and N is sufficiently large. For example, the naive composite mid-point 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 [24, Section 6.4.1] and Sinha [26, Chapt. 19] the complete procedure, recently used also in the context of traffic flow on networks [4]. Even if we will stick with the one-dimensional case, we recall now 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 (2.9), it is easy to recover a standard LP problem with equality constraints

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

simply defining

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

and $\mathbf{A}$ as 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 inspired to the distance defined by (2.1) between probability measures.

3.1. Figalli & Gigli approach [12]

In [12], 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 μ^s, μ^p on Ω is defined by

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

where infimum is over

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

In the latter, we used 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 denoted by $\pi^s_\sharp \gamma$ (resp., $\pi^p_\sharp \gamma$) the push-forward of γ under the projection defined on $\overline{\Omega} \times \overline{\Omega}$ by $\pi^s(x, y) = x$ (resp., $\pi^p(x, y) = y$). Equivalently, a non-negative finite Radon measure γ on $\overline{\Omega} \times \overline{\Omega}$ is an admissible competitor for the minimization in (3.1) if and only if

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

Let us note that (3.2) coincides with the set

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

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

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

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

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

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

The authors of [12] deal explicitly with the case of a quadratic unitary cost, i.e., $c(x, y) = \|x - y\|^2$ in (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 towards 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 (3.1) defines a distance on $\mathcal{M}^+(\Omega)$, and the infimum is achieved by an appropriate optimal plan γ.

The idea of minimizing (3.1) among all γ as in (3.2) is that $\partial\Omega$ models an infinite mass reservoir and the transportation policy described by 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 to integrate 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 had been already investigated especially in the case of a linear cost ($p = 1$), see [13, 14].

Remark 3. *It is clear from* (3.4) *that the distance between positive finite Radon measures introduced in [12] is not an extension of the classical Wasserstein distance: if the two prescribed marginals are concentrated close to the boundary with supports far apart from each other, then while solving* (3.4) *there will be a convenience for $\bar{\mu}^s$ and $\bar{\mu}^D$ in charging appropriately two corresponding disjoint boundary regions, after which the Wasserstein to be minimized will be achieved by transport plans that only move mass from or to the boundary. In fact, the goal of [12] was to generalize Otto's calculus to PDEs under Dirichlet conditions, rather than extending the optimal transport metrics to finite positive Radon measures.*

3.2. Piccoli & Rossi approach [21]

In [21], the set of positive finite Borel measures on Ω is endowed with a metrics, defined by

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

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

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

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

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

The infimum in (3.6), and hence in (3.5) too, is indeed achieved and it defines a distance [21, 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 $\mathring{\mu}^s(\Omega) \leq \mu^s(\Omega)$ and $\mathring{\mu}^D(\Omega) \leq \mu^D(\Omega)$ may be added in the minimization without changing the minimum value, see [22, Prop. 2].

---

[*]Namely, we are setting $W_p(\mathring{\mu}^s, \mathring{\mu}^D) = m W_p(\mathring{\mu}^s/m, \mathring{\mu}^D/m)$, where $m = \mathring{\mu}^D(\Omega) = \mathring{\mu}^s(\Omega)$ and $\mathring{\mu}^s, \mathring{\mu}^D$ 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 above actions are admissible for the minimization: an arbitrary fraction of mass may be removed from the marginals, along with transporting the remaining mass. Accordingly, (3.5) does not extend the Wasserstein distance because in the case of two marginals with equal masses the allowed actions include the possibility of removing mass from both. On the other hand, in the case $p = 1$ with $a = b = 1$ it still offers a natural generalization of the 1-WD, being equivalent to the flat metric.*

The Piccoli & Rossi distance (3.5) differs from the Figalli & Gigli one (3.1). Recalling (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 geometry plays no role in (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 (3.5) is equivalent to the partial optimal transport problem [6, 11], i.e.,

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

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}, that 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 (3.5) is inspired by the dynamical formulation of the Monge–Kantorovich problem: by generalizing the Benamou and Brenier's approach [3], the supply and the demand distributions are thought of as the initial and the final prescribed configuration for a fluid motion and the minimal value of (3.5) is obtained by minimizing the total kinetic energy necessary to connect the two configurations so as to balance mass conservation with 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 [16]

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

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

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

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

Such a distance is called in [16] 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 (3.7) can be equivalently written in the form

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

,

where the infimum is among all positive finite measures $\hat{\mu}^s$ and $\hat{\mu}^p$ of equal mass on Ω. In order to define the last term in the right hand-side of (3.8), we have implicitly understood both $\hat{\mu}^s$ and $\hat{\mu}^p$ 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^p themselves happen to have equal mass, they form an admissible pair for the minimization: then, the right hand side of (3.8) is the classical (quadratic) WD, whose the value is a computed by solving a balanced optimal transport problem, because in view of (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 (3.7) belongs to the wide class of *optimal entropy-transport problems* considered in [16], 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 (3.7b). That large class covers the Piccoli & Rossi distance (3.6) for $p = 1$, with entropy function $U(\sigma) = |\sigma - 1|$, and is comprised of several models, alternative to [21] but equally inspired by dynamical formulations with source terms in the differential constraints, which have appeared in the literature since.

In particular, the model case of the *logarithmic optimal entropy-transport problem* (3.7) has been investigated under various perspectives: its value coincides with that of the metrics studied in [7, 15] (also called Wasserstein–Fisher–Rao distance, as it interpolates between the quadratic Wasserstein and the Fisher–Rao distance), for which a static definition had been provided in [8].

Remark 5. *It is known, e.g. by [16, Part II], that the logarithmic entropy-transport minimization* (3.7) *defines a distance that metrizes the weak/narrow convergence of non-negative finite measure, extending the quadratic WD (otherwise defined only for probability measures).*

3.4. Savaré & Sodini approach [25]

In [16, Part II] the logarithmic entropy-transport minimization (3.7) is proved to be equivalent to an optimal transport type problem in the *metric cone* $\mathbb{C}[X]$ over X. More precisely, one defines $\mathbb{C}[X]$ 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

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

where

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

Then, one considers the minimization of

$$\int_{\mathbb{C}[X] \times \mathbb{C}[X]} d_{\mathbb{C},g}^2([x^s, r^s], [x^p, r^p])\, d\alpha$$

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

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

,

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

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

4. Discrete and computational approaches for Generalized Wasserstein distances

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 towards a numerical computation. Secondly, we propose original computational approaches to get algorithms 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 instead postponed 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 one-dimensional domain, i.e.,

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

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

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

where

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

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

$$\tilde{W}_p^{\text{FG}}(m^{\text{s}}, m^{\text{D}}) := \min \left\{ \sum_{j,k=1}^{N} c_{jk} \gamma_{jk} \,\Big|\, \gamma \in [0, +\infty)^{N \times N}, \ \sum_{k=1}^{N} \gamma_{jk} = m_j^{\text{s}}, \text{ for } j \in \{2, \dots, N-1\}, \right.$$

$$\sum_{j=1}^{N} \gamma_{jk} = m_k^{\text{D}}, \text{ for } k \in \{2, \dots, N-1\},$$

$$\left. \sum_{k=1}^{N} (\gamma_{1k} + \gamma_{Nk}) - \sum_{j=1}^{N} (\gamma_{j1} + \gamma_{jN}) = \Delta M \right\}^{\frac{1}{p}}.$$

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

,

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.

4.1.2. Computational approach

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

$$\mathbf{x} := (\gamma_{11}, \gamma_{12}, \ldots, \gamma_{1N}, \gamma_{21}, \gamma_{22}, \ldots, \gamma_{2N}, \ldots, \gamma_{N1}, \ldots, \gamma_{NN})^\top$$

$$\mathbf{c} := (c_{11}, c_{12}, \ldots, c_{1N}, c_{21}, c_{22}, \ldots, c_{2N}, \ldots, c_{N1}, \ldots, c_{NN})^\top$$

$$\mathbf{b} := (m_2^s, \ldots, m_{N-1}^s, m_2^D, \ldots, m_{N-1}^D, \Delta M)^\top$$

and $\mathbf{A}$ as 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 $[\mathbb{0}_{N-2}^\top\ I_{N-2}\ \mathbb{0}_{N-2}^\top]$.

The LP formulation described above can be slightly generalized to account for the costs of creating/destroying mass at boundaries. Still keeping the simplifying assumption (2.7), the vector $\mathbf{x}$ of the unknowns is modified adding four new unknowns $M_1^s, M_N^s, M_1^D, M_N^D$, 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 $c_1^s, c_N^s, c_1^D, c_N^D$ related to the four new unknowns are added. For example, the quantity $c_1^s M_1^s$ will give the cost to create the mass M_1^s at the left boundary (N.B., just creating it, not transporting it inside the domain), while the quantity $c_N^D M_N^D$ will give the cost to destroy the mass M_N^D at the right boundary.

