Unbalanced L1 optimal transport for vector valued measures and application to Full Waveform Inversion

Gabriele Todeschi, Ludovic Métivier, Jean-Marie Mirebeau

UNBALANCED L1 OPTIMAL TRANSPORT FOR VECTOR VALUED MEASURES AND APPLICATION TO FULL WAVEFORM INVERSION

GABRIELE TODESCHI, LUDOVIC MÉTIVIER, AND JEAN-MARIE MIREBEAU

ABSTRACT. Optimal transport has recently started to be successfully employed to define misfit or loss functions in inverse problems. However, it is a problem intrinsically defined for positive (probability) measures and therefore strategies are needed for its applications in more general settings of interest. In this paper we introduce an unbalanced optimal transport problem for vector valued measures starting from the L^1 optimal transport. By lifting data in a self-dual cone of a higher dimensional vector space, we show that one can recover a meaningful transport problem. We show that the favorable computational complexity of the L^1 problem, an advantage compared to other formulations of optimal transport, is inherited by our vector extension. We consider both a one-homogeneous and a two-homogeneous penalization for the imbalance of mass, the latter being potentially relevant for applications to physics based problems. In particular, we demonstrate the potential of our strategy for full waveform inversion, an inverse problem for high resolution seismic imaging.

Keywords: Optimal transport, vector valued measures, inverse problems, seismic imaging, proximal splitting

1. INTRODUCTION

Optimal transport is the problem of finding the optimal way of reallocating distributions of mass. Among its many interesting features, it allows to define, in a natural and meaningful way, distances between probability measures. For this reason, it has started recently to be considered for applications to inverse problems to propose misfits between real and synthetic data. Inverse problems are usually highly non-linear and non-convex variational problems and a fundamental role is played by the misfit or loss function. The use of optimal transport based distances has revealed to be helpful in attenuating the lack of well-posedness of the problem and increasing the robustness of the inversion approach. Full Waveform Inversion (FWI), a seismic inversion problem, is one particular example, with a lot of recent literature (see for example the review paper [22]).

1.1. **Optimal transport.** Consider $\Omega \subset \mathbb{R}^d$, compact and convex with non-empty interior, and two positive probability measures $\mu, \nu \in \mathcal{P}(\Omega)$, that is non-negative measures with the same total mass, $\int_\Omega \mu = \int_\Omega \nu$. Given a scalar non-negative real valued continuous function $c : \Omega \times \Omega \to \mathbb{R}_+$, the optimal transport problem between μ and ν is

$$(1.1) \qquad \mathrm{OT}_c(\mu, \nu) := \min_{\gamma \in \Pi(\mu, \nu)} \int_{\Omega \times \Omega} c(x, y) \mathrm{d}\gamma,$$

where the set of admissible transport plans γ is defined as

$$(1.2) \qquad \Pi(\mu, \nu) = \left\{ \gamma \in \mathcal{P}(\Omega \times \Omega), (\pi_1)_{\#}\gamma = \mu, (\pi_2)_{\#}\gamma = \nu \right\},$$

i.e. the subset of probability measures on the product space $\Omega \times \Omega$ that have first and second marginals μ and ν. The ground cost $c(x, y)$ is the cost for transporting one unit of mass from

x to y. An interesting case is provided by the choice $c(x, y) = |x - y|^p$. In this case (1.1) is called L^p optimal transport problem and can be shown to provide a (p-th power of a) distance on the space of probability measures which is referred to as the p-Wasserstein distance, W_p. This distance translates in a natural way to the measure setting properties of the ground cost function. Importantly, it characterizes weak convergence of measures. We suggest for example the monographs [36, 39] for an introduction to the subject.

In this paper we deal with the specific case $p = 1$, namely the L^1 optimal transport problem. In this case problem (1.1) enjoys three equivalent formulations:

$$(1.3) \qquad W_1(\mu, \nu) := \min_{\gamma \in \Pi(\mu, \nu)} \int_\Omega |y - x| \mathrm{d}\gamma$$

$$(1.4) \qquad = \max_{\phi \in \mathcal{C}(\Omega)} \left\{ \int_\Omega \phi(\mu - \nu), \ |\nabla \phi|_\infty \le 1 \right\}$$

$$(1.5) \qquad = \min_{\sigma \in \mathcal{M}(\Omega; \mathbb{R}^d)} \left\{ \int_\Omega |\sigma| , \ \mathrm{div}(\sigma) = \mu - \nu \right\}.$$

The maximization problem (1.4) is known as the dual formulation of the L^1 transport whereas (1.5) is known as Beckmann's minimal flow problem. These two formulations are particularly interesting for at least two reasons. Firstly, they appear as simpler problems as they present only local constraints differently from (1.3). In particular, notice that given a solution γ it is possible to recover solutions σ and ϕ (see [36, Chapter 4]) but the opposite is not true. For this reason they are much more amenable for numerical computations. Furthermore, they can be generalized straightforwardly to more general signed measures, contrary to the original form of the problem.

1.2. Applications to inverse problems.

Two major issues arise for the application of optimal transport to inverse problems. The first one is the computational cost. Problem (1.1) is evidently more expensive than evaluating simple L^p distances which are usually considered. However, there has been a significant improvement on this side in the past years (see for example [30, 20]), which mostly contributed to the recent development of optimal transport. We highlight also that for the specific applications to inverse problems computing an optimizer γ is not strictly necessary: the only requirement is to be able to evaluate the misfit function, i.e. compute the transport cost, and compute its gradient with respect to the measures. This may allow to devise a cheaper numerical strategy. In this regard, the L^1 problem offers a cheaper alternative as previously remarked.

The second issue is that problem (1.1) is defined for probability measures, i.e. non-negative measures with fixed total mass. Generalizations have been proposed to deal with non-negative Radon measures [31, 12]. However, inverse problems may have to deal in general with signed data which therefore need to be handled properly. Trivial transformation such as transporting separately positive and negative parts [17], taking the exponential or powers of the data [34], adding positive mass everywhere [41], may not result effective. These transformations tend to destroy the structure of the data. Two effective and general ideas have been proposed in [1, 21].

In [1] the authors started from the dual formulation (1.4) of the L^1 problem to propose an optimal transport misfit for signed data. As previously remarked, this optimal transport problem has the advantage that it can be extended directly. They considered then the misfit

provided by:

$$(1.6) \qquad \mathrm{KR}(\mu, \nu) = \max_{\phi \in \mathcal{C}(\Omega)} \left\{ \int_{\Omega} \phi(\mu - \nu), \ |\nabla \phi|_{\infty} \leq 1, \ |\phi|_{\infty} \leq \lambda \right\}$$

for some parameter $\lambda \in \mathbb{R}_{+}$. The L^{∞} bound on the potential is added in order to be able to deal with measures that do not have the same total mass. Problem (1.6) is known as the Kantorovich-Rubenstein norm, or bounded-Lipschitz metric, and is a distance on the space of signed measures. This misfit function provides good results and the approach is also computationally appealing, since the problem can be solved effectively up to high dimensions. However, as problem (1.6) is not framed on positive measures it does not ensure sensitivity to horizontal shifts of the data (see Example 3.7). The growth of the optimal transport cost with respect to shifts of the measures is a key feature for its application to inverse problems.

In [21], the authors proposed to lift signed data to a higher dimensional space, where it can be regarded as a positive measure on its trace. Then, the positive lifted measures can be discretized and transported according to (1.1). Since the value assigned to the measure is arbitrary, the ground cost $c(x, y)$ needs to be tuned properly in order to encode transport along the time direction and not along the amplitude one. The approach in the end is effective and leads to a robust inversion problem. On the other hand, it is computationally expensive and not suited to scale to high dimensions.

1.3. **Contribution and structure of the paper.** In this paper we propose a generalization of the L^1 optimal transport problem (1.3) to vector valued measures. Assuming to be able to lift, without loss of information, signed data into a (self-dual) cone C of a general vector space V, this generalization can serve as a meaningful misfit function in inverse problems. In this sense, our approach can be seen as an extension of the Kantorovich-Rubenstein approach (1.6) proposed in [1], aiming at recovering a transport problem between positive measures. By extending the L^1 transport we take advantage of the lower computational complexity this problem offers. We consider as a specific application the FWI problem, where one can consider multi-component signals as data and naturally lift them in the space of Symmetric Positive Definite (SPD) valued measures. Some good preliminary results we obtained indicate the effectiveness of the approach.

Extensions of the optimal transport problem to vector valued measures (and more specifically SPD valued measures) have already been proposed recently, starting from the seminal work [26], where the authors proposed a generalization of the formulation (1.1) inspired by quantum mechanics arguments. Later on, several other works appeared, considering either a static formulation [29] or extensions of the dynamical formulation for the L^2 transport problem [10, 7, 24, 4, 11]. Dynamical formulations of optimal transport are not suited for applications to inverse problems as they demand a higher computational cost to fully compute trajectories. On the other hand, the entropic regularization approach can be extended to this setting as shown in [29]. However, the blurring effect of entropic regularization may require low values of the temperature parameter to provide meaningful descent directions in the inverse problem, vanishing the favorable computational complexity. The most natural way of extending optimal transport to a vector setting is probably to consider the dual formulation (1.4) (or the Beckmann's one (1.5)) of the L^1 transport. This has been originally done in [25, 9, 35], where the authors proposed an extension of L^1 transport to the setting of SPD valued measures inspired again from quantum mechanics. More recently, in [15, 14] a similar extension has been considered for functional analysis applications.

After setting our notation in Section 2, we present the unbalanced L^1 optimal transport problem for vector valued measures in Section 3. We first introduce the straightforward

extension of the dual (1.4) and the Beckmann's (1.5) formulations, in the balanced setting. These coincide with the ones already considered in [14]. Differently from these and the other works we mentioned just above, we consider however a general setting and justify our results via convex analysis arguments. In order to deal with general measures, we propose an unbalanced problem similarly to the (scalar) unbalanced L^1 problem introduced in [31]. We show its equivalence to the vector extension of the Kantorovich-Rubenstein norm (1.6) (analogously to what has been done in [32] in the scalar case) and also to a relaxed version of the Beckmann's formulation. We then consider a different way of penalizing transport in the unbalanced model, namely an L^2 penalization instead of an L^1, which can be beneficial for applications to inverse problems. In Section 4 we present our finite difference discretization and show that the use of proximal-splitting algorithms, an efficient strategy to solve problem (1.6) in the scalar setting, can be extended to the vector setting with a simple linear increase in the computational complexity. Finally, in Section 5 we present some numerical results and the application to the inverse problem of FWI.

2. Notations

The standard Euclidean scalar product and norm are denoted by $x \cdot y$ and $|\cdot|$, for $x, y \in \mathbb{R}^d$, where d is the dimension of the compact optimal transport domain $\Omega \subset \mathbb{R}^d$. We denote by V another real vector space of dimension n endowed with its scalar product $\langle \cdot, \cdot \rangle_V$ and the respective norm $|\cdot|_V$. In V we consider a convex cone C. We recall that the dual cone of C is defined as

$$C^* = \{u \in V : \langle u, v \rangle_V \geq 0, \ \forall v \in V\},$$

and we restrict to considering cones such that $C \subseteq C^*$. In particular, this implies that $\langle v_1, v_2 \rangle_V \geq 0, \ \forall v_1, v_2 \in C$. Possible examples are $V = \mathbb{R}^n$ and C equal to the non-negative orthant $[\mathbb{R}_+]^n$, or $V = \mathrm{S}^{n'}$ the space of symmetric matrices and $C = \mathrm{S}_+^{n'}$ the subspace of positive semi-definite matrices (where $n = n'(n'+1)/2$). For these two cases in particular it holds $C = C^*$ and the cones are called self-dual. We further introduce the vector space V^d endowed with the scalar product $\langle \cdot, \cdot \rangle_{V^d} = \sum_{l=1}^{d} \langle \cdot, \cdot \rangle_V$ and the respective norm $|\cdot|_{V^d}$. One has a natural isomorphism of vector spaces $V^d \simeq L(\mathbb{R}^d; V)$, and as a result V^d is also equipped with the operator norm $|\cdot|_{L(\mathbb{R}^d;V)}$. The two norms are comparable, namely

$$(2.1) \qquad |\cdot|_{L(\mathbb{R}^d;V)} \leq |\cdot|_{V^d} \leq \sqrt{d} |\cdot|_{L(\mathbb{R}^d;V)},$$

which is easily established using the singular value decomposition. However in dimension $d > 1$ these norms are distinct, and $|\cdot|_{L(\mathbb{R}^d;V)}$ is not related to an inner product. For $A, B \in L(\mathbb{R}^d; V)$, $\langle A, B \rangle_{V^d} = \mathrm{Tr}(A^\top B)$ is known as the Frobenius inner product.

We denote $\mathcal{C}(\Omega; V)$ and $\mathcal{C}(\Omega; V^d)$ the spaces of continuous functions defined on the compact domain $\Omega \subset \mathbb{R}^d$ and valued in V and V^d. These spaces are endowed respectively with the sup norms

$$|\phi|_{V,\infty} = \sup_{x \in \Omega} |\phi(x)|_V, \quad |\psi|_{V^d,\infty} = \sup_{x \in \Omega} |\psi(x)|_{V^d}, \quad \text{for } \phi \in \mathcal{C}(\Omega; V), \ \psi \in \mathcal{C}(\Omega; V^d).$$

The space of finite vectorial measures valued in V is $\mathcal{M}(\Omega; V)$ and is the dual space of $\mathcal{C}(\Omega; V)$ via the duality product $(\phi, \mu) \mapsto \int_\Omega \langle \phi, \mu \rangle_V$. The convex subspace of measures which take value in the cone C is $\mathcal{M}(\Omega; C)$. A measure $\mu \in \mathcal{M}(\Omega; V)$ belongs to $\mathcal{M}(\Omega; C)$ if and only if $\int_\Omega \langle \varphi, \mu \rangle_V \geq 0$ for all $\varphi \in C^0(\Omega; C^*)$. To any measure μ we can associate a positive

scalar measure $|\mu|_V \in \mathcal{M}(\Omega; \mathbb{R}_+)$ defined as

$$(2.2) \qquad |\mu|_V(K) = \sup_{\substack{\varphi \in \mathcal{C}(K;V) \\ |\phi|_{V,\infty} \leq 1}} \int_K \langle \varphi, \mu \rangle_V, \quad \text{for all Lebesgue measurable subsets } K \subset \Omega.$$

The map

$$\mu \in \mathcal{M}(\Omega; V) \longmapsto |\mu|_{\mathcal{M}(\Omega;V)} := |\mu|_V(\Omega)$$

is a norm on $\mathcal{M}(\Omega; V)$ and is called the total variation norm. On $\mathcal{M}(\Omega; V)$ we consider also the weak-* topology. Importantly, closed balls of $\mathcal{M}(\Omega; V)$ with respect to the strong topology are weakly-* compact. Furthermore, from definition (2.2), the total variation norm is lower semi-continuous with respect to the weak-* convergence. Finally, we define the space of finite vectorial measures valued in V^d, $\mathcal{M}(\Omega; V^d)$, in duality with $\mathcal{C}(\Omega; V^d)$ through $(\psi, \sigma) \mapsto \int_\Omega \langle \psi, \sigma \rangle_{V^d}$. The space $\mathcal{M}(\Omega; V^d)$ endowed with the total variation norm $|\sigma|_{\mathcal{M}(\Omega;V^d)}$, defined in the same way as above, is again a normed vector space. In the following we will use the notation:

$$|\mu|_{\mathcal{M}(\Omega;V)} = \int_\Omega |\mu|_V, \quad |\sigma|_{\mathcal{M}(\Omega;V^d)} = \int_\Omega |\sigma|_{V^d}, \quad \text{for } \mu \in \mathcal{M}(\Omega; V) \text{ and } \sigma \in \mathcal{M}(\Omega; V^d).$$

On V and V^d we can consider two orthonormal bases, respectively $(v_k)_{k=1}^n$ and $w_{k,l} = v_k \otimes e_l$, for $(e_l)_{l=1}^d$ canonical basis of $\mathbb{R}^d$. Using these bases, a function ϕ belongs to $\mathcal{C}(\Omega; V)$ if and only if there exists $(a_k)_{k=1}^n \subset \mathcal{C}(\Omega; \mathbb{R})$ such that $\phi(x) = \sum_{k=1}^n a_k(x) v_k, \forall x \in \Omega$, and analogously for $\psi \in \mathcal{C}(\Omega; V^d)$. In the same way a measure $\mu \in \mathcal{M}(\Omega; V)$ can be written as $\mu = \sum_{k=1}^n b_k v_k$ for a sequence of real valued measures $(b_k)_{k=1}^n \subset \mathcal{M}(\Omega; \mathbb{R})$, and analogously for $\sigma \in \mathcal{M}(\Omega; V^d)$.

The gradient $\nabla'\phi \in C(\Omega, V^d)$ of a function $\phi \in C^1(\Omega, V)$ is defined componentwise as follows

$$\nabla'\phi = \sum_k v_k \otimes \nabla\phi_k \quad \text{for } \phi = \sum_k \phi_k v_k.$$

Likewise, for $\psi \in C^1(\Omega, V^d)$, the divergence $\text{div}' \psi \in C(\Omega, V)$ is defined as

$$\text{div}' \psi = \sum_k \text{div}(\psi_k) v_k \quad \text{for } \psi = \sum_k v_k \otimes \psi_k.$$

In the following, we drop the superscript over ∇ and div for the sake of notational simplicity. A function $\phi \in \mathcal{C}(\Omega; V)$ is 1-Lipschitz if

$$|\phi(y) - \phi(x)|_V \leq |y - x|, \quad \forall x, y \in \mathbb{R}^d.$$

Since Ω is supposed to be convex,

$$(2.3) \qquad \phi \text{ is 1-Lipschitz} \iff |\nabla\phi|_{L(\mathbb{R}^d;V),\infty} \leq 1,$$

where $|\nabla\phi|_{L(\mathbb{R}^d;V)}$ is the operator norm. By convention, $|\nabla\phi|_{L(\mathbb{R}^d;V),\infty} = \infty$ if the distributional derivative $\nabla\phi$ does not admit an L^∞ representative; the quantity $|\nabla\phi|_{L(\mathbb{R}^d;V),\infty}$ can also be expressed using duality, similarly to (2.5) below. There is no counterpart of the equivalence (2.3) involving the Frobenius $|\cdot|_{V^d}$ norm, and thus we define instead the set

$$(2.4) \qquad \text{Fr-Lip}(\Omega, V) := \{\phi \in \mathcal{C}(\Omega; V), \ |\nabla\phi|_{V^d,\infty} \leq 1\}.$$

We refer to $|\nabla\phi|_{V^d,\infty}$ as the Fr-Lipschitz constant, and note again that it is defined as ∞ by convention if the distributional derivative $\nabla\phi$ has no L^∞ representative. By definition of the

distributional derivative, and since $L^\infty = (L^1)^*$, one has for any $\phi \in \mathcal{C}(\Omega; V)$

$$(2.5) \qquad |\nabla\phi|_{V^d,\infty} := \sup_{\psi \in \mathcal{C}_c^\infty(\text{int}(\Omega); V^d)} \left\{ \int_\Omega \langle \phi, \, \text{div}\,\psi \rangle_V \,, \, |\psi|_{V^d,1} \leq 1 \right\},$$

where $\mathcal{C}_c^\infty$ denotes the space of smooth and compactly supported test functions, $\text{int}(\Omega)$ stands for the interior of the domain and $|\psi|_{V^d,1} := \int_\Omega |\psi|_{V^d}$. Notice that $1\text{-}\operatorname{Lip}(\Omega, V) \subset$ $\operatorname{Fr-Lip}(\Omega, V) \subset \sqrt{d}\text{-}\operatorname{Lip}(\Omega, V)$. Furthermore, $\operatorname{Fr-Lip}(\Omega, V)$ is closed w.r.t. uniform convergence, since (2.5) is lower-semi-continuous for this topology (as a supremum of continuous maps).