The PL formulation becomes:

$$\tilde{W}_p^{\text{FG}'}(m^s, m^D) := \min \left\{ \sum_{j,k=1}^{N} c_{jk} \gamma_{jk} + c_1^s M_1^s + c_N^s M_N^s + c_1^D M_1^D + c_N^D M_N^D \,\middle|\, \gamma \in [0, +\infty)^{N \times N}, \right.$$

$$\sum_{k=1}^{N} \gamma_{jk} = m_j^s, \text{ for } j \in \{2, \ldots, N-1\}, \quad \sum_{j=1}^{N} \gamma_{jk} = m_k^D, \text{ for } k \in \{2, \ldots, N-1\},$$

$$\sum_{k=1}^{N} x_{jk} = M_j^s, \quad \text{for } j \in \{1, N\}, \quad \sum_{j=1}^{N} x_{jk} = M_k^D, \quad \text{for } k \in \{1, N\},$$

$$\left. M_1^s, M_N^s, M_1^D, M_N^D \geq 0, \quad c_{11} = c_{1N} = c_{N1} = c_{NN} = +\infty \right\}^{\frac{1}{p}},$$

$$\tag{4.2}$$

,

which corresponds to (2.10) with

$$\mathbf{x} := (x_{11}, x_{12}, \ldots, x_{1N}, x_{21}, x_{22}, \ldots, x_{2N}, \ldots, x_{N1}, \ldots, x_{NN}, M_1^{\text{s}}, M_N^{\text{s}}, M_1^{\text{D}}, M_N^{\text{D}})^{\top}$$

$$\mathbf{c} := (c_{11}, c_{12}, \ldots, c_{1N}, c_{21}, c_{22}, \ldots, c_{2N}, \ldots, c_{N1}, \ldots, c_{NN}, c_1^{\text{s}}, c_N^{\text{s}}, c_1^{\text{D}}, c_N^{\text{D}})^{\top}$$

$$\mathbf{b} := (0, m_2^{\text{s}}, \ldots, m_{N-1}^{\text{s}}, 0, 0, m_2^{\text{D}}, \ldots, m_{N-1}^{\text{D}}, 0)^{\top}$$

and $\mathbf{A}$ is a $2N \times (N^2 + 4)$ block matrix, obtained from matrix (2.12) by extending it with four columns. The extra columns have a single non-zero 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 (4.2). The last constraints of (4.2) are needed to assure that it never convenient creating mass at boundaries just to keep it there or to transport it on 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

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

the distance (3.5) is computed by minimizing

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

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

$$\sum_{k=1}^{N} \gamma_{jk} = \mathring{m}_j^{\text{s}}, \quad \text{for } j \in \{1, \ldots, N\},$$

$$\sum_{j=1}^{N} \gamma_{jk} = \mathring{m}_k^{\text{D}}, \quad \text{for } k \in \{1, \ldots, N\},$$

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

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

,

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

$$\sum_{k=1}^{N} \gamma_{jk} \leq m_j^{\mathrm{s}}, \qquad \text{for all } j = 1, \dots, N,$$

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

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

$$
\begin{aligned}
\text{minimize} \quad & \hat{\mathbf{c}}^{\mathsf{T}}\mathbf{x} \\
\text{subject to} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\
& \mathbf{x} \geq 0,
\end{aligned}
\tag{4.4}
$$

where $\hat{\mathbf{c}} := b\mathbf{c} - 2a\mathbb{1}_{N^2}$, $\mathbf{x}$, $\mathbf{c}$, $\mathbf{b}$ are defined as in (2.11), and $\mathbf{A}$ is defined as in (2.12). Then, once the solution $\mathbf{x}^* = \mathbf{x}^*(a, b)$ of (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 (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 [16, 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

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

,

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 $f : \mathbb{R}^{2N} \to \mathbb{R}$,

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

Then, we aim at finding

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

under constraints

$$\mathbf{A}^{\mathsf{T}} \mathbf{z} \le \mathbf{c},$$

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

$$\mathbf{A} := \left[\begin{array}{ccccc} \mathbf{e}_1 \, \mathbf{e}_1 \cdots \mathbf{e}_1 & \mathbf{e}_2 \, \mathbf{e}_2 \cdots \mathbf{e}_2 & \cdots & \mathbf{e}_N \, \mathbf{e}_N \cdots \mathbf{e}_N \\ \mathbf{e}_1 \, \mathbf{e}_2 \cdots \mathbf{e}_N & \mathbf{e}_1 \, \mathbf{e}_2 \cdots \mathbf{e}_N & \cdots & \mathbf{e}_1 \, \mathbf{e}_2 \cdots \mathbf{e}_N \end{array} \right],$$

with $\mathbf{e}_i = (0 \, 0 \, \cdots \, \underbrace{1}_{i^{th}} \, \cdots \, 0 \, 0)^{\mathsf{T}} \in \mathbb{R}^N$ 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 [25]. In particular, in view of [25, 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 $\gamma_{(jh,k\ell)} \ge 0$, $\sum_{jhk\ell} \gamma_{(jh,k\ell)} = 1$, and for the homogeneous marginals that

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

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.
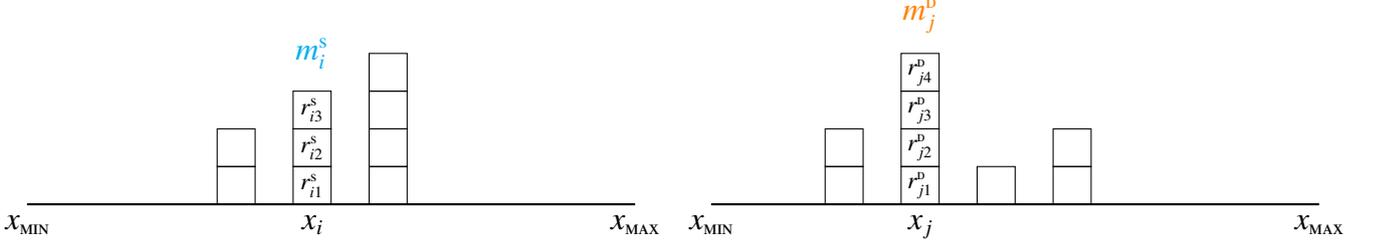

Figure 4. Additional discretization of the concentrated masses.

At each grid point x_i, the mass m_i^{s} is divided in $n_i^{\mathrm{s}} \geq 0$ intervals (reasonably, $n_i^{\mathrm{s}} = 0$ iff $m_i^{\mathrm{s}} = 0$). Analogously, the mass m_i^{D} is divided in $n_i^{\mathrm{D}} \geq 0$ intervals (again, $m_i^{\mathrm{D}} = 0$ iff $m_i^{\mathrm{D}} = 0$). 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, at each subdivision of the supply mass is associated the additional variable r_{jh}^{s}, $j = 1, \dots, N$, $h = 1, \dots, n_j^{\mathrm{s}}$, and at each subdivision of the demand mass is associated the additional variable $r_{k\ell}^{\mathrm{D}}$, $k = 1, \dots, N$, $\ell = 1, \dots, n_k^{\mathrm{D}}$.

Define also

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

as the vector of all the weights $\gamma_{(jh,k\ell)}$'s associated to 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

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

where $\mathcal{S}$ is defined in (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 (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 (4.7), we have devised two strategies: the first one is based on a slow but exhaustive search, which guarantees to find 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 $\tilde{\Gamma}$'s lead to linear systems with no solution (we have discarded them).

```
 1: for i₁ = 0, . . . , Q do
 2:     for i₂ = 0, . . . , Q − i₁ do
 3:         for i₃ = 0, . . . , Q − i₁ − i₂ do
 4:             for i₄ = 0, . . . , Q − i₁ − i₂ − i₃ do
 5:                 i₅ = Q − i₁ − i₂ − i₃ − i₄
 6:                 Γ̃ = (i₁, i₂, i₃, i₄, i₅)/Q
 7:             end for
 8:         end for
 9:     end for
10: end for
```

3. For each $\tilde{\Gamma}$, find $\{r^s_{jh}\}_{jh}$ and $\{r^p_{k\ell}\}_{k\ell}$ (depending on that $\tilde{\Gamma}$) solving the underdetermined linear systems, selecting the solution with minimal norm.
4. For each $\tilde{\Gamma}$ (and corresponding $\{r^s_{jh}\}_{jh}$, $\{r^p_{k\ell}\}_{k\ell}$), 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 a largest-as-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.

```
 1: for z = 1, . . . , d do
 2:     Extract a uniform random variable ωz ∈ {1, . . . , Q}
 3: end for
 4: Γ̃ = (ω₁, . . . , ω_d)/ ∑_z ωz
```

2. For each $\tilde{\Gamma}$, find $\{r^s_{jh}\}_{jh}$ and $\{r^p_{k\ell}\}_{k\ell}$ (depending on that $\tilde{\Gamma}$) solving the underdetermined linear systems, selecting the solution with minimal norm.
3. For each $\tilde{\Gamma}$ (and corresponding $\{r^s_{jh}\}_{jh}$, $\{r^p_{k\ell}\}_{k\ell}$), 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 initial guess, explore the control space perturbing the best Γ obtained so far with 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 by FG, PR, SS (variant B only), and GHK, with obvious meaning.

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

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

in the domain $X = [-T-\eta, T+\eta]$. The two step functions start with perfectly overlapping support, then supply mass moves leftward while demand mass moves rightward, until they both reach the boundary of the domain, see Figure 5.
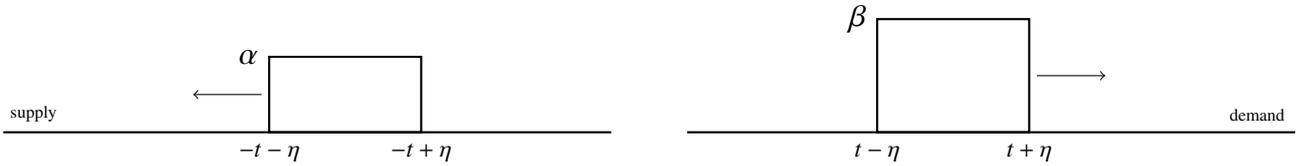
Figure 5. Moving distributions ρ^{s} and ρ^{D} as defined in (4.8).

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, $Q = 1,000$, $\epsilon = 0.001$, iterations 5,000 (random searches) + 40,000 (random descend), average on 10 runs.

Figure 6 shows the behavior of the three GWDs between ρ^{s} and ρ^{D} as defined in (4.8), as a function of time. 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 additional divisions, therefore it will be discarded in the following analysis.

Among the others, PR and GHK share with L^p an important drawback: the distances saturate when 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.
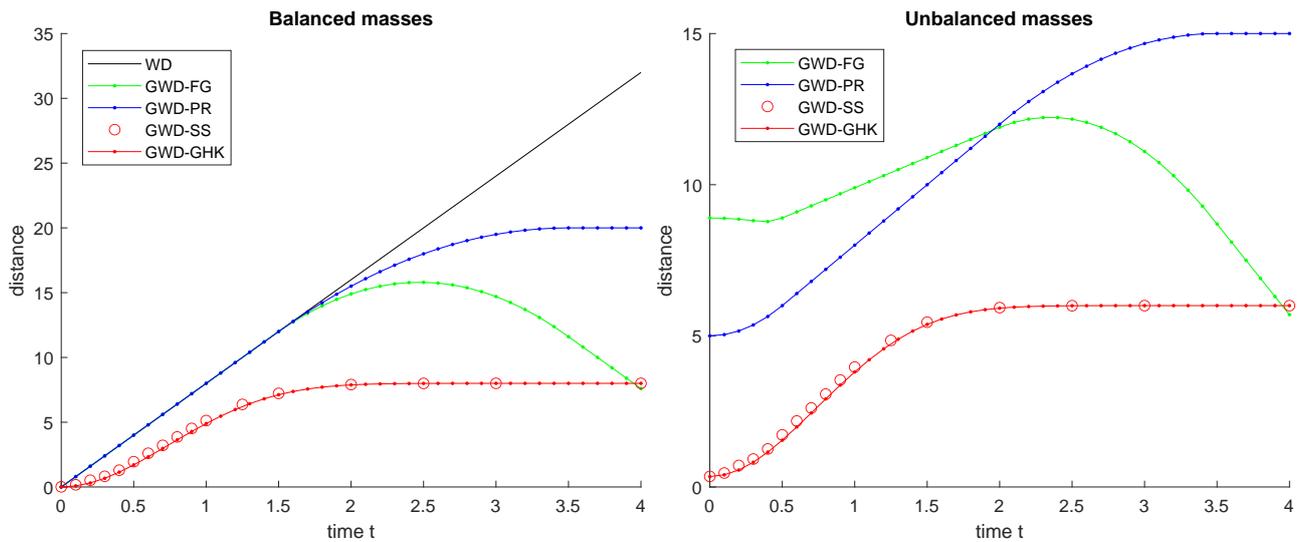
,

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

Lastly, FG has an important feature which deserves a discussion: the distance decreases whenever 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 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 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 [4] authors have already extensively studied the sensitivity of the LWR model [17, 23] using the WD. More precisely, many simulations were performed varying some key parameters like the initial vehicle density, the road capacity (maximal flux), the critical density (vehicle density associated to the maximal flux), the size of the road network, and the distribution coefficients at junctions (how traffic distributes across a junction with more than one outgoing road). Interestingly, 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 WD have 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 considered scenarios.

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

,

at a given time), and *velocity v*. Models can be *first-order* (i.e., velocity based), if 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 the road is a finite segment $X = [x_{\text{MIN}}, x_{\text{MAX}}]$ with a single lane and a single class of vehicles.

LWR model

The LWR model, introduced independently by Lighthill & Whitham [17] and by Richards [23], describes the evolution of the vehicle density $\rho(x,t)$ by means of the following IBVP 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, final time $T > t_0$, and concave flux function $f : [0, \rho_{\text{MAX}}] \rightarrow \mathbb{R}^+$ (called *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 $0 \leq \rho \leq \rho_{\text{MAX}}$ at any time, provided $0 \leq \rho_0, \rho_{\text{IN}}, \rho_{\text{OUT}} \leq \rho_{\text{MAX}}$.

Recalling the well known physical law $f = \rho v$, which is always valid for any x and t, we get 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 of 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, 2, 3 we will define

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

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 real-like 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

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

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 the Godunov's numerical flux defined by

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

,

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, e.g., if boundary data are given by fixed sensors counting vehicles passing through $x = x_{\text{MIN}}$ and $x = x_{\text{MAX}}$ in a time interval Δt. If this is the case, values $\rho_1^n = \rho_{\text{IN}}^n$ and $\rho_N^n = \rho_{\text{OUT}}^n$ are dropped, and incoming/outgoing fluxes F_{IN}^n, F_{OUT}^n are directly injected in the scheme (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.*

ARZ model

The ARZ model, introduced independently by Aw & Rascle [2] and by Zhang [29], describes the evolution of the vehicle density $\rho(x, t)$ and the velocity $v(x, t)$ by means of the following IBVP 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, an 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 $v_{eq}(\rho) := \frac{f(\rho)}{\rho}$, so that the desired equilibrium velocity of the ARZ model coincides with the instantaneous velocity of the LWR model. Also in the boundary conditions, we 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 boundary conditions. We consider the case described in Remark 6, which is a scenario often met in real life [5]. In the same spirit of [5], we assume that the time interval is divided in 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 a data because they are not yet available, but we assume to be able to (exactly) forecast the average value of them,

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

In short: we denote by ρ^s the "exact" solution obtained by using boundary fluxes $F_{\text{IN}}(t)$, $F_{\text{OUT}}(t)$ in the whole interval $[t_0, T]$, and by ρ^p the "predicted" solution obtained by using 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.

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 uniform distribution in $[0, \frac{1}{4}]$, F_{OUT} has uniform distribution in $[0, \mathcal{G}(0.2, 0.2)]$. 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
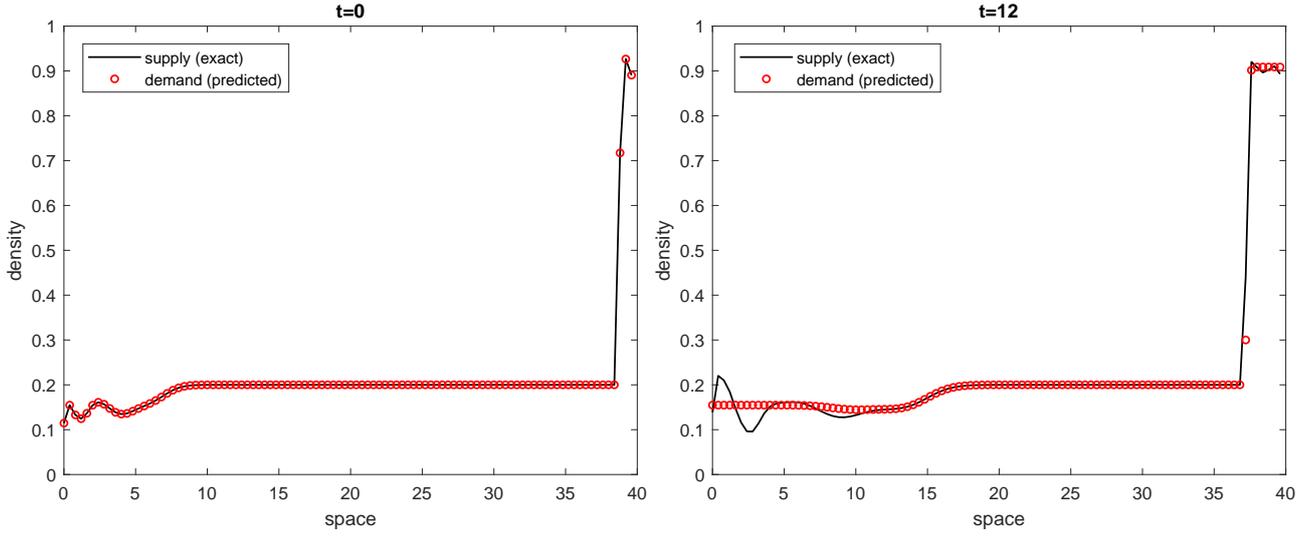
Figure 7. Test 1 on sensitivity to boundary conditions. Two screenshots at $t = 0$ ('now'), i.e. when solutions cease to be equal (left), and $t = 12$ (right).