3. Unbalanced L^1 optimal transport problem for vector valued measures

In this section we introduce an unbalanced L^1 optimal transport problem for vector valued measures. In order to ease the presentation of the unbalanced model, we first consider the balanced case, i.e. the case of measures μ and ν with the same total mass, and present the vector extension of the dual and Beckmann's formulation of the L^1 transport, problems (1.4) and (1.5). These extensions coincide with the ones introduced in [14]. We then extend these models to the unbalanced setting. We will consider two different penalization for the mass imbalance, an L^1 and an L^2 penalization, respectively in Sections 3.2 and 3.3.

For the sake of generality, we stick to considering $\mu, \nu \in \mathcal{M}(\Omega; V)$ instead of $\mu, \nu \in \mathcal{M}(\Omega; C)$. As in the scalar case in fact, the vectorial extension of the L^1 problem does not require the measures to be positive. Of course, considering positive measures will be key for the application to inverse problems, see Section 5 and Example 3.6.

3.1. **Balanced case.** We consider two measures $\mu, \nu \in \mathcal{M}(\Omega; V)$ with the same total mass, $\int_\Omega \mu = \int_\Omega \nu$, i.e.

$$(3.1) \qquad \int_\Omega \mu_k = \int_\Omega \nu_k \,, \quad \forall\, 1 \leq k \leq n \,, \quad \text{for } \mu = \sum_{k=1}^n \mu_k v_k \text{ and } \nu = \sum_{k=1}^n \nu_k v_k \,.$$

We can define straightforwardly in this setting the dual problem (1.4) as:

$$(3.2) \qquad \sup_{\phi \in \mathcal{C}(\Omega; V)} \left\{ \int_\Omega \langle \phi, \nu - \mu \rangle_V \,, \, |\nabla\phi|_{V^d,\infty} \leq 1 \right\}.$$

Proposition 3.1. *For all $\mu, \nu \in \mathcal{M}(\Omega; V)$, with the same total mass, problem (3.2) admits a solution $\phi \in \operatorname{Fr-Lip}(\Omega; V)$.*

Proof. The proof is similar to the scalar case (see e.g. [36, Proposition 1.11]). Thanks to the regularity constraint, a maximizing sequence ϕ_n is equi-continuous. Furthermore, since the problem is independent of the addition of a constant, we can fix an arbitrary one for each potential ϕ_n and derive a uniform bound using the equi-continuity and the boundedness of the domain Ω. By Ascoli-Arzelà one can extract a subsequence converging uniformly on Ω. For the uniform topology, the functional is continuous, and the space $\operatorname{Fr-Lip}$ closed as observed below (2.5), hence the limit solves (3.2). $\square$

The Beckmann formulation (1.5) can also be extended in a straighforward way. For $\sigma \in \mathcal{M}(\Omega; V^d)$ and $\delta \in \mathcal{M}(\Omega; V)$, the continuity equation $\text{div}(\sigma) = \delta$ with zero flux boundary conditions is defined in the following weak sense

$$-\int_\Omega \langle \nabla\phi, \sigma \rangle_{V^d} = \int_\Omega \langle \phi, \delta \rangle_V \,, \quad \forall\phi \in \mathcal{C}^1(\Omega; V) \,.$$

As a consequence of this definition, (σ, μ) can be a weak solution only if δ has zero total mass, i.e. $\int_\Omega \delta = 0$ according to (3.1). Then we can write the problem:

$$(3.3) \qquad \inf_{\sigma \in \mathcal{M}(\Omega; V^d)} \left\{ \int_\Omega |\sigma|_{V^d}, \ \mathrm{div}(\sigma) = \mu - \nu \right\}.$$

As in the scalar case, (3.2) and (3.3) are in duality and the problems are equivalent, as it can be noticed via formal computations. Enforcing the divergence constraint in (3.3),

$$\inf_\sigma \sup_\phi \int_\Omega |\sigma|_{V^d} + \int_\Omega \langle \phi, \mathrm{div}(\sigma) - \mu + \nu \rangle_V,$$

exchanging inf and sup and integrating by parts, we obtain the following dual problem

$$\sup_\phi \inf_\sigma \int_\Omega |\sigma|_{V^d} - \int_\Omega \langle \nabla\phi, \sigma \rangle_{V^d} + \int_\Omega \langle \phi, \nu - \mu \rangle_V.$$

The minimization in σ provides then

$$\inf_\sigma \int_\Omega |\sigma|_{V^d} - \int_\Omega \langle \nabla\phi, \sigma \rangle_{V^d} = \begin{cases} 0 & \text{if } |\nabla\phi|_{V^d,\infty} \le 1, \\ -\infty & \text{else}, \end{cases}$$

so that we obtain problem (3.2).

In order to rigorously prove the equivalence of problem (3.2) with (3.3) we need to justify the swapping of inf and sup. We can establish this duality result adapting an approach proposed in [36, Section 1.6.3] for proving the equivalence between the primal and dual Kantorovich problems (in the scalar setting). As a byproduct of this result, we also derive existence of a solution for the Beckmann formulation (3.3).

Theorem 3.2. *For all $\mu, \nu \in \mathcal{M}(\Omega; V)$, with the same total mass, problems (3.2) and (3.3) attain the same objective value. Furthermore, the latter admits a solution $\sigma \in \mathcal{M}(\Omega; V^d)$.*

Proof. Let us introduce the functional $\mathcal{F} : \mathcal{C}(\Omega; V^d) \to \mathbb{R}$,

$$\mathcal{F}(f) = - \sup_{\phi \in \mathcal{C}(\Omega; V)} \left\{ \int_\Omega \langle \phi, \nu - \mu \rangle_V, \ |\nabla\phi - f|_{V^d,\infty} \le 1 \right\}.$$

Note that $\nabla\phi - f$ is formally defined as a difference of distributions, and that it admits an L^∞ representative if and only if the distributional derivative $\nabla\phi$ does, since f is continuous. Following the same arguments of the proof of Proposition 3.1, for every f there exists a solution ϕ to the maximization problem (the equi-continuity is provided in this case by observing that $|\nabla\phi|_{V^d,\infty} \le 1 + |f|_{V^d,\infty}$) and for $f = 0$ we obtain the opposite of problem (3.2). We show that $\mathcal{F}$ is convex and lower semi-continuous.

Regarding convexity, take f_1, f_2, the corresponding optimal potentials ϕ_1, ϕ_2, and an interpolation parameter $t \in [0, 1]$. The potential $t\phi_1 + (1-t)\phi_2$ is admissible for the maximization problem with $tf_1 + (1-t)f_2$ since

$$|t\nabla\phi_1 + (1-t)\nabla\phi_2 - tf_1 - (1-t)f_2|_{V^d} \le t|\nabla\phi_1 - f_1| + (1-t)|\nabla\phi_2 - f_2|_{V^d} \le 1$$

pointwise on Ω. Therefore we obtain:

$$\mathcal{F}(tf_1 + (1-t)f_2) \le - \int_\Omega \langle t\phi_1 + (1-t)\phi_2, \nu - \mu \rangle_V = t\mathcal{F}(f_1) + (1-t)\mathcal{F}(f_2).$$

Regarding lower semi-continuity, consider a sequence f_n uniformly converging to f, and which is therefore uniformly bounded. We extract a subsequence (without relabeling it) such that

$\lim_{n\to\infty} \mathcal{F}(f_n) = \liminf_{n\to\infty} \mathcal{F}(f_n)$, still converging to f. If we take the corresponding sequence of optimal potentials ϕ_n, then

$$|\nabla\phi_n|_{V^d} \leq |\nabla\phi_n - f_n|_{V^d} + |f_n|_{V^d} \leq C\,,$$

pointwise on Ω, hence the sequence $(\phi_n)_{n\in\mathbb{N}}$ is uniformly equi-continuous. Up to adding a normalizing constant we may assume that $\int_\Omega \phi_n = 0$ for all $n \in \mathbb{N}$, hence the sequence of potentials is uniformly equi-bounded, and therefore converges uniformly (up to extracting a subsequence) towards a potential ϕ, by the Arzelà-Ascoli theorem. This latter is an admissible potential for f. Indeed, for any test function $\psi \in \mathcal{C}_c^\infty(\text{int}(\Omega), V^d)$ such that $|\psi|_{V^d,1} \leq 1$ we have

$$\int_\Omega \langle \phi_n, \text{div}(\psi)\rangle_V + \langle f_n, \psi\rangle_{V^d} \leq 1\,,$$

which passes to the limit by uniform convergence. This implies $|\nabla\phi - f|_{V^d,\infty} \leq 1$. In the end we get:

$$\mathcal{F}(f) \leq -\int_\Omega \langle \phi, \nu - \mu\rangle_V = \liminf_n \mathcal{F}(f_n)\,.$$

Let us compute the Legendre transform of $\mathcal{F}$:

$$\mathcal{F}^*(\sigma) = \sup_{f\in\mathcal{C}(\Omega;V^d)} \int_\Omega \langle \sigma, f\rangle_{V^d} - \mathcal{F}(f)$$

$$= \sup_{\substack{f\in\mathcal{C}(\Omega;V^d)\\ \phi\in\mathcal{C}(\Omega;V)}} \left\{ \int_\Omega \langle\sigma,f\rangle_{V^d} + \int_\Omega \langle\phi,\nu-\mu\rangle_V\,,\ |\nabla\phi-f|_{V^d,\infty} \leq 1 \right\},$$

$$= \sup_{\substack{f\in\mathcal{C}^\infty(\Omega;V^d)\\ \phi\in\mathcal{C}^\infty(\Omega;V)}} \left\{ \int_\Omega \langle\sigma,f\rangle_{V^d} + \int_\Omega \langle\phi,\nu-\mu\rangle_V\,,\ |\nabla\phi-f|_{V^d,\infty} \leq 1 \right\}.$$

The last equality is justified by the following density argument. Let $f \in \mathcal{C}(\Omega;V^d)$, $\phi \in \mathcal{C}(\Omega;V)$. Choose a non-negative mollifier $\rho \in C^\infty(\mathbb{R}^d)$ supported in the unit ball and with unit integral, and denote $\rho_\delta := \delta^{-d}\rho(\cdot/\delta)$. The convolutions $f_\delta := f * \rho_\delta$ and $\phi_\delta := \phi * \rho_\delta$, obey $|\nabla\phi_\delta - f_\delta|_{V^d,\infty} = |\rho_\delta * (\nabla\phi - f)|_{V^d,\infty} \leq 1$ and are defined on $\Omega_\delta := \{x \in \Omega, d(x,\partial\Omega) \geq \delta\}$. We assume, up to translating the domain, that $\overline{B}(0,r) \subset \Omega$ for some $r > 0$, and observe that $(1-\delta/r)\Omega \subset \Omega_\delta$ by convexity for all $0 < \delta < r$. Denoting $\tilde{f}_\delta(x) := f_\delta((1-\delta/r)x)$ and $\tilde{\phi}_\delta(x) := \phi_\delta((1-\delta/r)x)/(1-\delta/r)$ we have $\tilde{f}_\delta \in C^\infty(\Omega;V^d)$, $\tilde{\phi}_\delta \in C^\infty(\Omega;V)$, $|\nabla\tilde{\phi}_\delta - \tilde{f}_\delta|_{V^d,\infty} \leq 1$, and $\tilde{f}_\delta \to f$ and $\tilde{\phi}_\delta \to \phi$ uniformly on Ω as $\delta \to 0$, hence the equality.

Changing variables to $\hat{f} := f - \nabla\phi \in C^\infty(\Omega;V^d)$ we obtain, by density of smooth functions in the set of continuous functions

$$\mathcal{F}^*(\sigma) = \sup_{\substack{\hat{f}\in C^\infty(\Omega;V^d)\\ \phi\in\mathcal{C}^\infty(\Omega;V)}} \left\{ \int_\Omega \langle\sigma,\hat{f}\rangle + \int_\Omega \langle\sigma,\nabla\phi\rangle_{V^d} + \int_\Omega \langle\phi,\nu-\mu\rangle_V,\ |\hat{f}|_{V^d,\infty} \leq 1 \right\}$$

$$= \sup_{\phi\in\mathcal{C}^\infty(\Omega;V)} \left\{ \int_\Omega |\sigma|_{V^d} + \int_\Omega \langle\sigma,\nabla\phi\rangle_{V^d} + \int_\Omega \langle\phi,\nu-\mu\rangle_V \right\}$$

$$= \begin{cases} \int_\Omega |\sigma|_{V^d} & \text{if } \text{div}(\sigma) = \mu - \nu \\ \infty & \text{else,} \end{cases}$$

which corresponds to the objective functional of problem (3.3).

As $\mathcal{F}$ is convex and lower semi-continuous, it holds

$$-\mathcal{F}(0) = -\mathcal{F}^{**}(0) = -\sup_{\sigma} \left\{ \int_{\Omega} \langle \sigma, 0 \rangle_{V^d} - \mathcal{F}^*(\sigma) \right\} = \inf_{\sigma} \mathcal{F}^*(\sigma) \,,$$

which provides the equivalence between problems (3.3) and (3.2). The existence of a solution σ to the dual problem follows from the fact that $\mathcal{F}$ being convex and lower semi-continuous,

$$\sigma \in \partial \mathcal{F}(0) \iff 0 \in \partial \mathcal{F}^*(\sigma) \iff \mathcal{F}(0) + \mathcal{F}^*(\sigma) = 0 \,.$$

Since 0 is in the interior of the domain of definition of $\mathcal{F}$, the subdifferential in this point is not empty. $\qquad\square$

We introduce the functional $W_1^V : \mathcal{M}(\Omega; V) \times \mathcal{M}(\Omega; V) \to \mathbb{R}_+$,

$$(3.4) \qquad W_1^V(\mu, \nu) := \sup_{\phi \in \mathcal{C}(\Omega; V)} \left\{ \int_{\Omega} \langle \phi, \nu - \mu \rangle_V \,, \ |\nabla \phi|_{V^d, \infty} \le 1 \right\}$$

which associates to any two measures μ and ν the transport cost, if they have the same total mass, else the value $+\infty$. The following proposition is a simple consequence of the above duality result.

Proposition 3.3. $W_1^V(\cdot, \cdot)$ *is jointly convex and lower semi-continuous. Furthermore, it provides a distance on any subset of $\mathcal{M}(\Omega; V)$ of measures with the same total mass.*

3.2. **Unbalanced case.** Requiring the two vector valued measures μ and ν to have the same mass may be in some cases too restrictive, as it requires the strong geometric constraint (3.1). We introduce for this reason an unbalanced version of the problem. For $\mu, \nu \in \mathcal{M}(\Omega; V)$ we define:

$$(3.5) \qquad \mathcal{T}_{1,1}^{\lambda}(\mu, \nu) := \inf_{\delta \in \mathcal{M}(\Omega; V)} W_1^V(\mu + \delta, \nu) + \lambda \int_{\Omega} |\delta|_V$$

for a constant $\lambda > 0$ (fixed in this section). The penalization of the total variation of δ determines the deformation price of the measures. Problem (3.5) aims at finding a balance between transporting mass from μ to ν or deforming it and for $\lambda \to +\infty$ one recovers (formally) a pure transport if this exists. Problem (3.5) is an extension to vector valued measures of the Piccoli-Rossi unbalanced L^1 distance [31]. From the equivalent formulations (3.2)-(3.3) one deduces that $W_1^V(\mu, \nu)$ only depends on the difference of the two measures. The choice of $W_1^V(\mu + \delta, \nu)$ in the definition of problem (3.5) is therefore not restrictive.

Theorem 3.4. *For all $\mu, \nu \in \mathcal{M}(\Omega; V)$, there exists a solution $\delta \in \mathcal{M}(\Omega; V)$ to problem (3.5).*

Proof. The functional $W_1^V(\mu + \delta, \nu)$ is well-posed for any $\delta \in \mathcal{M}(\Omega; V)$ such that $\mu + \delta$ has the same total mass as ν, according to (3.1). It is also lower semi-continuous for the weak-* topology, as a supremum of lower semi-continuous functionals (3.4), and non-negative. Since the total variation norm is lower semi-continuous as well, the existence of a solution follows by the direct method of calculus of variations: consider a minimizing sequence $(\delta_n)_{n \in \mathbb{N}}$, note that $\int_{\Omega} |\delta_n|_V$ is uniformly bounded since $\lambda > 0$ and $W_1^V(\mu + \delta_n, \nu) \ge 0$, hence there is a weak-* converging subsequence by compactness of the closed balls of $\mathcal{M}(\Omega, V)$, whose limit is a minimizer to (3.5) by lower semi-continuity. $\qquad\square$

In the rest of this section we show that (3.5) admits analogous unbalanced versions of (3.2) and (3.3). The latter is evident, it simply requires to use the Beckmann formulation of W_1^V

to obtain:

$$(3.6) \qquad \mathcal{T}_{1,1}^{\lambda}(\mu,\nu) = \min_{\substack{\sigma \in \mathcal{M}(\Omega;V^d) \\ \delta \in \mathcal{M}(\Omega;V)}} \left\{ \int_{\Omega} |\sigma|_{V^d} + \lambda \int_{\Omega} |\delta|_V, \ \mathrm{div}(\sigma) = \mu - \nu + \delta \right\}.$$

Due to (3.6), it can be immediately recognized that $\mathcal{T}_{1,1}^{\lambda}(\cdot,\cdot)$ is again a distance on $\mathcal{M}(\Omega;V)$. Using instead (3.2) to rewrite the W_1^V distance, we obtain the saddle point problem

$$\inf_{\delta \in \mathcal{M}(\Omega;V)} \sup_{\phi \in \mathcal{C}(\Omega;V)} \left\{ \int_{\Omega} \langle \phi, \nu - \mu - \delta \rangle_V + \lambda \int_{\Omega} |\delta|_V \, , \ |\nabla\phi|_{V^d,\infty} \le 1 \right\}.$$

Exchanging inf and sup and minimizing in δ provides

$$(3.7) \qquad \sup_{\phi \in \mathcal{C}(\Omega;V)} \left\{ \int_{\Omega} \langle \phi, \nu - \mu \rangle_V \, , \ |\nabla\phi|_{V^d,\infty} \le 1, \ |\phi|_{V,\infty} \le \lambda \right\},$$

by the same computation used in the previous section. The dual problem (3.7) is a vector extension of the Kantorovich-Rubenstein norm (1.6). The following theorem shows its equivalence with problem (3.5).

Theorem 3.5. *For all* $\mu, \nu \in \mathcal{M}(\Omega;V)$*, there exists a solution* $\phi \in \mathrm{Fr\text{-}Lip}(\Omega;V)$ *to problem* (3.7) *and it holds:*

$$\mathcal{T}_{1,1}^{\lambda}(\mu,\nu) = \max_{\phi \in \mathcal{C}(\Omega;V)} \left\{ \int_{\Omega} \langle \phi, \nu - \mu \rangle_V \, , \ |\nabla\phi|_{V^d,\infty} \le 1, \ |\phi|_{V,\infty} \le \lambda \right\}.$$

Proof. The proof of existence of a solution is similar to that of problem (3.2) in Proposition 3.1, the explicit uniform bound on the potential providing in this case directly the equi-boundedness.

For $\phi \in \mathcal{C}(\Omega,V)$, we introduce the functionals

$$\mathcal{F}(\phi) = \int_{\Omega} \langle \phi, \mu - \nu \rangle_V + i_{|\nabla\phi|_{V^d,\infty} \le 1} \, , \quad \mathcal{G}(\phi) = i_{|\phi|_{V,\infty} \le \lambda} \, ,$$

where we denote by

$$(3.8) \qquad\qquad i_A(\phi) := \begin{cases} 0 & \text{if } \phi \in A, \\ \infty & \text{otherwise,} \end{cases}$$

the indicator function of a set $A \subset \mathcal{C}(\Omega,V)$. If A is convex and closed (for the strong topology), as it is the case in the two functionals above, then i_A is convex and (weakly-*) lower semi-continuous. Therefore, both $\mathcal{F}$ and $\mathcal{G}$ are convex and (weakly-*) lower semi-continuous. Furthermore, for $\phi = 0$ both $\mathcal{F}$ and $\mathcal{G}$ are bounded and $\mathcal{G}$ is also continuous in this point. Then, from the Fenchel-Rockafellar duality theorem [5, Theorem 1.12],

$$\max_{\phi \in \mathcal{C}(\Omega;V)} -\mathcal{F}(\phi) - \mathcal{G}(\phi) = \min_{\delta \in \mathcal{M}(\Omega;V)} \mathcal{F}^*(-\delta) + \mathcal{G}^*(\delta) \, ,$$

where

$$\mathcal{F}^*(-\delta) = \sup_{\phi \in \mathcal{C}(\Omega;V)} \int_{\Omega} \langle \phi, \nu - \mu - \delta \rangle_V - i_{|\nabla\phi|_{V^d,\infty} \le 1} = W_1^V(\mu + \delta, \nu) \, ,$$

by definition of W_1^V, and

$$\mathcal{G}^*(\delta) = \sup_{\phi \in \mathcal{C}(\Omega;V)} \int_{\Omega} \langle \phi, \delta \rangle_V - i_{|\phi|_{V,\infty} \le \lambda} = \lambda \int_{\Omega} |\delta|_V \, ,$$

showing the equivalence between problems (3.5) and (3.7). $\square$

The following example clarifies the role played by the penalization term and is helpful in understanding how the parameter λ can be tuned in practice.

Example 3.6 (Exact solution for two delta measures)**.** Let us consider the two measures $\mu = M_1 \delta_{x_1}, \nu = M_2 \delta_{x_2}$ for $M_1, M_2 \in V$ and two distinct points $x_1, x_2 \in \Omega$. The Kantorovich-Rubinstein norm (3.7) can be written in this case as:

$$(3.9) \qquad |x_1 - x_2| \max_{A,B \in V} \{ \langle A, M_1 \rangle_V - \langle B, M_2 \rangle_V, |A|_V \leq \gamma, |B|_V \leq \gamma, |A - B|_V \leq 1 \},$$

where $\gamma = \frac{\lambda}{|x_1 - x_2|}$. Indeed, let ϕ be such that $|\nabla \phi|_{V^d} \leq 1$ pointwise on the convex domain Ω, which implies $|\nabla \phi|_{L(\mathbb{R}^d;V)} \leq 1$ since the Frobenius norm is larger than the operator norm. Thus ϕ is 1-Lipschitz, and therefore $A := \phi(x_1)$ and $B := \phi(x_2)$ satisfy $|A - B|_V \leq |x_1 - x_2|$. Conversely, assume that $|A - B|_V \leq |x_1 - x_2|$, and define for any $z \in \Omega$

$$\phi(z) = \begin{cases} (1-t)A + tB & \text{if } t \in [0,1], \\ A & \text{if } t \leq 0, \\ B & \text{if } t \geq 1, \end{cases} \qquad \text{where } t := (z - x_1) \cdot n,$$

and $n := (x_2 - x_1)/|x_2 - x_1|^2$. Then ϕ is piecewise linear hence Lipschitz, and when $0 < t < 1$ the gradient $\nabla \phi(z) = (B - A) \otimes n$ is a rank one matrix which satisfies $|\nabla \phi(z)|_{V^d} = |\nabla \phi(z)|_{L(\mathbb{R}^d;V)} = |B - A|_V |n| \leq 1$ as required, since the Frobenius and operator norms coincide on rank one matrices.

We can rewrite (3.9) by means of an infsup formulation as

$$\max_{A,B} \min_{C_1,C_2,C_3} \langle A, M_1 \rangle_V - \langle B, M_2 \rangle_V + |C_1|_V - \frac{1}{\gamma} \langle A, C_1 \rangle_V + |C_2|_V - \frac{1}{\gamma} \langle B, C_2 \rangle_V + |C_3|_V - \langle A - B, C_3 \rangle_V.$$

Swapping min and max by strong duality[1] and optimizing in A and B we obtain

$$M_1 - \frac{1}{\gamma}C_1 - C_3 = 0, \quad -M_2 - \frac{1}{\gamma}C_2 + C_3 = 0,$$

and the problem reduces to

$$\min_{C \in V} \gamma |C - M_1|_V + \gamma |C - M_2|_V + |C|_V.$$

This is the generalized Fermat-Torricelli problem: finding the point that minimizes the weighted sum of the distances from the vertices of a triangle (which lies on the plane passing by the three points $0, M_1, M_2$). The analytical solution is available for this problem for

---

[1]The Fenchel-Rockafellar duality theorem can be applied again for this purpose, as in Theorem 3.5, for $f(A, B, D) := \gamma \langle A, M_1 \rangle_V - \gamma \langle B, M_2 \rangle_V + i_{D=\gamma(A-B)}$ and $g(A, B, D) := i_{|A| \leq 1} + i_{|B| \leq 1} + i_{|D| \leq 1}$ (up to a rescaling of A and B by γ), which are both finite for $(0, 0, 0)$ and the latter is furthermore continuous at this point.

example in [38]. The total cost is then:
(3.10)
$$
\begin{cases}
\lambda(|M_1|_V + |M_2|_V) & \text{if } \dfrac{\langle M_1, M_2\rangle_V}{|M_1|_V|M_2|_V} \leq \dfrac{1}{2\gamma^2} - 1 \\[2ex]
|x_1 - x_2||M_1|_V + \lambda|M_1 - M_2|_V & \text{if } \dfrac{\langle M_1, M_2 - M_1\rangle_V}{|M_1|_V|M_2 - M_1|_V} \geq \dfrac{1}{2\gamma^2} \\[2ex]
|x_1 - x_2||M_2|_V + \lambda|M_1 - M_2|_V & \text{if } \dfrac{\langle M_2, M_1 - M_2\rangle_V}{|M_2|_V|M_1 - M_2|_V} \geq \dfrac{1}{2\gamma^2} \\[2ex]
|x_1 - x_2|\Big((\gamma^2 - \dfrac{1}{2})|M_1 - M_2|_V^2 + \dfrac{1}{2}(|M_1|_V^2 + |M_2|_V^2) + & \\[1ex]
\qquad (4\gamma^2 - 1)^{\frac{1}{2}}(|M_1|_V^2|M_2|_V^2 - \langle M_1, M_2\rangle_V^2)^{\frac{1}{2}}\Big)^{\frac{1}{2}} & \text{else,}
\end{cases}
$$

with the three degenerate cases
$$
\begin{cases}
\lambda|M_1|_V & \text{if } M_2 = 0 \\
\lambda|M_2|_V & \text{if } M_1 = 0 \\
\min(2\lambda, |x_1 - x_2|)|M|_V & \text{if } M_1 = M_2 = M.
\end{cases}
$$

It would be furthermore possible to explicitly write the solution C and recover the optimal potential A, B.

In the scalar case $n = 1$, for $\mu = m_1\delta_{x_1}, \nu = m_2\delta_{x_2}$, with $m_1, m_2 \in \mathbb{R}$, the cost reduces to the already known (e.g. [13, Section 5.1]):

(3.11)
$$
\begin{cases}
\lambda(|m_1| + |m_2|) & \text{if } \gamma \leq \dfrac{1}{2} \text{ or } \operatorname{sign}(m_1) \neq \operatorname{sign}(m_2) \\[1ex]
\lambda(|m_2| - |m_1|) + |x_1 - x_2||m_1| & \text{if } m_2 \geq m_1 \geq 0 \text{ or } 0 \geq m_1 \geq m_2 \\[1ex]
\lambda(|m_1| - |m_2|) + |x_1 - x_2||m_2| & \text{if } m_1 \geq m_2 \geq 0 \text{ or } 0 \geq m_2 \geq m_1.
\end{cases}
$$

The first condition in (3.10) discriminates between transporting or not the mass, depending on the parameter λ. If it is satisfied, then the mass is not transported (and the cost corresponds to its destruction at x_1 and its reconstruction at x_2). If the product $\langle M_1, M_2\rangle_V$ is non-negative, then there is never transport for $\lambda \leq \frac{|x_1 - x_2|}{2}$, whereas transport occurs for $\lambda \geq \frac{|x_1 - x_2|}{\sqrt{2}}$, with an intermediate regime in between depending on the angle. If on the contrary $\langle M_1, M_2\rangle_V$ is negative, λ needs to increase further to ensure transport, depending on the angle between M_1 and M_2, tending to $+\infty$ in the limiting case they have opposite directions, in which case transport never occurs. This latter phenomenon is the equivalent of the scalar case with a positive dirac and a negative dirac (3.11). This motivates our choice of considering measures valued in a cone $C \subseteq C^*$, in order to have a meaningful transport problem. Indeed, in this case $\langle M_1, M_2\rangle_V \geq 0$ for any $M_1, M_2 \in C$.

The previous example shows that transport may not occur if the data are not signed. The next example shows a more interesting case where this can lead to a somewhat pathological behavior.

Example 3.7 (Insensitivity of KR to horizontal shifts for signed measures)**.** Consider for simplicity $V = \mathbb{R}$, in which case (1.6) and (3.7) coincide, and two measures $\mu, \nu \in \mathcal{M}(\Omega; V)$. Notice that (1.6) only depends on the difference of μ and ν and can be therefore formulated between the two positive measures $(\mu - \nu)_+$ and $(\mu - \nu)_-$. In terms of optimal transport, this means that positive mass can be transported on negative one and viceversa. Consider the
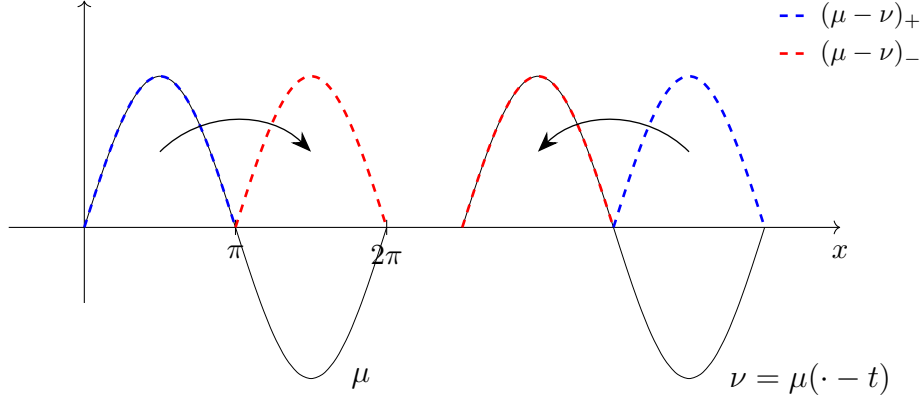
FIGURE 1. Representation of the transport problem of Example 3.7. The L^1 optimal transport between $(\mu - \nu)_+$ and $(\mu - \nu)_-$ is independent of the shift $t \geq 2\pi$. The solution is the shift, of length π, of the leftmost blue bump of mass to the right and of the rightmost blue bump of mass to the left. Notice that, if μ has not zero mean, this is not possible and some leftover mass needs to travel from the leftmost to the rightmost side (or viceversa).

two measures $\mu = \sin_{2\pi}(x) \coloneqq \sin(\max\{0, \min\{2\pi, x\}\})$, and $\nu = \mu(\cdot - t)$ for $t \in \mathbb{R}_+, t \geq 2\pi$ (see Figure 1). Since $\int_\Omega \mu = \int_\Omega \nu = 0$, from the previous example we can deduce that $\mathrm{KR}(\mu, \nu) = \mathrm{KR}((\mu - \nu)_+, (\mu - \nu)_-) = W_1((\mu - \nu)_+, (\mu - \nu)_-) = 2\pi$, for a sufficiently big value of the parameter λ. The transport cost is independent of the parameter t, i.e. the cost is independent of the shift between the two measures. We highlight that this is true since μ has zero mean. If this is not the case, one recovers sensitivity to the time shift, since an amount of mass needs to be transported between μ_+ and ν_+ (or between μ_- and ν_-).

3.3. **Quadratic penalization.** As we have seen in the previous section, when using a one-homogeneous total variation norm penalization the parameter λ corresponds to half the transport distance up to which mass can be transported. This behavior is independent of the mass considered. In applications it may be more relevant instead to consider a penalization on the transport distance depending also on the mass. This can be achieved by considering a two-homogeneous penalization. Another advantage in the specific application of interest for this work, namely seismic imaging, is that the quadratic penalization corresponds to the physical energy of the signal (see Section 5). On the contrary, the total variation has no physical interpretation and thus sometimes induces undesirable artifacts, in which some regions of very low physical energy have an excessively strong influence on the transport problem.

We restrict in this section our attention to measures which are absolutely continuous w.r.t. the Lebesgue measure and whose density lies in $L^2(\Omega; V)$. By a slight abuse of notation, we identify in this case measures with their densities. Since the domain Ω is compact, $\mathcal{M}(\Omega; V) \cap L^2(\Omega; V)$ coincides with $L^2(\Omega; V)$ itself. We define the unbalanced L^1 transport cost with quadratic penalization as

$$(3.12) \qquad \mathcal{T}_{1,2}^\lambda(\mu, \nu) \coloneqq \inf_{\delta \in L^2(\Omega; V)} W_1^V(\mu + \delta, \nu) + \frac{\lambda}{2} \int_\Omega |\delta|_V^2,$$

for any $\mu, \nu \in L^2(\Omega; V)$. The following theorem provides the analogous of the unbalanced formulations (3.7) and (3.6) in this case.

Theorem 3.8. *For all $\mu, \nu \in L^2(\Omega; V)$, problem (3.12) admits a minimizer $\delta \in L^2(\Omega; V)$ and it holds:*

$$(3.13) \qquad \mathcal{T}_{1,2}^\lambda(\mu, \nu) = \max_{\phi \in \mathcal{C}(\Omega; V)} \left\{ \int_\Omega \langle \phi, \nu - \mu \rangle_V - \frac{1}{2\lambda} \int_\Omega |\phi|_V^2, \ |\nabla\phi|_{V^d, \infty} \le 1 \right\}$$

$$(3.14) \qquad = \min_{\substack{\sigma \in \mathcal{M}(\Omega; V^d) \\ \delta \in L^2(\Omega; V)}} \left\{ \int_\Omega |\sigma|_{V^d} + \frac{\lambda}{2} \int_\Omega |\delta|_V^2, \ \mathrm{div}(\sigma) = \mu - \nu + \delta \right\}.$$

Furthermore, the optimal potential ϕ of (3.13) is unique.

Proof. Let δ_n be a minimizing sequence for (3.12). Since $W_1^V(\mu + \delta, \nu) \ge 0$, we can assume that the sequence is bounded in $L^2(\Omega; V)$. The $L^2(\Omega; V)$ space is not closed with respect to the weak-* topology. We can consider however in this case the weak L^2 topology to extract a converging subsequence. From definition (3.2), since $\mathcal{C}(\Omega; V) \subset L^2(\Omega; V)$, Ω being compact, the distance W_1^V is clearly lower semi-continuous with respect to this topology. We obtain therefore existence of an optimal δ for (3.12). The equivalence with (3.14) is trivial.