the total mass on the road and the GWDs between $\rho^{\text{s}}(\cdot, t)$ and $\rho^{\text{D}}(\cdot, t)$ at any time.
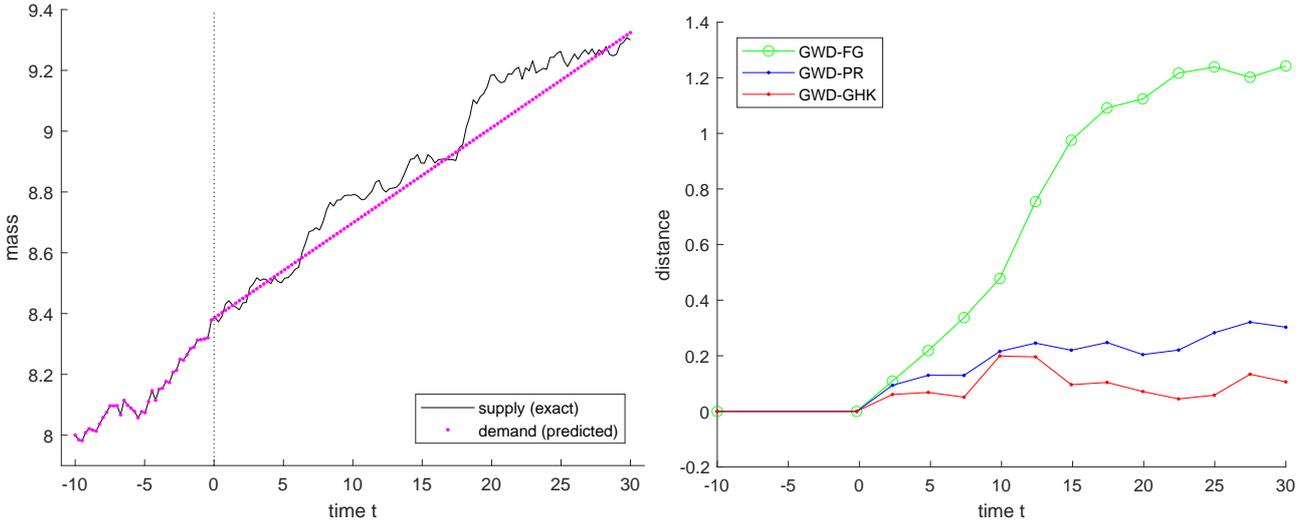
Figure 8. Test 1 on sensitivity to boundary conditions. Difference of mass between the two simulations (left) and comparison of GWDs (right).

Observing the numerical results, we tend to confirm what we have anticipated in Section 4.5, namely that the 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 the traffic prediction. This behavior of the error is clearly better caught by FG (see Figure 8(right)), which is the last distance that saturates.

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 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 at 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.

Initial conditions are $\rho^s(x, t_0) = \rho^D(x, t_0) = 0$ for any x. Left Dirichlet boundary conditions are $\rho^s_{\text{IN}} = \rho^D_{\text{IN}} = 0.4$. Right Dirichlet boundary conditions are $\rho^s_{\text{OUT}} = \rho^D_{\text{OUT}} = 0$. Other parameters are: $x_{\text{MIN}} = 0$, $x_{\text{MAX}} = 4$, $t_0 = 0$, $T = 20$, $\Delta x = 0.04$ (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 $t = 7.4$.
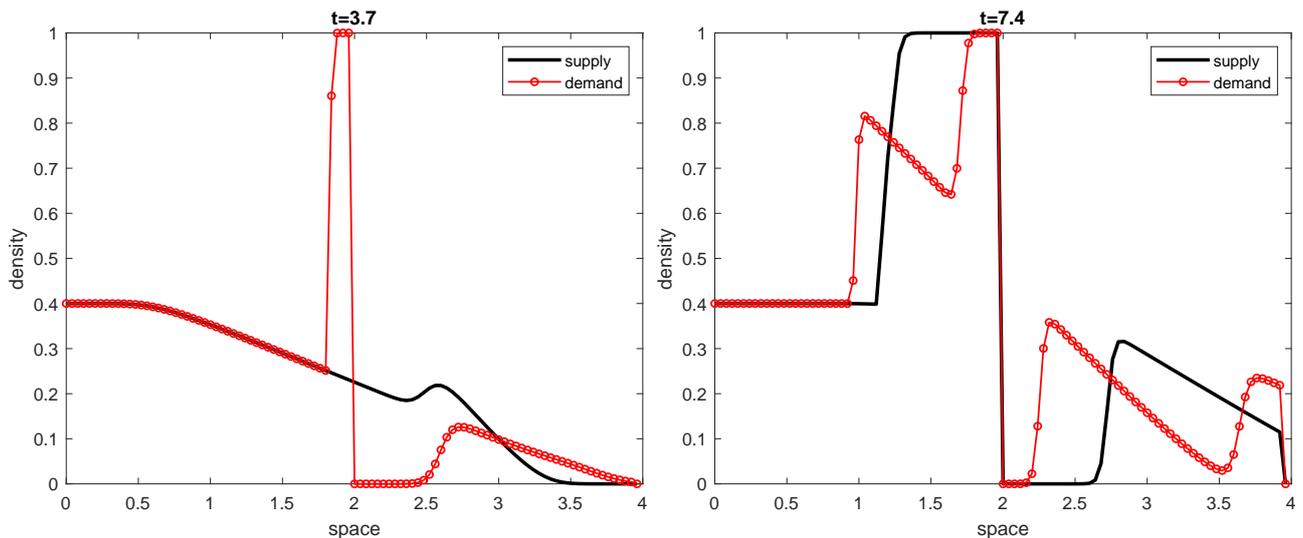
Figure 9. Test 2 on sensitivity to traffic light cycle. Two screenshots at $t = 3.7$ with red traffic light for demand and green traffic light for supply (left), and $t = 7.4$ with red traffic light for both distributions (right).

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

This test does not show much difference among the three GWDs, having them 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).
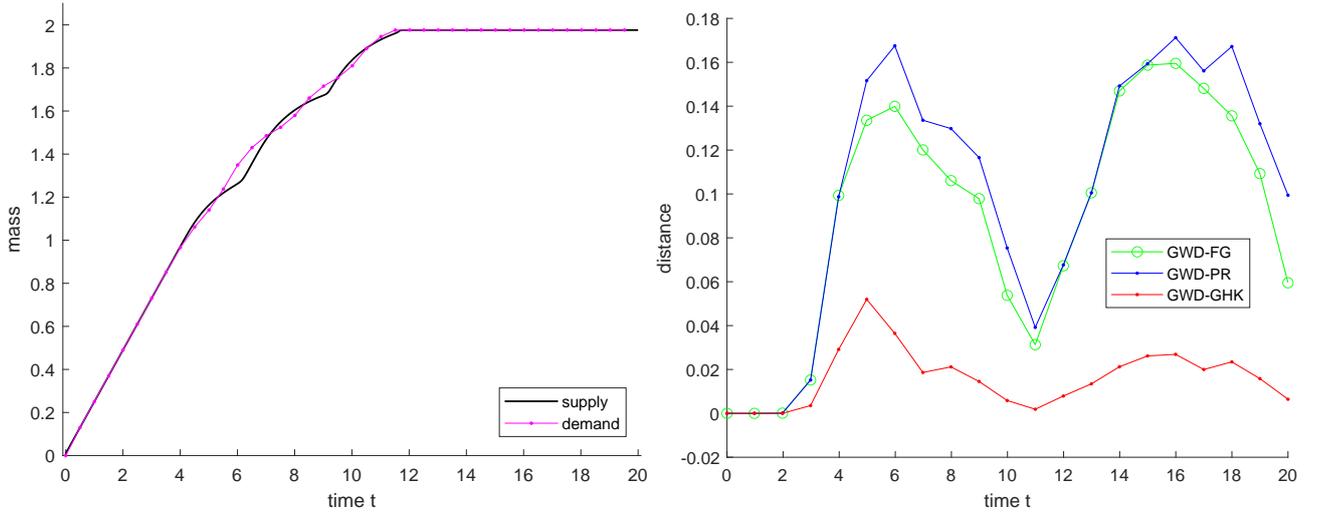
Figure 10. Test 2 on sensitivity to traffic light cycle. Difference of mass between the two simulations (correctly masses differ only in a certain time interval, left), and comparison of GWDs (right).