Let us consider existence for problem (3.13). In order to derive the uniform boundedness for a sequence of maximizers, we can use the uniform Lipschitz regularity. The problem is not independent of the addition of a constant to the potential but we can determine a uniform bound for it. Problem (3.13) can be rewritten as

$$\sup_{\substack{a \in \mathbb{R}^d, \hat\phi \in \mathcal{C}(\Omega; V) \\ \int_\Omega \hat\phi = 0}} \left\{ \int_\Omega \langle \hat\phi, \nu - \mu \rangle_V + a \int_\Omega (\nu - \mu) - \frac{1}{2\lambda} \int_\Omega |\hat\phi|_V^2 - \frac{a^2}{2\lambda}|\Omega|, \ |\nabla\hat\phi|_{V^d, \infty} \le 1 \right\},$$

by making explicit the average a of the potential ϕ, and introducing the zero-mean difference $\hat\phi := \phi - a/|\Omega|$. Consider then a maximizing sequence $(a_n, \hat\phi_n)$. The average a_n clearly converges to the optimal value $\lambda \int_\Omega (\nu - \mu)/|\Omega|$. The sequence of zero mean uniformly Lipschitz potential $\hat\phi_n$ is now also uniformly bounded, $|\hat\phi_n| \le \mathrm{diam}(\Omega) := \max\{|x - y|, x, y \in \Omega\}$. Indeed, $\hat\phi_n$ is Lipschitz because the Frobenius norm is larger than the operator norm. Second, let $v \in V$ with $|v|_V = 1$ be such that $|\hat\phi_n|_\infty = |\langle \hat\phi_n, v \rangle_V|_\infty$; then $x \mapsto \langle \hat\phi_n(x), v \rangle_V$ is 1-Lipschitz, zero-mean and scalar valued, hence bounded by $\mathrm{diam}(\Omega)$ as announced. Hence, we can apply Ascoli-Arzelà as in the proof of Proposition (3.1) to deduce convergence to an optimal $\hat\phi$. Finally, the optimal potential in (3.13) is $\phi = \lambda \int_\Omega (\nu - \mu)/|\Omega| + \hat\phi$. Uniqueness derives from the strict convexity of the objective functional.

To conclude, the proof of the duality result is analogous to the proof of Theorem 3.5. We introduce the two convex and lower semi-continuous functionals

$$\mathcal{F}(\phi) = \int_\Omega \langle \phi, \mu - \nu \rangle_V + i_{|\nabla\phi|_{V^d, \infty} \le 1}, \quad \mathcal{G}(\phi) = \frac{1}{2\lambda} \int_\Omega |\phi|_V^2,$$

which satisfies the hypotheses of the Fenchel-Rockafellar duality theorem [5, Theorem 1.12] (consider again $\phi = 0$). It holds again

$$\mathcal{F}^*(-\delta) = W_1^V(\mu + \delta, \nu),$$

whereas for the second term

$$\mathcal{G}^*(-\delta) = \sup_{\phi \in \mathcal{C}(\Omega; V)} - \int_\Omega \langle \phi, \delta \rangle_V - \frac{1}{2\lambda} \int_\Omega |\phi|_V^2 = \begin{cases} \frac{\lambda}{2} \int_\Omega |\delta|_V^2 & \text{if } \delta \in L^2(\Omega; V), \\ +\infty & \text{else,} \end{cases}$$

since $\int_\Omega \langle \phi, \delta \rangle_V$ is not a bounded functional on $L^2(\Omega; V)$ if $\delta \notin L^2(\Omega; V)$. The equivalence between (3.12) and (3.13) follows. □

Remark 3.9. *Differently from problems* (3.12) *and* (3.14) *which require* $\mu, \nu \in L^2(\Omega; V)$, *the dual formulation* (3.13) *is well-defined for every* $\mu, \nu \in \mathcal{M}(\Omega; V)$.

We can again shed light on the role played by the parameter λ, in this case with quadratic penalization, by considering an exact solution.

Example 3.10 (Exact solution for two delta measures). Consider the case $d = 1, n = 1$ and two measures of the form $\mu = m_1 \delta_{x_1}, \nu = m_2 \delta_{x_2}$ for $m_1, m_2 \in \mathbb{R}_+$. Following Example 3.6, we want to rewrite problem (3.13) in the variables $a = \phi(x_1)$ and $b = -\phi(x_2)$. Clearly, the L^2 norm squared of ϕ is minimized, under the gradient constraint, by a piecewise linear function which on its support has slope ± 1 almost everywhere, and may only have a local maximum at x_1 and a local minimum at x_2. Suppose that $m_1 > m_2$. We need to distinguish between the cases $\lambda \geq \frac{|x_1 - x_2|^2}{m_1 - m_2}$ and $\lambda \leq \frac{|x_1 - x_2|^2}{m_1 - m_2}$. In the first case the first mass dominates on the second one and we can therefore write the problem as:

$$\sup_a \left\{ am_1 + bm_2 - \frac{1}{3\lambda} a^3, \ a \geq 0, b = -a + |x_1 - x_2| \right\},$$

where $\phi(x) = \max\{0, a - |x - x_1|\}$ is the hat function centered at x_1, and thus $\frac{1}{3} a^3 = \frac{1}{2} \int \phi^2$. In the second case instead the problem writes:

$$\sup_{a,b} \left\{ am_1 + bm_2 - \frac{1}{3\lambda} (a^3 + b^3), \ a + b \leq |x_1 - x_2|, a \geq 0, b \geq 0 \right\},$$

where now $\phi(x) = \max\{0, a - |x - x_1|\} - \max\{0, b - |x - x_2|\}$ is the difference of two hat functions centered at x_1 and x_2 and whose supports have disjoint interiors, and thus $\frac{1}{2} \int \phi^2 = \frac{1}{3}(a^3 + b^3)$. Easy computations finally show that the solution has the following form:

(3.15)

$$\begin{cases} a = \sqrt{\lambda(m_1 - m_2)}, b = -a + |x_1 - x_2| & \text{if } \lambda \geq \dfrac{|x_1 - x_2|^2}{m_1 - m_2}, \\[2mm] a = \dfrac{|x_1 - x_2|}{2} + \dfrac{\lambda(m_1 - m_2)}{2|x_1 - x_2|}, b = \dfrac{|x_1 - x_2|}{2} - \dfrac{\lambda(m_1 - m_2)}{2|x_1 - x_2|} & \text{if } \lambda \leq \dfrac{|x_1 - x_2|^2}{m_1 - m_2}, \\[2mm] a = \sqrt{\lambda m_1}, b = \sqrt{\lambda m_2} & \text{if } \lambda \leq \dfrac{|x_1 - x_2|^2}{(\sqrt{m_1} + \sqrt{m_2})^2}. \end{cases}$$

As in Example 3.6, λ represents a threshold for transporting the mass. If the third condition is satisfied in (3.15), mass is not transported. In this case however λ represents the distance at which a unit of mass is transported. By adding a quadratic penalization term in (3.12), particles are transported according to their mass, allowing smaller masses to have a lower influence on the transport problem.

Remark 3.11 (The H^{-1} norm). *From* (3.14) *arises the natural idea of endowing also* $\mathcal{M}(\Omega; V^d)$ *with the* L^2 *structure and write*

(3.16)
$$\inf_{\substack{\sigma \in L^2(\Omega; V^d) \\ \delta \in L^2(\Omega; V)}} \left\{ \int_\Omega |\sigma|_{V^d}^2 + \frac{\lambda}{2} \int_\Omega |\delta|_V^2, \ \mathrm{div}(\sigma) = \mu - \nu + \delta \right\}.$$

In this case the problem coincides with the H^{-1} *norm of* $\mu - \nu$. *Indeed, the dual problem to* (3.16) *is given by*

(3.17)
$$\sup_\phi \int_\Omega \langle \phi, \nu - \mu \rangle_V - \frac{1}{2\lambda} \int_\Omega |\phi|_V^2 - \frac{1}{2} \int_\Omega |\nabla \phi|_{V^d}^2$$

which corresponds to the dual norm

$$\sup_{|\phi|_{H^1} \leq 1} \int_\Omega \langle \phi, \nu - \mu \rangle_V , \quad |\phi|_{H^1}^2 := \frac{1}{\lambda} \int_\Omega |\phi|_V^2 + \int_\Omega |\nabla \phi|_{V^d}^2 .$$

With this choice the optimal transport structure is lost. The solution ϕ is provided by the following elliptic equation:

$$(3.18) \qquad\qquad\qquad \Delta \phi + \frac{1}{\lambda} \phi = \nu - \mu .$$

By writing ϕ, ν and μ in the common basis of V, (3.18) is a set of n scalar equations that can be solved separately and (3.16) is decoupled in the different components of μ, ν. Problem (3.16) is not interesting for the purpose of this paper but it will be considered for comparison. We denote $\mathcal{T}_{2,2}^\lambda(\mu, \nu)$ the functional that associates the two measures $\mu, \nu \in L^2(\Omega; V)$ with the optimal value in (3.17).

4. Discretization and numerical solution

We assume in this section a rectangular domain of the form $\Omega = \Pi_{i=1}^d [a_i, b_i]^d$, where $a_i, b_i \in \mathbb{R}$ and $a_i < b_i$ for all $1 \leq i \leq d$, and consider a forward finite difference discretization of our vector valued unbalanced optimal transport problem. We restrict to this setting for simplicity, but a straightforward extension to more general polyhedral domains and unstructured partitioning may be obtained by relying on finite volume discretizations.

For simplicity, let $d := 2$ in the rest of this section (the extension to arbitrary dimension is straightforward). The domain $\Omega = [a_x, b_x] \times [a_y, b_y]$ is discretized using a uniform Cartesian grid $\mathcal{G}$, with mesh sizes $h_x = \frac{b_x - a_x}{N_x - 1}$ and $h_y = \frac{b_y - a_y}{N_y - 1}$, where N_x and N_y are the number of grid points, respectively along the first and second direction. A point on the grid is denoted by $(x_i, y_j) = (a_x + (i-1)h_x, a_y + (j-1)h_y) \in \mathcal{G}$, for $1 \leq i \leq N_x$, $1 \leq j \leq N_y$. The set of discrete functions defined on $\mathcal{G}$, with values respectively in V and V^d, are denoted by $\mathcal{X} := \mathfrak{F}(\mathcal{G}; V)$ and $\mathcal{X}^d := \mathfrak{F}(\mathcal{G}; V^d)$ (we will use the compact notation $\mathcal{X}$ and $\mathcal{X}^d$ in subscripts for simplicity). We use the convention $\phi_{i,j} \in V$ to denote the (i,j)-th component for every discrete function $\phi \in \mathfrak{F}(\mathcal{G}; V)$ and similarly for functions $\psi \in \mathfrak{F}(\mathcal{G}; V^d)$. These spaces are endowed with the scalar products

$$\langle u, v \rangle_\mathcal{X} = h_x h_y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \langle u_{i,j}, v_{i,j} \rangle_V, \qquad \forall u, v \in \mathfrak{F}(\mathcal{G}; V) ,$$

$$\langle w, z \rangle_{\mathcal{X}^d} = h_x h_y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \langle w_{i,j}, z_{i,j} \rangle_{V^d}, \quad \forall w, z \in \mathfrak{F}(\mathcal{G}; V^d) ,$$

and the respective norms $| \cdot |_\mathcal{X}, | \cdot |_{\mathcal{X}^d}$.

We also introduce the discrete l^p and l^∞ norms on the discrete spaces $\mathfrak{F}(\mathcal{G}; V)$ and $\mathfrak{F}(\mathcal{G}; V^d)$ as

$$|u|_{\mathcal{X},p}^p = h_x h_y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} |u_{i,j}|_V^p , \qquad |w|_{\mathcal{X}^d,p}^p = h_x h_y \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} |w_{i,j}|_{V^d}^p,$$

$$|u|_{\mathcal{X},\infty} = \max_{i,j} |u_{i,j}|_V , \qquad\qquad |w|_{\mathcal{X}^d,\infty} = \max_{i,j} |w_{i,j}|_{V^d} ,$$

for $u \in \mathfrak{F}(\mathcal{G}; V), w \in \mathfrak{F}(\mathcal{G}; V^d)$.

The discrete gradient operator $\nabla_{\mathcal{X}} : \mathfrak{F}(\mathcal{G}; V) \longrightarrow \mathfrak{F}(\mathcal{G}; V^d)$ (we recall that $d = 2$) is defined as

$$(\nabla_{\mathcal{X}}(u))_{i,j} := \left(\frac{u_{i+1,j} - u_{i,j}}{h_x} , \frac{u_{i,j+1} - u_{i,j}}{h_y} \right) \qquad 1 \leq i \leq N_x , 1 \leq j \leq N_y ,$$

with the convention $u_{N_x+1,j} = u_{N_x,j}$ and $u_{i,N_y+1} = u_{i,N_y}$. The discrete divergence operator $\mathrm{div}_{\mathcal{X}} : \mathfrak{F}(\mathcal{G}; V^d) \longrightarrow \mathfrak{F}(\mathcal{G}; V)$ is instead defined by duality as

$$\langle \mathrm{div}_{\mathcal{X}}(w), u \rangle_{\mathcal{X}} = -\langle w, \nabla_{\mathcal{X}}(u) \rangle_{\mathcal{X}^d} .$$

It automatically inherits the discrete Neumann boundary conditions.

The discrete setting allows us to uniformize the notation for the discrete transport problem. Given two discrete measures $\mu, \nu \in \mathfrak{F}(\mathcal{G}; V)$, and two natural numbers $p, q \geq 1$, the discrete (generalized) Beckmann's minimal flow problem is

$$(4.1) \qquad (\mathcal{T}_{\mathcal{X}})_{p,q}^{\lambda}(\mu, \nu) := \inf_{\substack{\sigma \in \mathfrak{F}(\mathcal{G}; V^d) \\ \delta \in \mathfrak{F}(\mathcal{G}; V)}} \left\{ \frac{1}{p} |\sigma|_{\mathcal{X}^d,p}^p + \frac{\lambda}{q} |\delta|_{\mathcal{X},q}^q , \; \mathrm{div}_{\mathcal{X}}(\sigma) = \mu - \nu + \delta \right\}.$$

The following proposition justifies the existence of a solution and gives the equivalent discrete dual formulation of (4.1).

Proposition 4.1. *For all $\mu, \nu \in \mathfrak{F}(\mathcal{G}; V)$, the following duality formulas hold:*

$$(4.2) \qquad (\mathcal{T}_{\mathcal{X}})_{1,1}^{\lambda}(\mu, \nu) = \max_{\phi \in \mathfrak{F}(\mathcal{G}; V)} \left\{ \langle \phi, \nu - \mu \rangle_{\mathcal{X}} , \; |\nabla_{\mathcal{X}}(\phi)|_{\mathcal{X}^d,\infty} \leq 1, \; |\phi|_{\mathcal{X},\infty} \leq \lambda \right\},$$

$$(4.3) \qquad (\mathcal{T}_{\mathcal{X}})_{1,2}^{\lambda}(\mu, \nu) = \max_{\phi \in \mathfrak{F}(\mathcal{G}; V)} \left\{ \langle \phi, \nu - \mu \rangle_{\mathcal{X}} - \frac{1}{2\lambda} |\phi|_{\mathcal{X},2}^2 , \; |\nabla_{\mathcal{X}}(\phi)|_{\mathcal{X}^d,\infty} \leq 1 \right\},$$

$$(4.4) \qquad (\mathcal{T}_{\mathcal{X}})_{2,2}^{\lambda}(\mu, \nu) = \max_{\phi \in \mathfrak{F}(\mathcal{G}; V)} \left\{ \langle \phi, \nu - \mu \rangle_{\mathcal{X}} - \frac{1}{2\lambda} |\phi|_{\mathcal{X},2}^2 - \frac{1}{2} |\nabla_{\mathcal{X}}(\phi)|_{\mathcal{X}^d,2}^2 \right\}.$$

Furthermore, the primal problem (4.1) admits a solution as well.

Proof. The proof of duality can be achieved as a straightforward adaptation of the proof strategies of Section 3 to the discrete setting, providing at the same time existence for (4.1). Existence for problems (4.2)-(4.3)-(4.4) follows from discrete compactness arguments. $\square$

Remark 4.2. *The discretization of the KR norm proposed in the original works [1, 23], on its application to seismic imaging, corresponds to a staggered grid finite difference scheme. As a result, the horizontal and vertical components of the gradient are stored in different locations of the grid, which leads these authors to use an axis-biased discretization of the gradient norm, and thus to implement a Lipschitz regularity constraint w.r.t. the Manhattan distance. Our discretization corresponds instead to a forward finite difference scheme: although being in general less accurate, this is consistent with the correct isotropic gradient norm $|\cdot|_{V^d}$, helping to eliminate substantial artifacts in the optimal transport solution.*

We present now an optimization strategy for solving problem (4.1). We focus here on the two cases $(p, q) = (1, 1)$ and $(p, q) = (1, 2)$ since the case $(p, q) = (2, 2)$ can be tackled in an easier way. Indeed, the latter can be addressed by solving directly the system of optimality conditions, namely the discrete version of the Poisson equation (3.18):

$$(4.5) \qquad \Delta_{\mathcal{X}}(\phi) + \frac{1}{\lambda} \phi = \nu - \mu , \quad \text{where } \Delta_{\mathcal{X}}\phi := \mathrm{div}_{\mathcal{X}}(\nabla_{\mathcal{X}}\phi).$$

Problems (4.2) and (4.3) can be written as minimization problems of the following form

$$(4.6) \qquad \min_{\phi} \sum_{i}^{N} f_i (L_i \phi),$$

for $(f_i)_1^N$ convex functions and $(L_i)_1^N$ full rank operators. The SDMM algorithm [16] is designed to solve problems in this form. The algorithm is given in 1. One of the main

---

Algorithm 1: SDMM

Starting from $\boldsymbol{y}^0, \boldsymbol{z}^0$ and $\tau > 0$;
for $n = 0, 1, \ldots$ **do**

$\quad \phi^n = \left(\sum_i^N L_i^T L_i \right)^{-1} \left(\sum_i^N L_i^T (y_i^n - z_i^n) \right)$;

$\quad$ **for** $i = 1, .., N$ **do**

$\quad\quad y_i^{n+1} = \operatorname{prox}_{\tau f_i}(L_i \phi^n + z_i^n)$;

$\quad\quad z_i^{n+1} = z_i^n + L_i \phi^n - y_i^{n+1}$;

$\quad$ **end**

end

---

building blocks is the computation of the proximal operators of $(f_i)_1^N$. For a real valued convex function f and a real parameter $\tau > 0$, the proximal operator (with respect to the rescaled Euclidean norm $|\cdot|_{\mathcal{X},2}$) is defined as

$$x^* = \operatorname{prox}_{\tau f}(y) = \operatorname*{argmin}_{x} \frac{1}{2\tau} |x - y|_{\mathcal{X},2}^2 + f(x).$$

The approach can be particularly effective as long as the computation of the proximal operators and the inversion of the operator $\sum_i^N L_i^T L_i$ can be done efficiently.

For problem (4.2), the operators L_i and the functions f_i are

$$L_1 = \operatorname{Id}_{\mathcal{X}}, \quad L_2 = \nabla_{\mathcal{X}}, \quad L_3 = \frac{1}{\lambda} \operatorname{Id}_{\mathcal{X}},$$

$$f_1(u) = -\langle u, \nu - \mu \rangle_{\mathcal{X}}, \quad f_2(w) = i_{\mathcal{B}_{\mathcal{X}^d}}(w), \quad f_3(u) = i_{\mathcal{B}_{\mathcal{X}}}(u),$$

for $u \in \mathfrak{F}(\mathcal{G}; V), w \in \mathfrak{F}(\mathcal{G}; V^d)$, where $\operatorname{Id}_{\mathcal{X}}$ denotes the identity operator on $\mathfrak{F}(\mathcal{G}; V)$ and $i_{\mathcal{B}_{\mathcal{X}}}$ (respectively $i_{\mathcal{B}_{\mathcal{X}^d}}$) is the convex indicator function (3.8) of the set

$$\mathcal{B}_{\mathcal{X}} = \{u \in \mathfrak{F}(\mathcal{G}; V), |u|_{\mathcal{X},\infty} \le 1\} \ \left(\text{respectively } \mathcal{B}_{\mathcal{X}^d} = \{w \in \mathfrak{F}(\mathcal{G}; V^d), |w|_{\mathcal{X}^d,\infty} \le 1\}\right).$$

For problem (4.3) instead

$$L_1 = \operatorname{Id}_{\mathcal{X}}, \quad L_2 = \nabla_{\mathcal{X}},$$

$$f_1(u) = -\langle u, \nu - \mu \rangle_{\mathcal{X}} + \frac{1}{2\lambda} |u|_{\mathcal{X}^d,2}^2, \quad f_2(w) = i_{\mathcal{B}_{\mathcal{X}^d}}(w).$$

The three proximal steps are inexpensive operations as they can be solved component-wise in closed form. For $q = 1$, the first one amounts to solve the problem

$$\bar{v} = \operatorname{prox}_{\tau f_1}(u) = \operatorname*{argmin}_{v} \frac{1}{2\tau} |v - u|_{\mathcal{X},2}^2 - \langle v, \nu - \mu \rangle_{\mathcal{X}}$$

$$= \left(\operatorname*{argmin}_{v_{i,j}} \frac{1}{2\tau} |v_{i,j} - u_{i,j}|_V^2 - \langle v_{i,j}, \nu_{i,j} - \mu_{i,j} \rangle_V \right)_{i,j},$$

whose solution is simply

$$\bar{v}_{i,j} = u_{i,j} + \tau(\nu_{i,j} - \mu_{i,j}), \quad \forall (i,j).$$

For the case $q = 2$ the proximal operator is analogous, with an additional quadratic term, and the solution becomes

$$\bar{v}_{i,j} = \frac{1}{1 + \tau/\lambda} \left(u_{i,j} + \tau(\nu_{i,j} - \mu_{i,j}) \right), \quad \forall (i,j).$$

The proximal operator of $i_{\mathcal{B}_\mathcal{X}}$ and $i_{\mathcal{B}_{\mathcal{X}^d}}$ are component-wise block soft-thresholding [28] with respect to the norms $|\cdot|_V$ and $|\cdot|_{V^d}$,

$$\bar{v}_{i,j} = \left(\mathrm{prox}_{\tau i_{\mathcal{B}_\mathcal{X}}}(u) \right)_{i,j} = \left(\operatorname*{argmin}_{v \in i_{\mathcal{B}_\mathcal{X}}} |v_{i,j} - u_{i,j}|_V^2 \right)_{i,j},$$

whose solution is

(4.7)
$$\bar{v}_{i,j} = \begin{cases} u_{i,j} & \text{if } |u_{i,j}|_V \le 1 \\ \frac{u_{i,j}}{|u_{i,j}|_V} & \text{else} \end{cases} \quad \forall (i,j),$$

and analogously for $i_{\mathcal{B}_{\mathcal{X}^d}}$. Finally, the operator $\sum_i^N L_i^T L_i$ is

$$\sum_{i=1}^N L_i^T L_i = \Delta_\mathcal{X} + \left(1 + 1/\lambda^2\right) \mathrm{Id}_\mathcal{X}, \qquad \text{for } q = 1,$$

$$\sum_{i=1}^N L_i^T L_i = \Delta_\mathcal{X} + \mathrm{Id}_\mathcal{X}, \qquad \text{for } q = 2.$$

Remark 4.3. *As already pointed out in Section 2, by expanding the discrete functions of $\mathfrak{F}(\mathcal{G}; V)$ in a basis of V, the operators $\mathrm{div}_\mathcal{X}$ and $\nabla_\mathcal{X}$ act component-wise on the coefficients of the basis and so does their composition $\Delta_\mathcal{X}$, meaning that they can be represented as block-diagonal matrices. The corresponding linear systems can be therefore solved separately in each component, which implies solving scalar Poisson equations. This latter is the most expensive operation and can be done very efficiently, either via Fast Fourier Transform [37] or multigrid solvers [3]. The computational complexity for solving problems (4.2) and (4.3) increases therefore only linearly with the dimension of the space V. We stress also that, as in the scalar case, requiring $\mu, \nu \in \mathfrak{F}(\mathcal{G}; C)$ does not reflect on the computational complexity of the algorithm, an advantage of the L^1 optimal transport problem compared to other formulations.*

Algorithm 1 is provably convergent. Since f_i are all lower semi-continuous, proper and convex, for $n \to +\infty$ it holds [2]

$$z_i^n \longrightarrow z_i, \quad y_i^n - L_i \phi^n \longrightarrow 0, \quad \sum_i^N f_i\left(L_i \phi^n\right) \longrightarrow \min_\phi \sum_i^N f_i\left(L_i \phi\right).$$

Relying on the fact that f_1 is differentiable with Lipschitz gradient (in both cases), we further obtain $\phi^n \to \phi$ solution to problem (4.6) and $y_i = L_i \phi$ [18]. The following proposition relates the fixed point of the algorithm to the solution of the Beckmann problem (4.1).

Proposition 4.4. *Let (ϕ, y, z) be a fixed point of the algorithm (1) for either problem (4.2) (q=1) or (4.3) (q=2). Then:*

(i) $z_1 = \tau(\mu - \nu), z_2 = \tau\sigma, z_3 = \tau\lambda\delta$ (q = 1),

(ii) $z_1 = \tau(\mu - \nu + \delta), z_2 = \tau\sigma$ (q = 2),

where (σ, δ) is a solution to (4.1).

Proof. We consider first the case $q = 1$. The identity $y_1 = \text{prox}_{\tau f_1}(\phi + z_1)$ implies

$$(4.8) \qquad \frac{1}{\tau}(y_1 - \phi - z_1) - (\nu - \mu) = 0 \,,$$

which means $z_1 = \tau(\mu - \nu)$ as $y_1 = \phi$. The identity $y_2 = \text{prox}_{\tau i_{\mathcal{B}_{\mathcal{X}^d}}}(\nabla_{\mathcal{X}}(\phi) + z_2)$ implies instead

$$-\frac{1}{\tau}(y_2 - \nabla_{\mathcal{X}}(\phi) - z_2) = \frac{1}{\tau}z_2 \in \partial i_{\mathcal{B}_{\mathcal{X}^d}}(\nabla_{\mathcal{X}}(\phi)) \,.$$

Using the definition of the subdifferential and recalling that by optimality of ϕ it holds $|\nabla_{\mathcal{X}}(\phi)|_{\mathcal{X}^d,\infty} \le 1$, i.e. $i_{\mathcal{B}_{\mathcal{X}^d}}(\nabla_{\mathcal{X}}(\phi)) = 0$, we obtain

$$\langle z_2/\tau, w - \nabla_{\mathcal{X}}(\phi)\rangle_{\mathcal{X}^d} \le i_{\mathcal{B}_{\mathcal{X}^d}}(w), \quad \forall w \in \mathfrak{F}(\mathcal{G}; V^d) \,.$$

If we take then $w = \frac{z_2}{|z_2|}$ this implies

$$0 \le \frac{1}{\tau}|z_2|_{\mathcal{X}^d,1} - \langle \frac{1}{\tau}z_2, \nabla_{\mathcal{X}}(\phi)\rangle_{\mathcal{X}^d} \le 0 \,,$$

where the first inequality is Cauchy-Schwarz. Repeating the same computation for the third proximal operator we obtain $\frac{1}{\tau}|z_3|_{\mathcal{X},1} - \langle \frac{1}{\tau}z_3, \phi/\lambda\rangle_{\mathcal{X}^d} = 0$. Finally, the Poisson equation reduces to

$$\Delta_{\mathcal{X}}(\phi) + \left(1 + 1/\lambda^2\right)\phi = (y_1 - z_1) - \text{div}_{\mathcal{X}}(y_2 - z_2) + (y_3 - z_3)/\lambda$$
$$= \phi + \tau(\nu - \mu) - \text{div}_{\mathcal{X}}(\nabla_{\mathcal{X}}(\phi) - z_2) + (\phi/\lambda - z_3)/\lambda \,,$$

which, simplifying the potential, leads to

$$\text{div}_{\mathcal{X}}(z_2) = \tau(\mu - \nu) + z_3/\lambda \,.$$

In the end we obtain:

$$\begin{cases} |z_2|_{\mathcal{X}^d,1} - \langle z_2, \nabla_{\mathcal{X}}(\phi)\rangle_{\mathcal{X}^d} = 0 \,, \\ |z_3|_{\mathcal{X},1} - \langle z_3, \phi/\lambda\rangle_{\mathcal{X}^d} = 0 \,, \\ \text{div}_{\mathcal{X}}(z_2/\tau) = \mu - \nu + z_3/(\lambda\tau) \,. \end{cases}$$

These three conditions, together with $|\nabla_{\mathcal{X}}\phi|_{\mathcal{X},\infty} \le 1$, $|\phi|_{\mathcal{X},\infty} \le \lambda$, imply that $(\phi, \sigma = z_2/\tau, \delta = z_3/\tau\lambda)$ is a saddle point for problems (4.2)-(4.1). Indeed:

$$|\sigma|_{\mathcal{X}^d,1} + \lambda|\delta|_{\mathcal{X},1} = |\sigma|_{\mathcal{X}^d,1} + \lambda|\delta|_{\mathcal{X},1} + \langle \phi, \text{div}_{\mathcal{X}}(\sigma) + \nu - \mu - \delta\rangle_{\mathcal{X}}$$
$$= |\sigma|_{\mathcal{X}^d,1} - \langle \nabla_{\mathcal{X}}(\phi), \sigma\rangle_{\mathcal{X}^d} + \lambda|\delta|_{\mathcal{X},1} - \langle \phi, \delta\rangle_{\mathcal{X}} + \langle \phi, \nu - \mu\rangle_{\mathcal{X}}$$
$$= \langle \phi, \nu - \mu\rangle_{\mathcal{X}} \,.$$

Therefore (σ, δ) defined in this way is optimal for the primal problem (4.1).

Consider now the case $q = 2$. Repeating the previous reasoning we obtain $z_1 = \tau(\mu - \nu) + \frac{\tau}{\lambda}\phi$,

$$\begin{cases} |z_2|_{\mathcal{X}^d,1} - \langle z_2, \nabla_{\mathcal{X}}(\phi)\rangle_{\mathcal{X}^d} = 0 \,, \\ \text{div}_{\mathcal{X}}(z_2/\tau) = (\mu - \nu) + \phi/\lambda \,, \end{cases}$$

with the further condition $|\nabla_{\mathcal{X}}(\phi)|_{\mathcal{X}^d,\infty} \le 1$. Therefore, in the same way as before, we deduce that $(\sigma = z_2/\tau, \delta = \phi/\lambda)$ is a solution to (4.1). $\qquad\square$

Remark 4.5 (Stopping criterion)**.** *Thanks to Proposition 4.4, we can consider*

$$(\sigma^n, \delta^n) = (z_2^n/\tau, \text{div}_{\mathcal{X}}(z_2^n)/\tau + \nu - \mu)$$

as a candidate solution to (4.1) *in order to estimate the duality gap. The SDMM algorithm solves the problem in the form*

$$\min_{\phi, \boldsymbol{y}} \left\{ \sum_i^N f_i\left(y_i\right), \, L_i\phi = y_i, \, 1 \le i \le N \right\},$$

and $\sum_i^N f_i\left(y_i^n\right) = f_1(y_1^n)$ *by definition of the algorithm. We can therefore control the (relative) duality gap by*

$$\Delta^n = \frac{1}{|f_1(y_1^n)|} \left| |\sigma^n|_{\mathcal{X}^d, 1} + \frac{\lambda}{q} |\delta^n|_{\mathcal{X}, q}^q + f_1(y_1^n) \right|.$$

Of course, without a control on the constraints $L_i\phi = y_i$ *this information is meaningless. We can evaluate the (relative) error on the constraints as*

$$r^n = \frac{\sum_i^N |L_i\phi^n - y_i^n|_{\mathcal{X}, 2}}{\max\left(\sum_i^N |L_i\phi^n|_{\mathcal{X}, 2}, \, \sum_i^N |y_i^n|_{\mathcal{X}, 2} \right)}.$$

Then, for $\varepsilon \in \mathbb{R}_+$, *we can set for simplicity as stopping criterion* $\max(\Delta^n, r^n) \le \varepsilon$.

Remark 4.6. *In order to solve problem* (4.1) *for* $(p, q) = (1, 2)$ *we may try to adopt proximal-splitting techniques that take advantage of the higher regularity of the objective functional, such as the accelerated forward-backward algorithm* [16] *or the Chambolle-Pock algorithm* [8]. *Nevertheless, we have not experienced a sensible speed-up by using these algorithms with respect to the SDMM, at least for the regimes of* λ *of interest. Some of these methods can avoid the inversion of the Laplacian, which makes individual iterations faster. However, the number of iterations required to reach convergence increases so much that it is not cost effective eventually.*

Remark 4.7 (Operator norm)**.** *Throughout this paper, we chose to formulate vector-valued* L^1 *optimal transport problem* (3.2) *using the gradient norm* $|\nabla\phi|_{V^d}$, *which corresponds to the Frobenius norm in the identification* $V^d = L(\mathbb{R}^d; V)$; *the flow norm* $|\sigma|_{V^d}$ *also appears in the dual formulation* (3.3). *Alternatively, we may have used the operator norm* $|\nabla\phi|_{L(\mathbb{R}^d; V)}$ *for the gradient, and thus the dual norm for* $|\sigma|_{L(\mathbb{R}^d; V)^*} = \mathrm{Tr}(\sqrt{\sigma^\top \sigma})$ *for the flow. The operator norm has the advantage of preserving the connection* (2.3) *with Lipschitz regularity, which only holds up to a multiplicative constant in the case of the Frobenius norm* (2.1). *The adaptation of the theoretical results and algorithms presented in the paper is straightforward, except for the proximal operator of* $i_{\mathcal{B}_{\mathcal{X}^d}}$, *which is obtained by pointwise orthogonal projection* (4.7) *at each site* (i, j) *of the grid onto the unit ball of* $V^d = L(\mathbb{R}^d; V)$ *w.r.t. the chosen norm. In this specific step, the normalisation procedure* (4.7) *which is adequate for* $|\cdot|_{V^d}$, *should be replaced with the orthogonal projection*

$$(4.9) \qquad\qquad \mathrm{argmin}\{|A - B|_{V^d}, \, |B|_{L(\mathbb{R}^d; V)} \le 1\}.$$

Assuming that $V = \mathbb{R}^n$, *up to choosing an orthonormal basis, the solution to* (4.9) *is obtained* [6, Theorem 2.1] *as* $B_{\mathrm{opt}} = U \, \mathrm{diag}(\min(\lambda_1, 1), \cdots, \min(\lambda_d, 1))V$, *where* $A = U \, \mathrm{diag}(\lambda_1, \cdots, \lambda_d)V$ *is the singular value decomposition (with* $U \in \mathbb{R}^{d \times d}$ *and* $V \in \mathbb{R}^{d \times n}$ *obeying* $UU^\top = VV^\top = \mathrm{Id}_d$, *and* $\lambda_1, \cdots, \lambda_d \ge 0$). *From the numerical standpoint the added numerical cost is linear in the problem size, since* d *and* n *are small and independent of the grid dimensions.*

5. Numerical results and application to FWI

In this section we consider the application of the L^1 transport for vector valued measures to a specific inverse problem: the full waveform inversion (FWI) problem.

5.1. **The acoustic wave equation and the FWI inverse problem.** The acoustic wave equation with constant density ρ can be written as (where here ∇ and div are the classical gradient and divergence operator, corresponding to the case $V = \mathbb{R}$ in Section 2)

$$(5.1) \qquad \begin{cases} \partial_t v - \frac{1}{\rho} \nabla p = 0 \\ \partial_t p - \rho c^2 \operatorname{div}(v) = f \end{cases}$$

where $p : D \subset \mathbb{R}^m \to \mathbb{R}$ is the scalar pressure field, $v : D \to \mathbb{R}^m$ is the vector field of particles' velocity and f is a given source term. The objective of FWI, in the simpler acoustic setting, is to reconstruct the velocity of the medium $c : D \to \mathbb{R}_+$ from observed data v^{obs}, p^{obs}, which are signed oscillatory functions.