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

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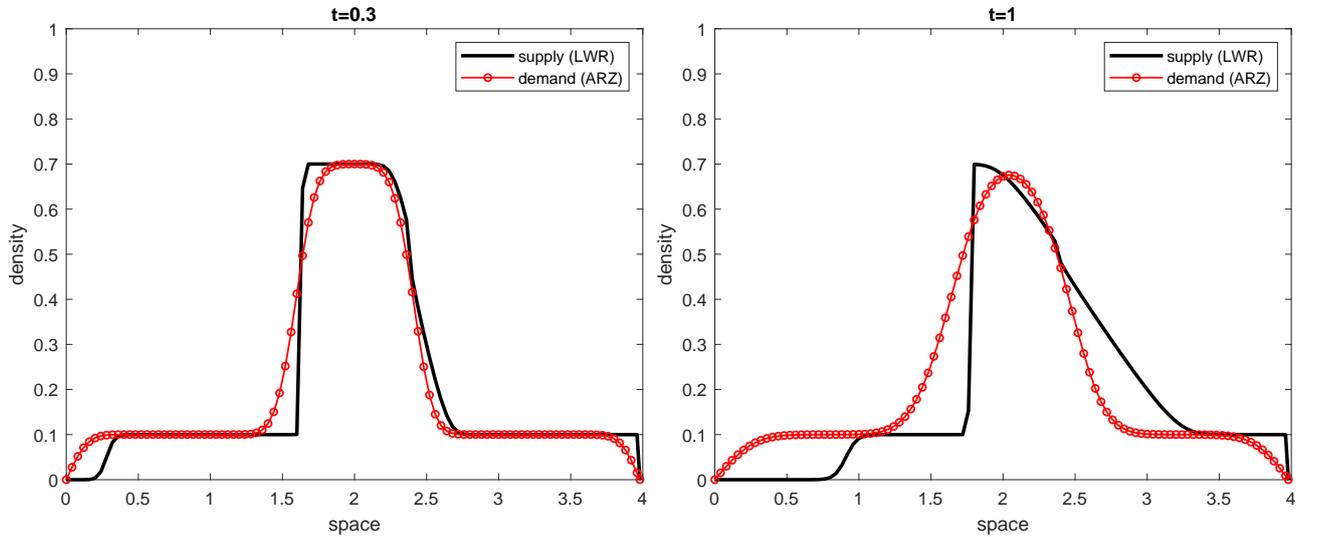
Figure 11. Test 3 on sensitivity to model order. Two screenshots at $t = 0.3$ (left), and $t = 1$ (right).

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

This Test is similar to Test 1, even if conclusions are less evident: here the two solutions start differing immediately, both at the boundary and inside the road. Moreover, intuition tells that the difference increases as the time goes by. Among the three GWDs, FG is the one which increases more
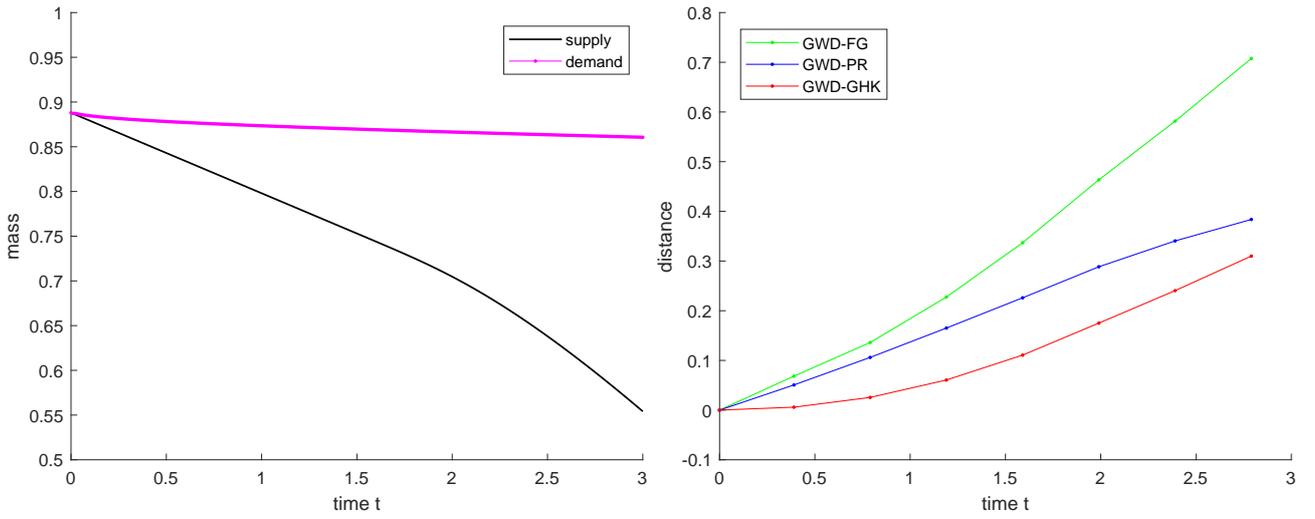
Figure 12. Test 3 on sensitivity to model order. Difference of mass between the two simulations (left) and comparison of GWDs (right).

rapidly at the initial time, and it is the one with the major relative increment at later times.

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 as much as possible some reference solution, typically coming from real observations (real traffic data). Once the calibration is done, one has to *validate* the model, checking if it is able to compute accurate solutions also 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 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 (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 (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 to measure several times the distance between the reference solution and the solution obtained with tentative parameters/conditions, until the minimum distance (best match) is found.

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 $[100, 200]$ veh/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 FG, PR, and GHK approaches.

It is interesting to note that in all the three cases the optimization problem is able to find quite precisely the optimal pair $(\rho_{\text{MAX}}^*, v_{\text{MAX}}^*)$, 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.

Test 4b   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, $x_{\text{MAX}} = 2$ km, $T = 3$ min, $\rho_{\text{IN}}^* = 40$ veh/km, $\rho_{\text{OUT}}^* = 0$ veh/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 unknown too (this is a typical situation which can happen during model validation). The intervals for the search are $[0, \rho_{\text{MAX}}/2]$ veh/km for the inflow boundary conditions and $[0, \rho_{\text{MAX}}]$ km/h for the outflow boundary conditions.

Conversely to the 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 at all considered in the inverse problem. Indeed, the optimization problem could tend to compensate 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 FG, PR, and GHK approaches.

Since the accident happens close to the boundary, and control parameters $\rho_{\text{IN}}, \rho_{\text{OUT}}$ act at boundaries as well, FG struggles to find the optimal parameters, being basically insensitive to any possible variation. This is evident 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 boundaries and are able to locate the exact solution even if the accident is ignored.

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) results to be numerically unfeasible, while the other three (FG, PR, GHK) are all relative simple to implement once the computational procedure is arranged.
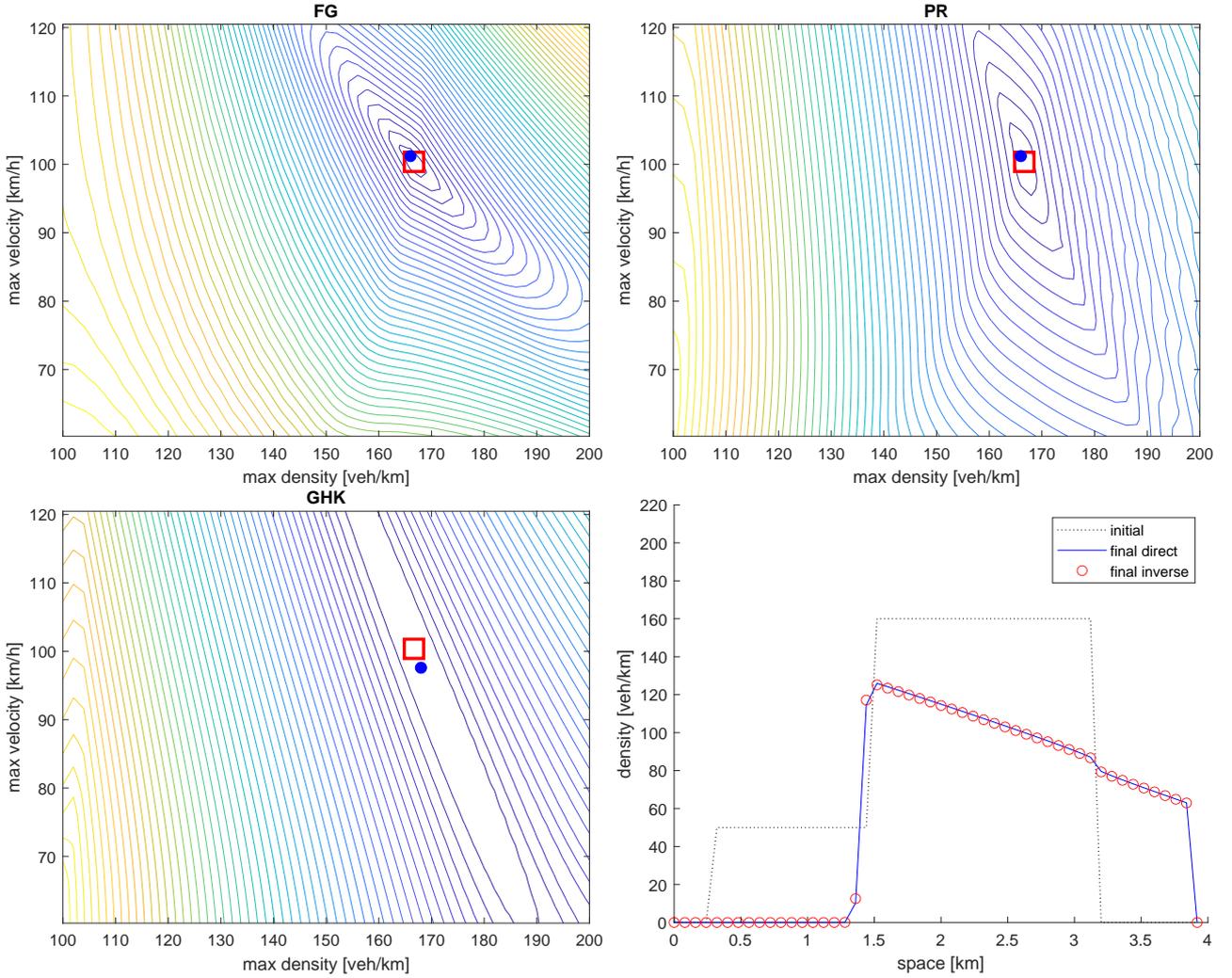
Figure 13. Test 4a on calibration. Level sets of the GWD functions $(\rho_{\text{MAX}}, v_{\text{MAX}}) \rightarrow \tilde{W}(\rho(\cdot, T; \rho_{\text{MAX}}, v_{\text{MAX}}), \rho_T^{\text{REF}})$ for FG (top-left), PR (top-right), and GHK (bottom-left). The red square locates the exact pair $(\rho_{\text{MAX}}^*, v_{\text{MAX}}^*)$, 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 final time corresponding to the solution to the optimization problem.