We restrict here to the two dimensional case for simplicity, $m = 2$, and we consider two-dimensional vector valued data $v^{obs} = (v_x^{obs}, v_z^{obs})$. In this setting, we can introduce the space of symmetric matrices $V = \mathrm{S}^2$ and lift v^{obs} in the self-dual cone of symmetric positive semi-definite matrices $C = \mathrm{S}_+^2$ (thus the lifted data dimension is $n = \dim(\mathrm{S}^2) = 3$). We consider for this purpose Pauli's transformation (in dimension 2) as a lift operator L:

$$(5.2) \qquad v = (v_x, v_z) \longmapsto L(v_x, v_z) = \begin{bmatrix} \alpha - v_x & v_z \\ v_z & \alpha + v_x \end{bmatrix}, \quad \text{with } \alpha = \sqrt{v_x^2 + v_z^2}.$$

The advantage of (5.2) is that it is bijective from $\mathbb{R}^2$ to its domain, i.e. no information is lost in the transformation. Another possibility could be to consider the tensor product

$$v = (v_x, v_z) \longmapsto v^T v = \begin{bmatrix} v_x^2 & v_x v_z \\ v_x v_z & v_z^2 \end{bmatrix},$$

however this lift is invariant with respect to sign changes of v which limits its effectiveness. This sign invariance amounts to doubling the frequency of oscillating signals, and thus to a greater risk of *cycle skipping* in the application of interest (see Section 5.3 below).

We write the FWI problem as

$$(5.3) \qquad \min_c \sum_{s=1}^{N_s} \mathcal{T}_{p,q}^\lambda(d_s[c], d_s^{obs}), \quad \text{such that (5.1) holds,}$$

where $d_s[c] = L(R v_{x,s}, R v_{z,s})$ and $v_{x,s}, v_{z,s}$ are the computed particles' velocities given the model c (for simplicity omitting the dependencies), $d_s^{obs} = L(R v_{x,s}^{obs}, R v_{z,s}^{obs})$. The operator R evaluates the velocity at specific points in the domain, the receivers locations, for a certain time span. The peculiarity of FWI is that the whole recording, i.e. the whole form of the wave, is used in the inversion problem. The index s runs instead over all the possible sources f_s: the FWI problem is usually performed by running several simulations in parallel for different sources. Finally, L is the lift function defined in (5.2).

Remark 5.1. *The space of symmetric matrices $\mathrm{S}^{n'}$, endowed with the Frobenius scalar product, is a finite dimensional Euclidean space and is thus isometric to $\mathbb{R}^{n'(n'+1)/2}$ endowed with the usual scalar product. When $n' = 2$, such an isometry is $\begin{pmatrix} a & b \\ b & c \end{pmatrix} \mapsto 2^{-\frac{1}{2}}(\frac{a-c}{2}, b, \frac{a+c}{2})$, and it maps the cone S_+^2 of symmetric positive semi-definite matrices onto the Lorentz cone $\{(x, z, t) \in \mathbb{R}^3, \sqrt{x^2 + z^2} \le t\}$. This isometry eases the numerical implementation. In these coordinates, Pauli's lift above reads as $(v_x, v_z) \mapsto (v_x, v_z, \alpha)$, up to a multiplicative constant.*

5.2. **Calibration of λ.** The parameter λ controls mass transport by tuning the penalization on the divergence constraint. On the one hand, if it is too small, mass can be deformed without being transported and the model is intuitively close to an L^q distance on the difference of the two measures: $\mathcal{T}_{p,q}^\lambda \sim \frac{\lambda}{q} \int_\Omega |\mu - \nu|_V^q$ as $\lambda \to 0$. On the other hand, it is important to notice that in the seismic imaging application the two measures can have different total mass. This mass is in fact related to the energy of the signals if one uses Pauli's lift above and considers $q = 2$ (which is our original motivation for considering $q = 2$), or the tensor product lift with $q = 1$. Observe nevertheless that the total energy of the signals can be quite different, since these are recorded only partially in the domain. In the other cases, such as Pauli's lift with $q = 1$, the total mass is a non-physical quantity, which has no reason to be the same. One can check that $\mathcal{T}_{p,q}^\lambda \sim \frac{\lambda}{q} |\int_\Omega \mu - \int_\Omega \nu|_V^q$ as $\lambda \to \infty$. The misfit function thus becomes proportional to the q-th power of the difference between the masses of μ and ν, for large λ, which is not a very relevant objective for the problem of interest. The L^q and the difference of masses behaviors lead to poor results in the inversion problem, and the parameter λ thus has to be carefully selected in order to avoid these.

Assume that the physical dimension of the measures is an amplitude, which is denoted $[\mu] = [\nu] = A$. Consequently the physical dimension of the flux is an amplitude times a length, $[\sigma] = Al$, and the physical dimension of λ is $[\lambda] = \frac{l^p}{A^{q-p}}$. For $(p, q) = (1, 1)$, the parameter λ thus represents a length, whereas for $(p, q) = (1, 2)$ it represents a length over the amplitude. Examples 3.6 and 3.10 help having a better understanding. For $(p, q) = (1, 1)$, λ provides half the maximum distance up to which mass can be transported, regardless of the mass involved. By knowing the maximum displacement we want to allow in the transport problem, it can be tuned accordingly. For $(p, q) = (1, 2)$, λ specifies the maximum distance up to which a unit of mass can be transported. This means that large amounts of mass can be transported at large scales, whereas small amounts only at small scales, proportionally to their mass. These considerations can be of help for the calibration of this parameter in practice. The maximum transport distance can be set by comparing the initial and final measures in order to allow the transport of the main structures (see for example Figure 3). In general, in the specific case of transporting signals, we may want to allow displacements of the order of the wavelengths involved. Smaller values are not relevant whereas bigger values may be dangerous due to the mass imbalance. The wavelength can be estimated from the frequency of the injected signal. The case $(p, q) = (1, 2)$ is more delicate and the parameter λ can be estimated taking into account the minimum value of mass allowed to be transported at a certain distance.

We conclude this section by highlighting that the calibration of λ, following the considerations just made, depends on the scaling of the domain and, in the case $q = 2$, also of the measures. A simple computation shows that if Ω is rescaled by a factor L then λ has to be scaled by the same factor, in both cases. If the measures are scaled by a factor M instead, in the case $q = 2$, then λ has to be rescaled by the same factor again. This scaling behavior, which is consistent with the above physical dimension considerations, has to be taken into account when solving practical instances (see Section 5.4).

5.3. **Sensitivity to shifts of $\mathcal{T}_{p,q}^\lambda$.** Optimal transport is generally used in inverse problems in order to acquire sensitivity to (time) shifts in the data, a feature which is not encoded using L^p distances. In seismic imaging, this is referred to as *cycle-skipping*. Due to signed data, optimal transport based misfit functions may fail to provide this sensitivity (see Examples 3.6 and 3.7). By lifting the data to the space $\mathcal{M}(\Omega; C)$ we aim at solving this issue. Here we test for this reason the different behavior of the misfit functions $\mathcal{T}_{p,q}^\lambda$, for the different values

of the parameters p, q and λ. We consider the following signal

$$v(t) = (v_x(t), v_z(t)) = \left(\frac{d}{dt} f(t), \frac{d^2}{dt^2} f(t) \right),$$

where $f(t) = \frac{1}{\sqrt{2\pi}\alpha} \exp^{-t^2/\alpha^2}$ with $\alpha = \frac{4}{3}$, and we evaluate the misfit between the signal itself and $v(t - t_0)$ for different values of t_0. The results are shown in Figure 2 and are compared with the L^2 and the KR norms.

The L^2 norm presents more local minima, explaining the drawback of using it in the FWI problem. The KR norm (1.6) presents several minima as well for small values of the parameter λ. For a sufficiently big value, there is a single well-defined minimum, but the misfit saturates to a constant value. The time average of $v_x(t)$ and $v_z(t)$ is in fact zero and the KR norm is therefore not sensitive to big shifts (see Example 3.7). For signals with non-zero mean the KR norm recovers sensitivity. However, typically signals in seismic imaging do have zero mean, which is a major issue for its application in this setting. The misfits $\mathcal{T}_{p,q}^{\lambda}$ are not convex with respect to the shift either, but for a sufficiently big value of λ they have a unique minimum, that is as long as λ is sufficiently big to allow sufficient transport of mass. We can also notice, empirically in this example, that $\mathcal{T}_{1,1}^{\lambda}$ is always increasing with respect to the shift, an advantage of our construction with respect to the KR norm (1.6) applied to signed measures. Finally, for $p = 1$ the misfit function is more selective around the minimum contrary to the case $p = 2$, where it is more flat, suggesting why the H^{-1} norm may not be a good choice.

5.4. **Transport between seismograms.** We perform now a test case computing the unbalanced L^1 optimal transport problem between two-dimensional vector valued seismograms, in order to further inspect the influence of the parameter λ and to compare the two cases $(p, q) = (1, 1)$ and $(p, q) = (1, 2)$. The setting for the test is as follows. We consider a two-dimensional physical domain $D = [a, b] \times [c, d] \subset \mathbb{R}^2$. In D, N_r receivers are positioned uniformly along the x axis at a certain constant depth z. The data domain, corresponding to the registration time of each receiver, is given by $\Omega = [0, T] \times \{1, .., N_r\}$ and can be considered as a semi-discrete bidimensional domain. This allows to interpret the measurements as semi-discrete bidimensional seismograms. By considering a regular time sampling for (v_x, v_z), this provides a regular Cartesian grid $\mathcal{G}$. The components of the lift (5.2) are presented in Figure 3 for two examples. They correspond to the lifted particles' velocities generated through the acoustic wave equation (5.1) with two different velocity models c. In these examples, the traces are registered by $N_r = 169$ receivers and regularly sampled 3000 times. The single point source f is placed in the same location as the 84-th receiver and injects a Ricker signal.

We solve the problem between the two vector valued measures μ and ν represented in Figure 3. In order to avoid anisotropic effects due to the different dimensions of Ω, and to have a meaningful transport problem, we rescale one of the two directions. Indeed, on the x direction the physical dimension is a length whereas on the z direction it is time. For example, we can rescale the time by the average velocity of the model, $\bar{c} \approx 2000 \, m/s$. Considering the recording time of ≈ 7 seconds and the horizontal size of the domain D of 16.9 kilometers, this provides approximately a square domain. We consider therefore $\Omega = [0, \ell]^2$, with $\ell = N_r - 1$. We stress that this scaling has an influence on the choice of the parameter λ, as highlighted in Section 5.2. In Figures 4, 5 and 6, we show the three components of the computed optimal δ measures for different values of λ, in the two cases $q = 1$ and $q = 2$. The value of δ determines how much mass is injected in the problem and therefore gives a measure of what is or is not transported. It can be compared with the difference of the two measures, i.e. the zero transport case (Figure 3).
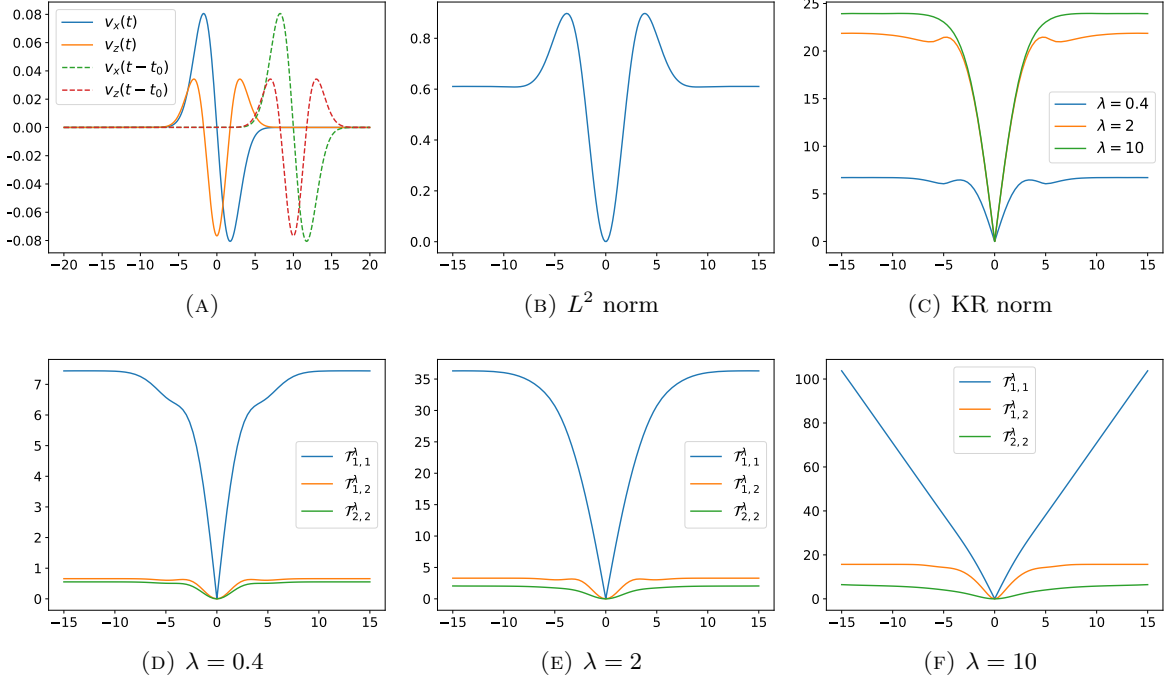
FIGURE 2. Sensitivity to cycle-skipping of $\mathcal{T}_{p,q}^{\lambda}$ (d-f), for different values of λ, and comparison with the L^2 distance (b) and the KR norm (1.6). In each graph (b-f) it is represented the evaluation of the misfit function between the multicomponent signals $(v_x(t), v_z(t))$ and $(v_x(t - t_0), v_z(t - t_0))$ (a), with respect to the time shift t_0. For the L^2 and KR norms, the distance is evaluated component-wise and summed.

First of all notice that, in both case $q = 1$ and $q = 2$, the order of magnitude of the first two components of δ (Figures 4 and 5) decreases for increasing values of λ. Mass is effectively transported on these components for increasing values of λ. On the contrary, the third component (Figure 6) does not change noticeably and for big values of λ it dominates on the first two by several orders of magnitude. The difference of mass between μ and ν concentrates in fact in the third component, since the velocities v_x and v_z have nearly zero average. If λ is too big, this component is therefore predominant in the transport problem and the misfit $\mathcal{T}_{p,q}^{\lambda}$ will have an undesired behavior, as mentioned in Section 5.2.

For $q = 1$, we can notice as expected that the mass can be transported only at a maximum distance, depending on the value of λ. For small values ($\lambda = 10^{-3}\ell$ or $\lambda = 10^{-2}\ell$), mass can be transported at very short distance and δ resembles the norm of the difference of the two measures, meaning that the model is close to an L^q distance. The main structures are well transported instead for a sufficiently big value ($\lambda = 10^{-1}\ell$). For $\lambda = \ell$ the mass is allowed to be transported everywhere in the domain (recall $\Omega = [0, \ell]^2$), however, due to the difference in total mass, δ is non zero. In this case, differently from the previous smaller values of λ, the third component of δ is signed and represents the unavoidable addition of mass. Increasing λ does not change at this point the solution but only makes this latter term more predominant in the overall cost.
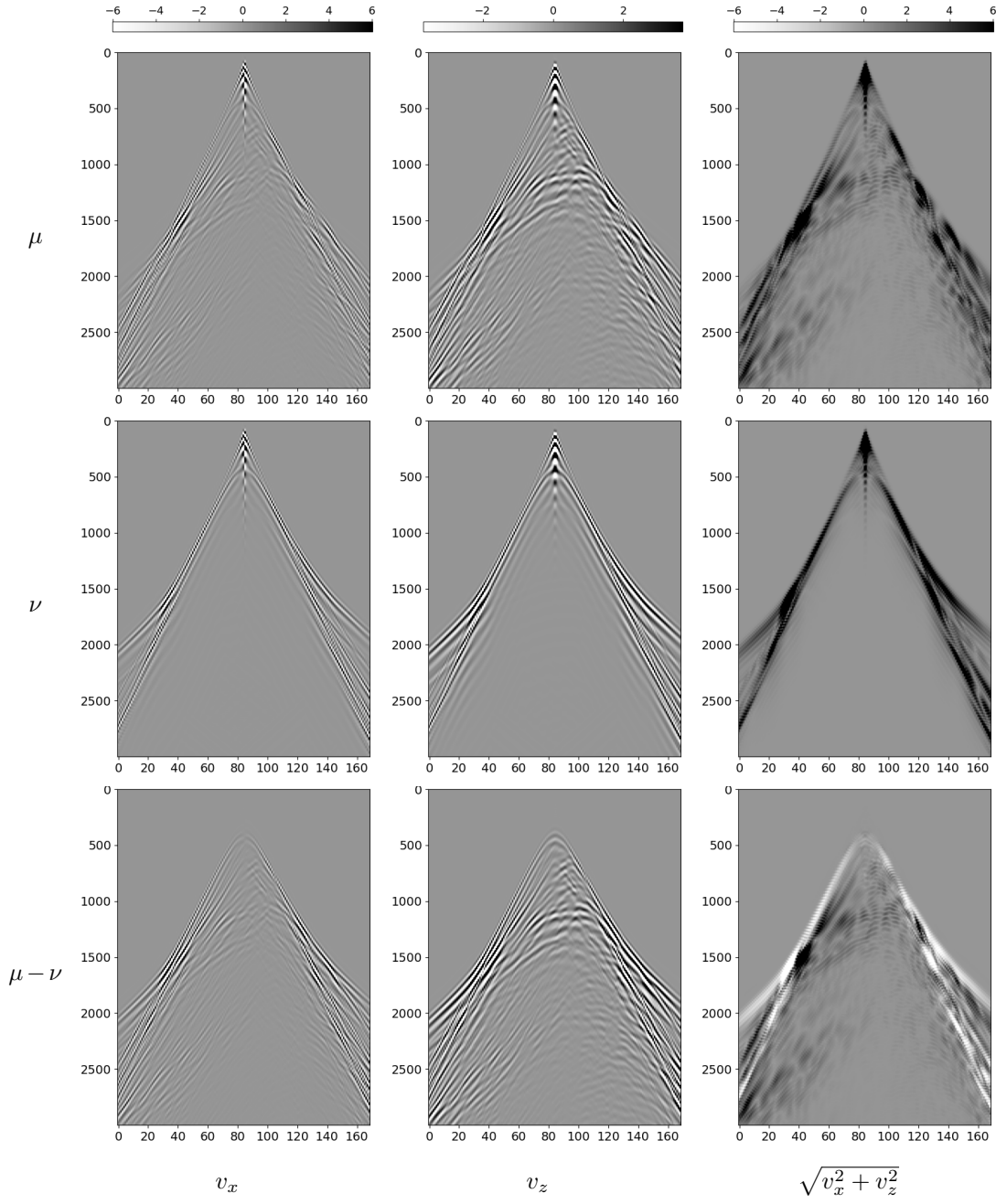
FIGURE 3. Example of two-dimensional seismograms of particles' velocities generated with two different velocity models c, respectively the *true model* and the *initial model 3* represented in Figure 8, lifted via (5.2) (see also Remark 5.1). Recording of 3000 regular time samples (y axis) from 169 evenly spaced receivers (x axis). For the sake of the representation, data are clipped to 10% of their maximum value and the color scale varies in each column.