Regarding traffic modeling, the different way the road boundaries are treated makes the choice of the GWD lean towards 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 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 boundaries. This feature is useful since traffic forecast at boundaries is necessarily inaccurate due to the impossibility to estimate inflow and outflow traffic with good precision.

In the future, one could exploring the possibility to compute the considered GWDs in more complicated domain like $\mathbb{R}^2$, $\mathbb{R}^3$, and (road) networks.
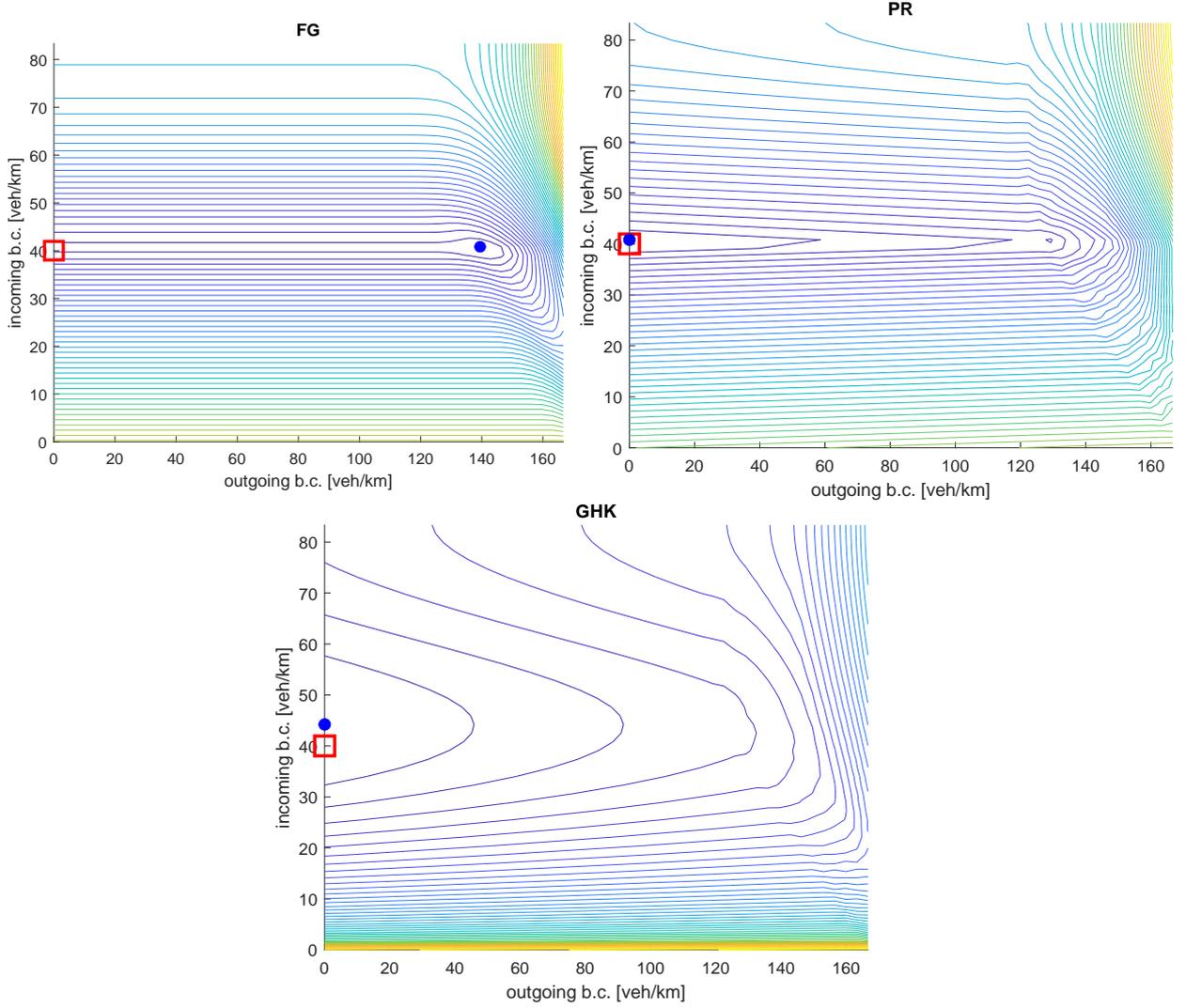
Figure 14. Test 4b on validation. Level sets of the GWD functions $(\rho_{\text{OUT}}, \rho_{\text{IN}}) \rightarrow \tilde{W}(\rho(\cdot, T; \rho_{\text{IN}}, \rho_{\text{OUT}}), \rho_T^{\text{REF}})$ for FG (top-left), PR (top-right), and GHK (bottom). The red square locates the exact pair $(\rho_{\text{OUT}}^*, \rho_{\text{IN}}^*)$, while the blue dot locates the argmin of the function.

Funding

Acknowledgments

Use of AI tools declaration

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

Conflict of interest

The authors declare there is no conflict of interest.

A. Appendices

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 $\Delta x = 0.1$. Total supply mass is 0.2, divided in two equal parts and concentrated in two points. Total demand mass is also 0.2, concentrated in one point only.

1.1. Here supply mass is located at $x = 1$ and $x = 2.5$, while demand mass is located at $x = 3$, see Figure 15(left). Optimal mass transport map prescribes to move supply mass onto demand mass, without resorting to mass at boundaries. We have $\tilde{W}_1^{\mathrm{FG}} = 0.25$.
1.2. Here supply mass is located at $x = 0.5$ and $x = 2.5$, while demand mass is located at $x = 4$, see Figure 15(right). Optimal mass transport map prescribes to move leftmost supply mass to the left boundary, rightmost supply mass onto the demand mass, and taking mass from right boundaries to bring them onto the demand mass. We have $\tilde{W}_1^{\mathrm{FG}} = 0.3$.
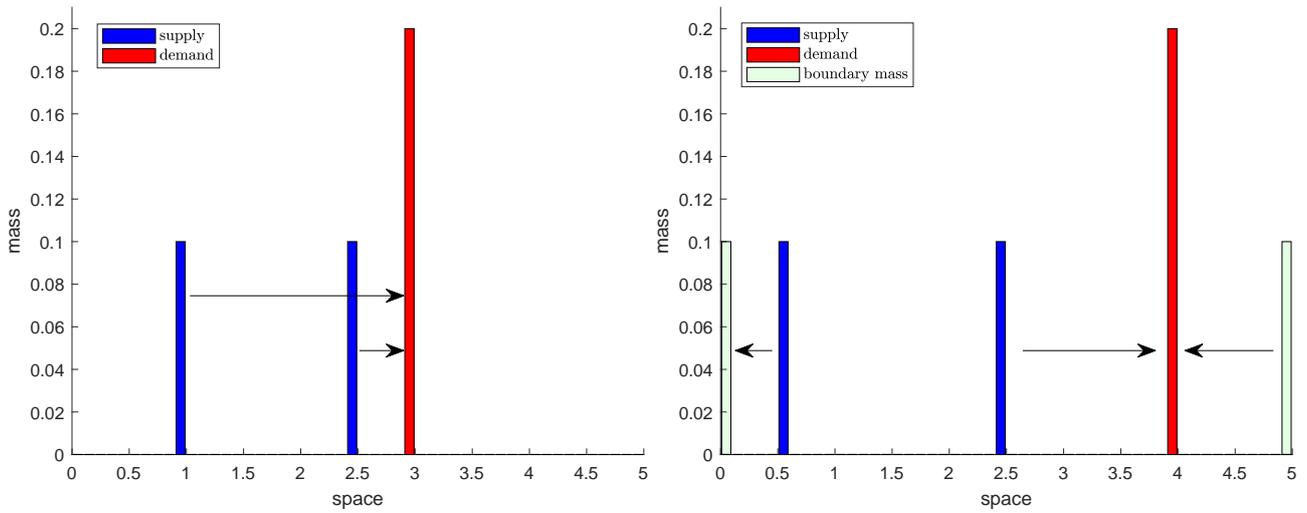
Figure 15. Test 1. Case 1.1 where boundaries are not used (left), and Case 1.2 where mass is both taken from and brought to boundaries (right).

Test 2

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

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

see Figure 16. We set $N = 50$ so that $\Delta x = 0.2$.
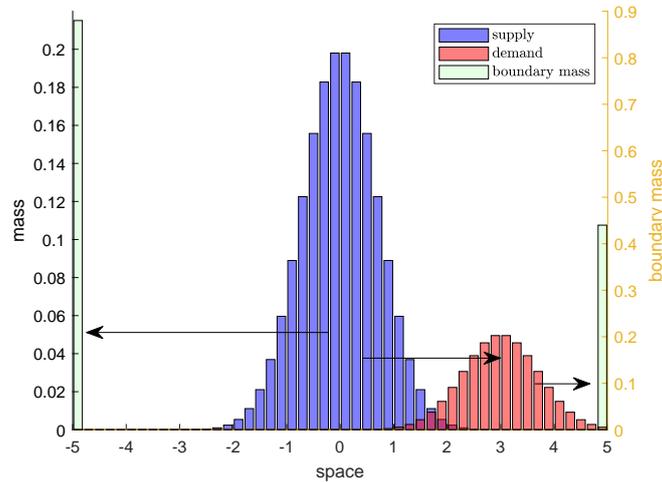

Figure 16. Test 2. Some supply (resp., demand) mass is brought to left (resp., right) boundary, the rest is transported.

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

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

Test 1

Consider the simple example proposed in [21, Section 2.2], where 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 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 17 shows the result of the numerical computation in the domain $X = [-2, 5]$ divided in $N = 200$ cells, for $\xi \in [0, 3]$.
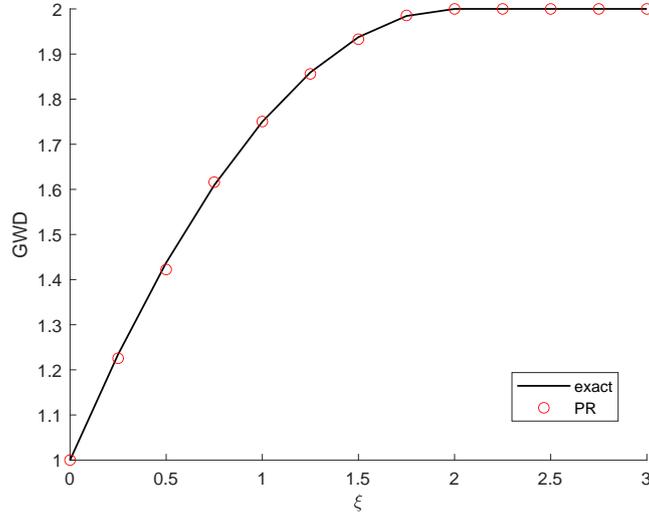

Figure 17. Test 1. Exact solution compared with numerical solution ($N = 200$).

Test 2

In this test we run the algorithm in a nontrivial case where exact solution is not known. More in 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 18 shows the result in terms of remaining mass after minimization. One can see that here the optimal solution ($\tilde{W}_1^{\mathrm{PR}} = 4.36$) requires to destroy only a part of the mass.

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

We present here numerical tests for both 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.
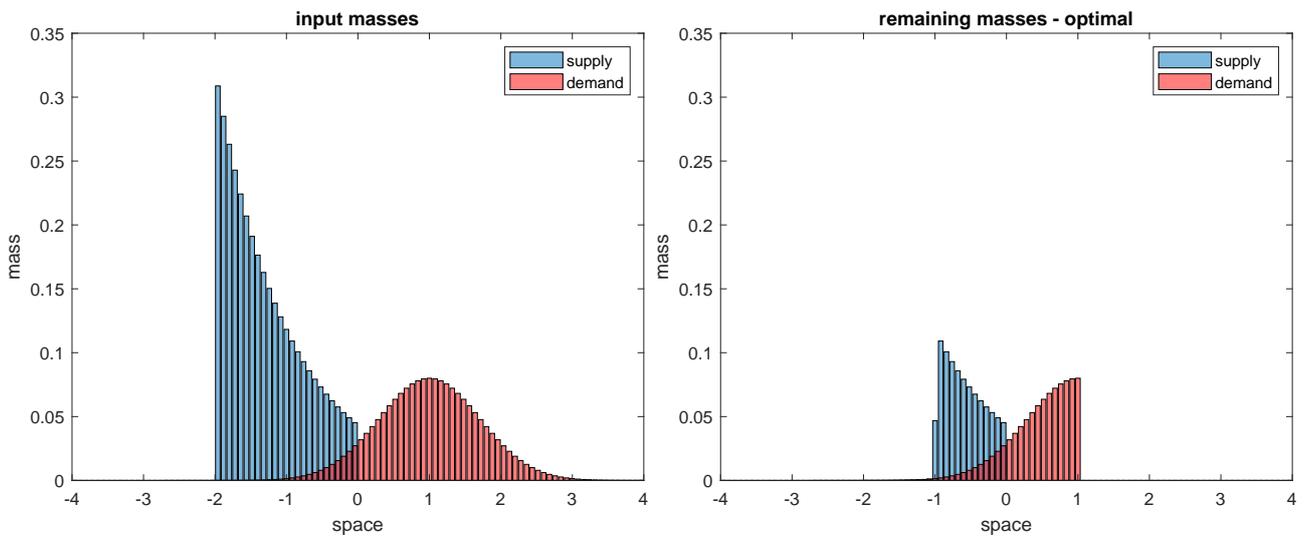
Figure 18. Test 2. Initial distributions m^{s} and m^{D} (left) and remaining masses $\mathring{m}^{\mathrm{s}}$ and $\mathring{m}^{\mathrm{D}}$ corresponding to the optimal solution (right).

Test 1

This test is mainly conceived 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 (3.9).

Figure 19 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.

More in detail, Figure 19(left) shows the GWD in the case 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 $\tilde{W}_2^{\mathrm{GHK}}$, that 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, 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 (3.9), the objective in (4.6) decays exponentially to $r_{jh}^{\mathrm{s}} + r_{k\ell}^{\mathrm{D}}$ which, in turn, equals at most $2\bar{m}$ due to the constraints on the homogeneous marginals.

Figure 19(right) shows instead the GWD in the case the two distributions are located at the same point and the difference of their masses progressively increases from 0 to 7.

Test 2

The goal of this test is mainly comparing 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 20. We assume that supply masses are located at $x = 0$ and $x = 1$, have masses $m = 2$ and $m = 1$, and have $n = 2$, $n = 1$ additional divisions, respectively. Demand mass is instead located at $x = 3$, has mass $m = 4$, and has $n = 4$ additional divisions. In this scenario it results $d = 12$.

Figure 21(top) shows the history of all the evaluations of $\tilde{W}_2^{\mathrm{SS}}$ done by the exhaustive algorithm
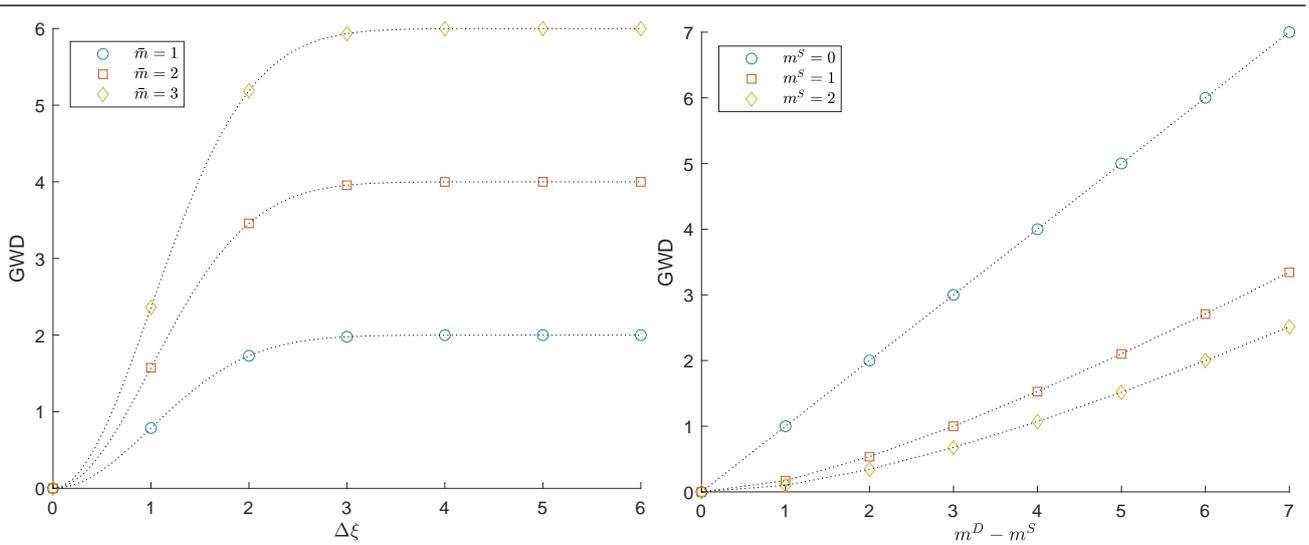
Figure 19. Test 1. GWD between two equal atomic mass distributions with mass $\bar{m}$ as a function of their distance $\Delta\xi$ (left), and GWD between two overlapping atomic mass distributions as a function of the difference of mass (right). Dotted black line is the exact solution.
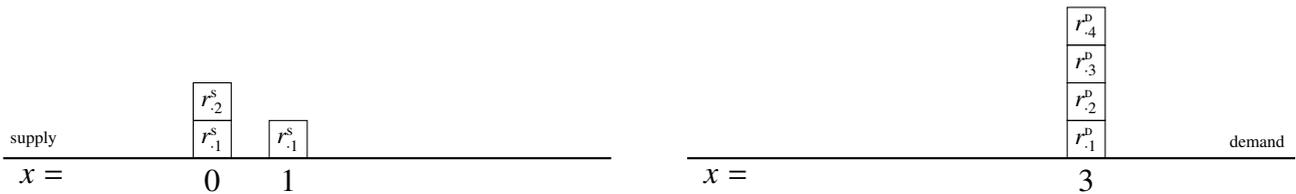

Figure 20. Test 2: two concentrated masses for supply mass and one for demand mass.