On the other hand, the situation changes completely in the case $q = 2$. No matter the value of λ, sufficiently big values of mass can always be transported at every distance, resulting in a rather diffused solution. On the contrary, sufficiently small values of mass will not be transported, which is the reason why the third component of δ is not signed even for $\lambda = \ell$, differently from the case $q = 1$. Transport of the whole amount of available mass can be recovered only in the limit $\lambda \to \infty$. For the small values of λ, δ resembles again the difference of the two measures, although more diffused than the case $q = 1$ since the bigger values of mass have already started to match. For the bigger value $\lambda = 10^{-1}\ell$ we can notice that more or less all the structures of the signals are matched. The biggest value $\lambda = \ell$ provides an extremely diffused solution corresponding to the residual mass (in the third component of δ).

In figure 7 we show the convergence profiles of Algorithm 1 for the different test cases here considered. In general, for the regimes of λ of interest, few hundreds iterations are sufficient to reach convergence up to a tolerance ε of $10^{-4}, 10^{-5}$. As one could expect, the case $(p, q) = (1, 2)$ converges faster than the case $(p, q) = (1, 1)$ thanks to the uniform convexity of the objective function, with better performance the smaller the parameter λ. For $(p, q) = (1, 1)$ and excessively big values of λ, the convergence profiles are extremely oscillatory and rather slow. This probably depends on the fact that $\mathcal{T}_{1,1}^{\lambda}$ is closer to a total variation norm for this values of λ, whose minimization is more involved. The convergence could probably be improved via a careful tuning of the parameter τ in Algorithm 1. In these tests we considered $\tau = 0.9$. However, we did not try to tune it since for the values of λ of interest the algorithm already performs quite well.

5.5. **Application to a realistic data set.** To validate our approach, we considered a realistic two-dimensional inversion test: the Marmousi case study [19]. The setting is the same one as in Section 5.4. We consider in this case $N_s = 124$ point sources, uniformly distributed along the x axis at the same depth as the receivers, all producing the same Ricker signal. The true velocity model $c : D \to \mathbb{R}_+$ we aim at reconstructing is shown in Figure 8.

The tests have been performed using the code *TOYxDAC_time* [40] of the *SEISCOPE project*[2] where the new misfit functions proposed in this article have been integrated. The acoustic wave equation is discretized using a finite difference approach. The optimization problem (5.3) is solved with a gradient based l-BFGS approach [27]. The computation of the gradient is obtained via the *adjoint state method* [33]. We considered a tolerance $\varepsilon = 10^{-2}$ in Algorithm 1, according to the stopping criterion presented in Remark 4.5. This relatively big value allows us to solve each optimization problem in a small number of iterations and limit in this way the computational cost for evaluating the misfit function. We have not experienced a sensitivity of the results of the test to considering smaller values of ε.

5.5.1. *Objective gradient via adjoint state method.* We present the method restricting to one source only for simplicity, f. We rewrite for the sake of notation the total cost as

$$(5.4) \qquad \mathcal{T}_{p,q}^{\lambda}(L(Rv_x[c], Rv_z[c]), L(v_x^{obs}, v_z^{obs})) = g(v_x[c], v_z[c]) \,.$$

The relation between c and v_x, v_z is given as a forward linear equation and can be seen as the constraint $F(c, v_x, v_z, p) = 0$, where

$$(5.5) \qquad \begin{cases} F_{v_x}(c, v_x, v_z, p) = \partial_t v_x - \frac{1}{\rho}\partial_x p = 0 \,, \\ F_{v_z}(c, v_x, v_z, p) = \partial_t v_z - \frac{1}{\rho}\partial_z p = 0 \,, \\ F_p(c, v_x, v_z, p) = \partial_t p - \rho c^2 \operatorname{div}(v) - f = 0 \,. \end{cases}$$
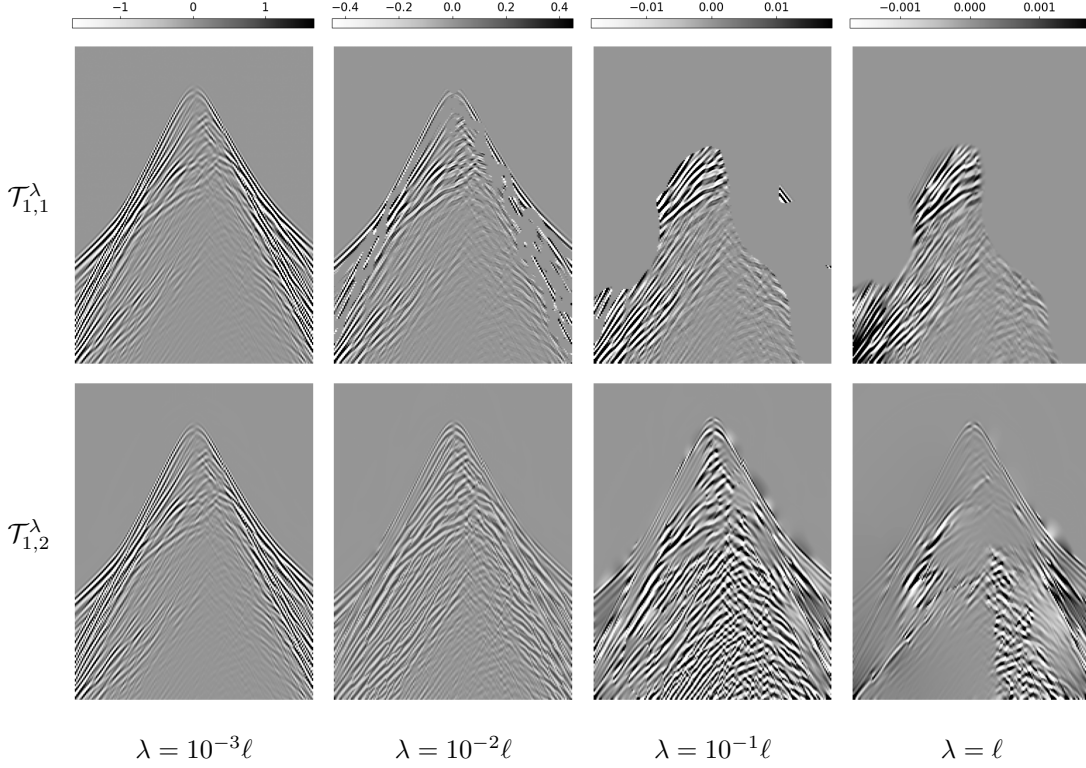
---

[2]https://seiscope2.osug.fr/

FIGURE 4. Plot of the first component of δ solution to the unbalanced L^1 optimal transport problem between μ and ν from Figure 3, for different values of the parameter λ (from first to fourth column), for $(p,q) = (1,1)$ (top row) and $(p,q) = (1,2)$ (bottom row). For the sake of the representation, data are clipped to 10% of their maximum value and the color scale varies in each column.

Considering then the velocities v_x, v_z and the pressure p as variables in the optimization problem, we can introduce the Lagrangian functional

$$(5.6) \qquad \mathcal{L}(c, v_x, v_z, p, \lambda) = g(v_x[c], v_z[c]) + \langle F(c, v_x, v_z, p), \lambda \rangle_D ,$$

where $\lambda = (\lambda_{v_x}, \lambda_{v_z}, \lambda_p)$ is the Lagrangian multiplier associated with the constraints and $\langle \cdot, \cdot \rangle_D$ is an appropriate scalar product in the model space. If we denote $v_x[c], v_z[c], p[c]$ the solution to the equation $F(c, v_x, v_z, p) = 0$ for a given velocity model c, then

$$\mathcal{L}(c, v_x[c], v_z[c], p[c], \lambda) = g(v_x[c], v_z[c])$$

and therefore

$$\frac{\mathrm{d}g(v_x[c], v_z[c])}{\mathrm{d}_c} = \partial_c \mathcal{L}(c, v_x[c], v_z[c], p[c], \lambda) + \partial_{v_x} \mathcal{L}(c, v_x[c], v_z[c], p[c], \lambda) \frac{\mathrm{d}v_x[c]}{\mathrm{d}_c}$$

$$+ \partial_{v_z} \mathcal{L}(c, v_x[c], v_z[c], p[c], \lambda) \frac{\mathrm{d}v_z[c]}{\mathrm{d}_c} + \partial_p \mathcal{L}(c, v_x[c], v_z[c], p[c], \lambda) \frac{\mathrm{d}p[c]}{\mathrm{d}_c} .$$

The adjoint state method consists in computing a specific multiplier $\bar{\lambda}[c]$ such that the partial derivatives with respect to v_x, v_z and p of the Lagrangian function vanish. In this way one
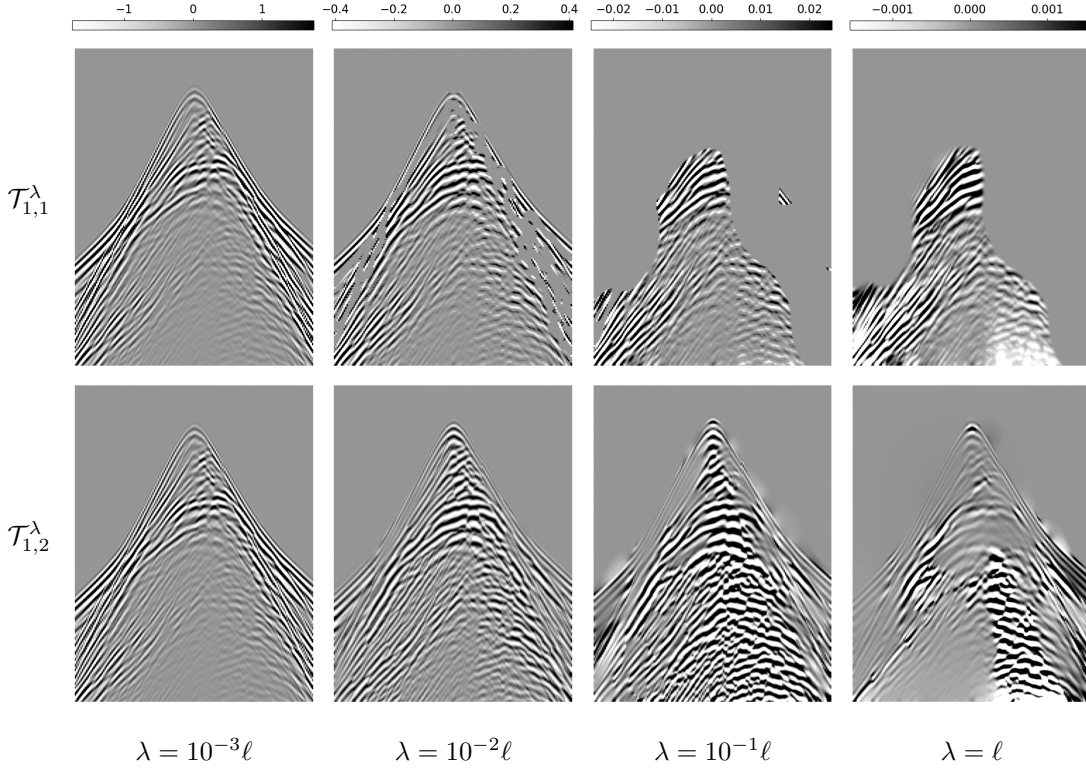
FIGURE 5. Plot of the second component of δ solution to the unbalanced L^1 optimal transport problem between μ and ν from Figure 3, for different values of the parameter λ (from first to fourth column), for $(p, q) = (1, 1)$ (top row) and $(p, q) = (1, 2)$ (bottom row). For the sake of the representation, data are clipped to 10% of their maximum value and the color scale varies in each column.

obtains the derivative of the objective function simply as

$$\frac{\mathrm{d}g(v_x[c], v_z[c])}{\mathrm{d}_c} = \partial_c \mathcal{L}(c, v_x[c], v_z[c], p[c], \bar{\lambda}) = \partial_c \langle F(c, v_x, v_z, p), \bar{\lambda} \rangle_D \,,$$

avoiding the expensive computation of the terms $\frac{\mathrm{d}v_x[c]}{\mathrm{d}_c}, \frac{\mathrm{d}v_z[c]}{\mathrm{d}_c}$ and $\frac{\mathrm{d}p[c]}{\mathrm{d}_c}$. From the definition of the Lagrangian function (5.6) and by the definition of the operator $F(c, v_x, v_z, p)$, $\bar{\lambda}$ can be computed as solution to:

$$(5.7) \quad \begin{cases} \partial_{v_x} \langle F^T(c, v_x, v_z, p), \bar{\lambda} \rangle_D = -\partial_t \bar{\lambda}_{v_x} + c^2 \rho \partial_x \bar{\lambda}_p = -\dfrac{\partial g(v_x[c], v_z[c])}{\partial v_x} \\[2mm] \partial_{v_z} \langle F^T(c, v_x, v_z, p), \bar{\lambda} \rangle_D = -\partial_t \bar{\lambda}_{v_z} + c^2 \rho \partial_z \bar{\lambda}_p = -\dfrac{\partial g(v_x[c], v_z[c])}{\partial v_z} \\[2mm] \partial_p \langle F^T(c, v_x, v_z, p), \bar{\lambda} \rangle_D = -\partial_t \bar{\lambda}_p + \dfrac{1}{\rho}(\partial_x \bar{\lambda}_{v_x} + \partial_z \bar{\lambda}_{v_z}) = 0 \,. \end{cases}$$

System (5.7) is the system of adjoint equations with respect to (5.5) and correspond to a backward acoustic wave equation.

From (5.4) and our definition of the cost function, taking advantage of the dual formulation ((3.7), (3.13) and (3.17) for the three different cases considered) which implies that $\mathcal{T}_{p,q}^\lambda$
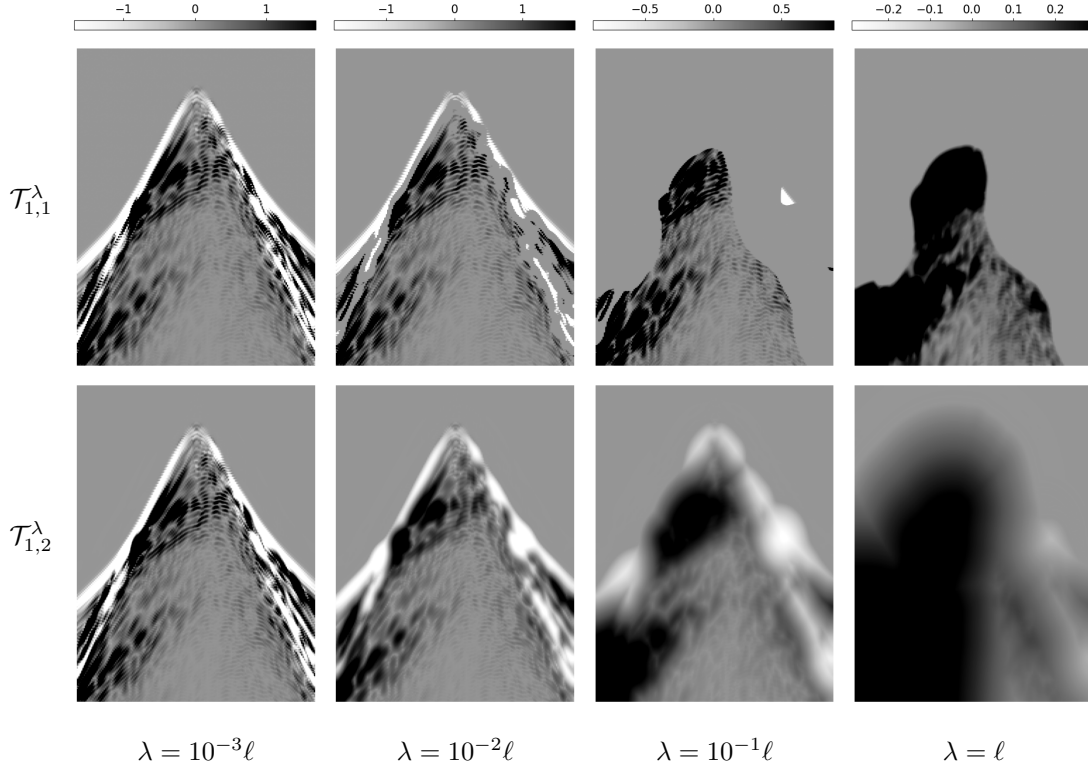
FIGURE 6. Plot of the third component of δ solution to the unbalanced L^1 optimal transport problem between μ and ν from Figure 3, for different values of the parameter λ (from first to fourth column), for $(p, q) = (1, 1)$ (top row) and $(p, q) = (1, 2)$ (bottom row). For the sake of the representation, data are clipped to 10% of their maximum value and the color scale varies in each column.
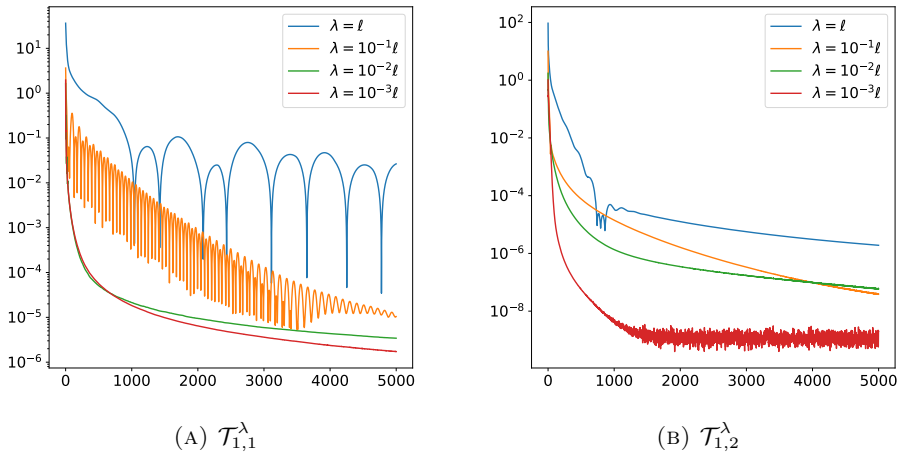

(A) $\mathcal{T}_{1,1}^{\lambda}$ (B) $\mathcal{T}_{1,2}^{\lambda}$

FIGURE 7. Convergence profiles of algorithm 1 for the error $\Delta^n + r^n$ in Remark 4.5, for the different test cases of Section 5.4.
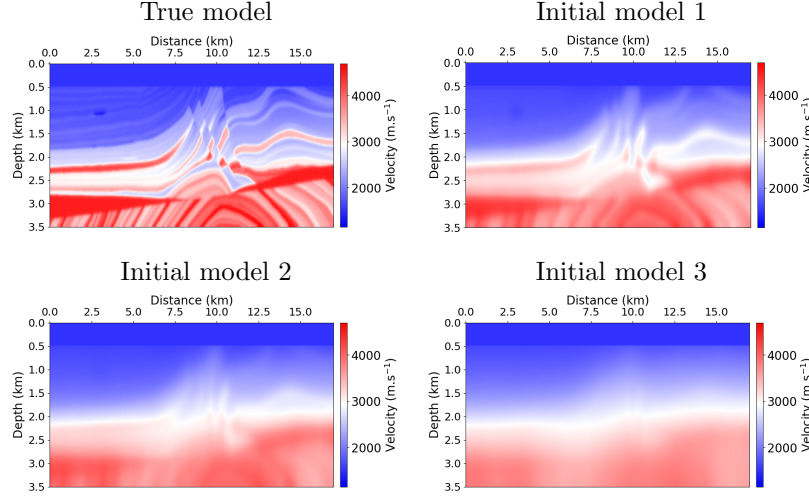
FIGURE 8. True velocity model of the Marmousi test case and several initial conditions for the inversion problem.

is a supremum of linear functionals in the measures, hence convex and almost everywhere differentiable, we can write

(5.8)
$$\frac{\partial \mathcal{T}_{p,q}^{\lambda}(L(Rv_x[c], Rv_z[c]), L(v_x^{obs}, v_z^{obs}))}{\partial v_x} = R^T \langle \partial_{v_x} L(Rv_x[c], Rv_z[c])), \bar{\phi} \rangle_V \,,$$
$$\frac{\partial \mathcal{T}_{p,q}^{\lambda}(L(Rv_x[c], Rv_z[c]), L(v_x^{obs}, v_z^{obs}))}{\partial v_z} = R^T \langle \partial_{v_z} L(Rv_x[c], Rv_z[c]))^T, \bar{\phi} \rangle_V \,,$$

where $\bar{\phi}$ is an (the) optimal potential for the dual formulation.

The computation of the gradient of the objective function in (5.3) requires solving each time the backward acoustic wave equation (5.7) with sources (5.8). The forward solution and the optimal potential are already provided by the evaluation of the objective function.

5.5.2. *Results.* We solve the problem for different values of the parameter (p, q) in the misfit $\mathcal{T}_{p,q}^{\lambda}$ and also compare with other misfit functions. We solve the inverse problem testing several initial models (Figure 8) of increasing difficulty, that is further and further away from the optimum we want to reconstruct. The objective is to measure the robustness of the inverse problem, that is the possibility to retrieve the minimum with less and less initial information. This is related to the convexity of the misfit function around the minimum. We do not consider in the comparison the computational cost related to the different misfit. The parameter λ is tuned manually to obtain the best performance, taking into account the considerations above.

In Figure 9 we show the results obtained for the three sets of parameters $(p, q) = (1, 1), (1, 2)$ and $(2, 2)$ for $\mathcal{T}_{p,q}^{\lambda}$. These results can be compared further with two other misfit functions: the classical L^2 distance and the KR norm (1.6). These latter are applied componentwise on the v_x and v_z signals. Figure 9 shows the final reconstructed velocity model c obtain from the (local) minimization of the misfit function, starting from several initial conditions (Figure 8, *initial models* from one to three) obtained as smoothed versions of the true velocity model we aim at recovering (Figure 8, *true model*).
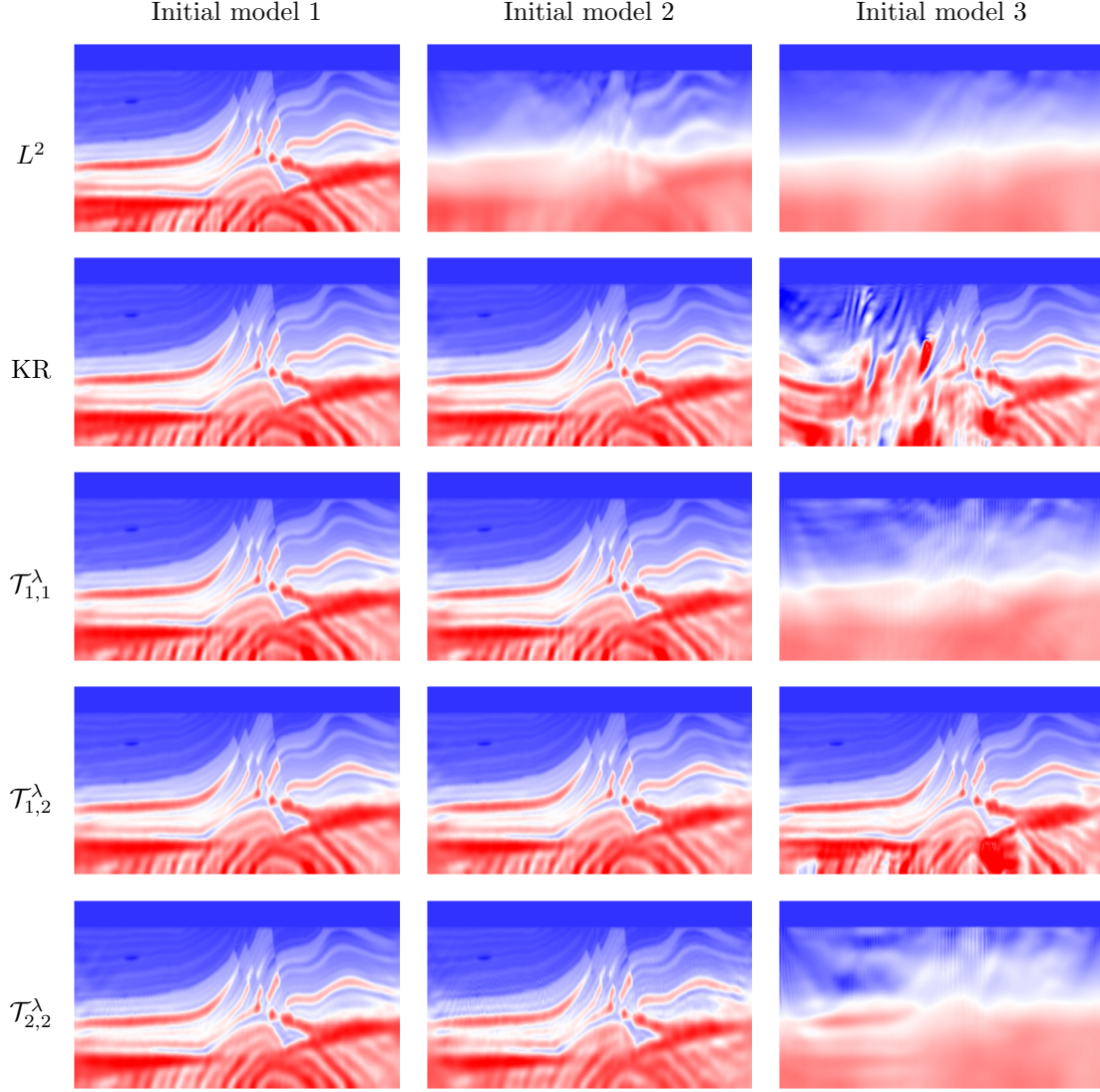
FIGURE 9. Comparison of the reconstruction of the Marmousi velocity model using the L^2 distance (first row), the KR norm (1.6) (second row) and $\mathcal{T}_{p,q}^{\lambda}$ (from third to fifth row), using several initial conditions for the inverse problem (Figure 8). For the misfit $\mathcal{T}_{p,q}^{\lambda}$, λ has been tuned manually to provide the better performance: we used $\lambda \approx 10^{-2}\ell$ for $(p,q) = (1,1)$, $\lambda \approx 0.05\ell$ for $(p,q) = (1,2)$ and $\lambda \approx 10^{-3}\ell$ for $(p,q) = (2,2)$, with ℓ size of the squared domain (see Section 5.4).

The L^2 distance does not present a big convexity region around the minimizer, as highlighted in Figure 2, which is the main drawback of using this simple misfit function. Consequently, the true velocity model can be retrieved only starting from a sufficiently close initial condition, else the l-BFGS algorithm would fall in a meaningless local minimum. Using optimal transport based distances may sensibly increase the convexity region, as recently

demonstrated, thanks to their horizontal sensitivity. Using the KR norm as misfit function allows indeed to solve the problem starting from further initial conditions. Notice that since the KR norm is applied separately to zero-mean scalar signals, therefore with no mass imbalance, it does not suffer from a higher bound on λ and the inverse problem is robust with respect to the choice of this parameter (as long as it is sufficiently high to allow proper displacement of mass, see Section 5.2). In this example we have chosen $\lambda = \ell$ (ℓ size of the squared domain, see Section 5.4). The horizontal sensitivity is however limited by the fact that the data are signed (see Examples 3.6 and 3.7). By lifting signed signals into the space of measures with symmetric positive semi-definite values, the misfit $\mathcal{T}_{p,q}^{\lambda}$ allows to go further. Among the three configurations, only the case $(p,q) = (1,2)$ provides considerably better results. The case $(p,q) = (1,1)$, the vectorial extension of the KR norm, suffers too much from the mass imbalance to allow a meaningful calibration of λ. The case $(p,q) = (2,2)$, as shown in Section 5.3, has a limited horizontal sensitivity and does not provide any advantage in terms of robustness with respect to the KR norm. On the other hand, it can be solved more efficiently with respect to the other two cases and the KR norm.

6. Conclusions

In this work we provided a misfit function for applications to inverse problems as a generalization of the L^1 optimal transport problem. By lifting general signed data to non-negative measures defined on a higher dimensional vector space, we recover a meaningful notion of transport leading to a more robust inversion problem. By choosing an appropriate lift function no information is lost on the data, differently from other previously proposed approaches. We considered the lift (5.2), that is using Pauli matrices to transform a vector with signed coordinates into a positive semi-definite matrix, but other ideas can be followed, our extension providing a general framework. The advantage of extending the L^1 transport, with respect to other optimal transport problems, is that it can be computed efficiently without a further increase in the computational complexity given by the vector extension. In particular, the quadratic penalization case may offer better optimization strategies with respect to the one considered here. On the other hand, the vector extension requires to deal with possible imbalances of mass which make the calibration of the model difficult. Our analysis of the role played by the parameter λ gives some insights on its calibration, but the model is quite sensitive to it. This in particular depends on the lift function considered and alternatives may be found to soften the problem. Moreover, the lift function may be tailored to the specific problem at hands taking into account also physical considerations in its design.

Acknowledgements

References

[1] Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 205(1):345–377, 2016.

[2] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[3] Achi Brandt. Multi-level adaptive solutions to boundary-value problems. *Mathematics of computation*, 31(138):333–390, 1977.

[4] Yann Brenier and Dmitry Vorotnikov. On optimal transport of matrix-valued measures. *SIAM Journal on Mathematical Analysis*, 52(3):2849–2873, 2020.

[5] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer New York, 2010.

[6] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on optimization*, 20(4):1956–1982, 2010.

[7] Eric A Carlen and Jan Maas. Gradient flow and entropy inequalities for quantum markov semigroups with detailed balance. *Journal of Functional Analysis*, 273(5):1810–1869, 2017.

[8] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40:120–145, 2011.

[9] Yongxin Chen, Tryphon T Georgiou, Lipeng Ning, and Allen Tannenbaum. Matricial wasserstein-1 distance. *IEEE control systems letters*, 1(1):14–19, 2017.

[10] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Matrix optimal mass transport: a quantum mechanical approach. *IEEE Transactions on Automatic Control*, 63(8):2612–2619, 2017.

[11] Yongxin Chen, Eldad Haber, Kaoru Yamamoto, Tryphon T Georgiou, and Allen Tannenbaum. An efficient algorithm for matrix-valued and vector-valued optimal mass transport. *Journal of Scientific Computing*, 77:79–100, 2018.

[12] Lénaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. An interpolating distance between optimal transport and Fisher–Rao metrics. *Foundations of Computational Mathematics*, 18(1):1–44, 2018.

[13] Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Unbalanced optimal transport: Dynamic and kantorovich formulations. *Journal of Functional Analysis*, 274(11):3090–3123, 2018.

[14] Krzysztof J Ciosmak. Matrix hölder's inequality and divergence formulation of optimal transport of vector measures. *SIAM Journal on Mathematical Analysis*, 53(6):6932–6958, 2021.

[15] Krzysztof J Ciosmak. Optimal transport of vector measures. *Calculus of Variations and Partial Differential Equations*, 60(6):230, 2021.

[16] Patrick L Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.

[17] Björn Engquist and Brittany D Froese. Application of the wasserstein metric to seismic signals. *Communications in Mathematical Sciences*, 12(5):979–988, 2014.

[18] Nikos Komodakis and Jean-Christophe Pesquet. Playing with duality: An overview of recent primal? dual approaches for solving large-scale optimization problems. *IEEE Signal Processing Magazine*, 32(6):31–54, 2015.

[19] Gary S Martin, Robert Wiley, and Kurt J Marfurt. Marmousi2: An elastic upgrade for marmousi. *The leading edge*, 25(2):156–166, 2006.

[20] Quentin Merigot and Boris Thibert. Optimal transport: discretization and algorithms. In *Handbook of Numerical Analysis*, volume 22, pages 133–212. Elsevier, 2021.

[21] Ludovic Métivier, Aude Allain, Romain Brossier, Quentin Mérigot, Edouard Oudet, and Jean Virieux. Optimal transport for mitigating cycle skipping in full-waveform inversion: A graph-space transform approach. *Geophysics*, 83(5):R515–R540, 2018.

[22] Ludovic Métivier, Romain Brossier, Félix Kpadonou, Jérémie Messud, and Arnaud Pladys. A review of the use of optimal transport distances for high resolution seismic imaging based on the full waveform. *arXiv preprint arXiv:2204.08514*, 2022.

[23] Ludovic Métivier, Romain Brossier, Quentin Merigot, Edouard Oudet, and Jean Virieux. An optimal transport approach for seismic tomography: Application to 3d full waveform inversion. *Inverse Problems*, 32(11):115008, 2016.

[24] Markus Mittnenzweig and Alexander Mielke. An entropic gradient structure for lindblad equations and couplings of quantum systems to macroscopic models. *Journal of Statistical Physics*, 167:205–233, 2017.

[25] Lipeng Ning and Tryphon T Georgiou. Metrics for matrix-valued measures via test functions. In *53rd IEEE Conference on Decision and Control*, pages 2642–2647. IEEE, 2014.

[26] Lipeng Ning, Tryphon T. Georgiou, and Allen R. Tannenbaum. Matrix-valued monge-kantorovich optimal mass transport. *52nd IEEE Conference on Decision and Control*, pages 3906–3911, 2013.

[27] Jorge Nocedal. Updating quasi-newton matrices with limited storage. *Mathematics of computation*, 35(151):773–782, 1980.

[28] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):127–239, 2014.

[29] Gabriel Peyré, Lenaïc Chizat, François-Xavier Vialard, and Justin Solomon. Quantum optimal transport for tensor field processing. *arXiv preprint arXiv:1612.08731*, 2016.

[30] Gabriel Peyré and Marco Cuturi. Computational optimal transport, 2020.

[31] Benedetto Piccoli and Francesco Rossi. Generalized wasserstein distance and its application to transport equations with source. *Archive for Rational Mechanics and Analysis*, 211:335–358, 2014.

[32] Benedetto Piccoli and Francesco Rossi. On properties of the generalized wasserstein distance. *Archive for Rational Mechanics and Analysis*, 222:1339–1365, 2016.

[33] R-E Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503, 2006.

[34] Lingyun Qiu, Jaime Ramos-Martínez, Alejandro Valenciano, Yunan Yang, and Björn Engquist. Full-waveform inversion with an exponentially encoded optimal-transport norm. In *SEG technical program expanded abstracts 2017*, pages 1286–1290. Society of Exploration Geophysicists, 2017.

[35] Ernest K Ryu, Yongxin Chen, Wuchen Li, and Stanley Osher. Vector and matrix optimal mass transport: theory, algorithm, and applications. *SIAM Journal on Scientific Computing*, 40(5):A3675–A3698, 2018.

[36] Filippo Santambrogio. *Optimal transport for applied mathematicians*. Springer, 2015.

[37] Paul N Swarztrauber. A direct method for the discrete solution of separable elliptic equations. *SIAM Journal on Numerical Analysis*, 11(6):1136–1150, 1974.

[38] Alexei Yu Uteshev. Analytical solution for the generalized fermat–toricelli problem. *The American Mathematical Monthly*, 121(4):318–331, 2014.

[39] C. Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003.

[40] Pengliang Yang, Romain Brossier, Ludovic Métivier, Jean Virieux, and Wei Zhou. A time-domain preconditioned truncated newton approach to visco-acoustic multiparameter full waveform inversion. *SIAM Journal on Scientific Computing*, 40(4):B1101–B1130, 2018.

[41] Yunan Yang and Björn Engquist. Analysis of optimal transport and related misfit functions in full-waveform inversion. *Geophysics*, 83(1):A7–A12, 2018.

Gabriele Todeschi (gabriele.todeschi@univ-eiffel.fr), Univ. Grenoble Alpes, ISTerre, F-38000 Grenoble, France; now LIGM, Univ. Gustave Eiffel, CNRS, F-77454 Marne-la-Vallée, France

Ludovic Métivier, (ludovic.metivier@univ-grenoble-alpes.fr) Univ. Grenoble Alpes, CNRS, LJK, ISTerre, F-38000 Grenoble, France

Jean-Marie Mirebeau, (jean-marie.mirebeau@ens-paris-saclay.fr) CNRS, Centre Borelli, ENS Paris-Saclay, University Paris-Saclay, 91190, Gif-sur-Yvette, France