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). Minimum is already found among the first evaluations and it is $\tilde{W}_2^{\text{ss}} = 6.45$.

Figure 21(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 descend, with $Q = 1,000$. The algorithm returns again $\tilde{W}_2^{\text{ss}} = 6.45$. 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, in the following we will use algorithm SS-B only, which is obviously faster.

References

1. L. Ambrosio, Mainini E., and Serfaty S. Gradient flow of the chapman-rubinstein-schatzman model for signed vortices. *Ann. Inst. H. Poincaré C Anal. Non Linéaire*, 28(2), 2011.

2. A. Aw and M. Rascle. Resurrection of "second order" models of traffic flow? *SIAM J. Appl. Math.*, 60:916–938, 2000.

3. J.-D. Benamou and Y. Brenier. A computational fluid mechanics solutions to the Monge–Kantoriovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.
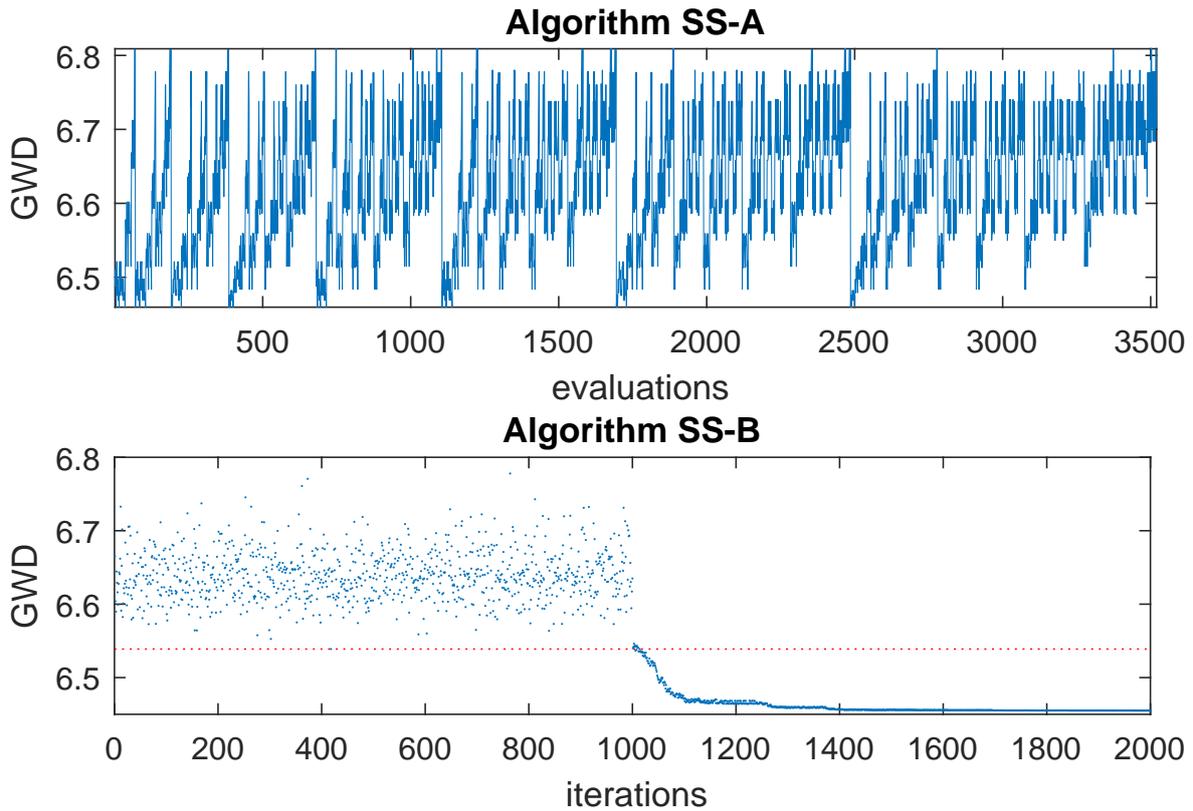
Figure 21. Test 2. Algorithm SS-A, complete view of all the evaluations of $\tilde{W}_2^{ss}$ (top), and Algorithm SS-B, 1,000 random searches followed by about 1,000 descend iterations (bottom). Dotted red line indicates the minimum value found in the first phase of SS-B.

4. M. Briani, E. Cristiani, and E. Iacomini. Sensitivity analysis of the LWR model for traffic forecast on large networks using Wasserstein distance. *Commun. Math. Sci.*, 16(1):123–144, 2018.

5. M. Briani, E. Cristiani, and E. Onofri. Inverting the fundamental diagram and forecasting boundary conditions: How Machine Learning can improve macroscopic models for traffic flow. To appear on Advances in Computational Mathematics. Preprint arXiv:2303.12740.

6. L. Caffarelli and R. J. McCann. Free boundaries in optimal transport and Monge-Ampère obstacle problems. *Annals of Mathematics*, 171(2):673–730, 2010.

7. L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. An interpolating distance between optimal transport and Fisher-Rao metrics. *Foundation of Computational Mathematics*, 18:1–44, 2018.

8. L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Unbalanced optimal transport: Dynamic and Kantorovich formulations. *Journal of Functional Analysis*, 274:3090–3123, 2018.

9. E. Cristiani, B. Piccoli, and A. Tosin. *Multiscale modeling of pedestrian dynamics*, volume 12 of *Series Modeling, Simulation & Applications*. Springer, 2014.

10. E. Cristiani and M. C. Saladino. Comparing comparisons between vehicular traffic states in microscopic and macroscopic first-order models. *Mathematical Methods in the Applied Sciences*, 42(3):918–934, 2019.

11. A. Figalli. The optimal partial transport problem. *Archive for Rational Mechanics and Analysis*, 195(2):533–560, 2010.

12. A. Figalli and N. Gigli. A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions. *Journal de Mathématiques Pures et Appliquées*, 94(2):107–130, 2010.

13. L. G. Hanin. Kantorovich-Rubinstein norm and its application in the theory of Lipschitz spaces. *Proceedings of the American Mathematical Society*, 115(2):345–352, 1992.

14. L. Kantorovich and G. S. Rubinstein. On a space of totally additive functions. *Vestnik Leningradskogo Universiteta*, 13:52–59, 1958.

15. S. Kondratyev, L. Monsaingeon, and D. Vortnikov. A new optimal transport distance on the space of finite Radon measures. *Adv. Differential Equations*, 21(11/12):1117–1164, 2016.

16. M. Liero, A. Mielke, and G. Savaré. Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures. *Invent. Math.*, 211(3):969–1117, 2018.

17. M. J. Lighthill and G. B. Whitham. On kinematic waves II. A theory of traffic flow on long crowded roads. *Proc. R. Soc. Lond. Ser. A*, 229:317–345, 1955.

18. L. Lombardini and F. Rossi. Obstructions to extension of Wasserstein distances for variable masses. *Proceedings of the American Mathematical Society*, 150(11):4879–4890, 2022.

19. Z. Ma, X. Wei, X. Hong, H. Lin, Y. Qiu, and Y. Gong. Learning to count via unbalanced optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 2319–2327, 2021.

20. E. Mainini. A description of transport cost for signed measures. *J. Math. Sci. (N.Y.)*, 181(6), 2012.

21. B. Piccoli and F. Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Archive for Rational Mechanics and Analysis*, 211:335–358, 2014.

22. B. Piccoli and F. Rossi. On properties of the generalized Wasserstein distance. *Arch. Rational Mech. Anal.*, 222:1339–1365, 2016.

23. P. I. Richards. Shock waves on the highway. *Oper. Res.*, 4:42–51, 1956.

24. F. Santambrogio. *Optimal Transport for Applied Mathematicians. Calculus of Variations, PDEs, and Modeling*, volume 87 of *Series Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, 2015.

25. G. Savaré and G. E. Sodini. A relaxation viewpoint to Unbalanced Optimal Transport: Duality, optimality and Monge formulation. *Journal de Mathématiques Pures et Appliquées*, 188:114–178, 2024.

26. S. M. Sinha. *Mathematical Programming. Theory and Methods*. Elsevier, 2006.

27. C. Villani. *Topics in Optimal Transportation*, volume 58 of *Series Graduate Studies in Mathematics*. American Mathematical Society, 2003.

28. C. Villani. *Optimal Transport. Old and New*, volume 338 of *Series Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

29. H. M. Zhang. A non-equilibrium traffic model devoid of gas-like behavior. *Transportation Res. Part B*, 36:275–290, 2002.